Taxonomic status of *Melaleuca argentea* "Ashburton biotype" and identification and evaluation of exon capture loci for Myrtaceae

Bokyung Choi

February 2018

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

Declaration

I, Bokyung Choi, declare that the research in this thesis is my original work except where due reference is given in the text. All the studies in each data chapters 2, 3, and 4 were completed in collaboration with coauthors. However, I was the main contributor in all chapters. No part of this thesis has been submitted for any previous degree.

Bokyung Choi

Bokyung Choi

February 2018

Acknowledgements

I am grateful for all the great opportunities I have had. This study opened up new perspectives and it was possible with all the people I have interacted or worked with along the way.

I appreciate Celeste Linde for her supervision. Thank you for being supportive and encouraging throughout different stages of my PhD with constructive advice and letting me work with Leon to troubleshoot in your lab.

I thank Adrienne Nicotra, who has been very encouraging and a big support throughout my PhD.

I thank Mike Crisp from whom I learned so much about *Melaleuca* and other Australian plants in the field. Thank you for all the discussion and reading my drafts.

I appreciate Lyn Cook for constructive feedback and questions and for having me at Cook lab to work on Chapter 2.

I thank Carsten Külheim for paper discussions and making the labwork and identifications of markers possible.

Brendan Lepschi, Lyn Craven and Joe Miller who are my past and present supervisors at the Australian National Herbarium. I thank Australian state herbaria (CANB, PERTH, NSW) and CSIRO Black Mountain for giving me an access to specimens. I thank Anna Monro, Chris Cargill, Alexander Schmidt-Lebuhn, Maggie Nightingale, Jo Palmer, Dave Mallinson, Rosemary Purdie, and Bronwyn Collins at ANBG and CANB.

I thank Leon Smith for his skills and insights in the labwork particularly for troubleshooting. Nay Chi Khin for your support and all the work through many long hours of lab work and discussion with me. Monica Ruibal and Amanda Padovan for their help with labwork and helpful discussion. I am also grateful for Renae Pratt and Sally Potter for helpful discussion on exon capture.

I am grateful for Megan McDonald and Karen Meusemann for being wonderful mentors and all their support.

Leon Deschamps, Vicky Long, Stephen Van Leeuwen, Russell Barrett and Matt Barrett kindly collected plant material for this study.

Thanks to Crisp lab, Cook lab and Linde lab mates.

I thank friends at ANU for their friendship, support and all their help: Belinda, Marta, Tom, Meredith, Sarah (for also providing RNA-Seq data), Amanda P, Amanda E, Buddhie, Angel, Carlos, Iliana, Xia, Nay Chi, Dan S., Dani, David K., David D., Josh, Moos, Huw, Nina, Angel, Ana, Anna, Jessie, Sonya, Liam, Dom, Sandra, Sophie, Heejin, Bee, Melita and friends at GH (Jo, Bogdan, Savitri, Channa, Maria, Sam, Kimlong, Yoko, Mark), Bort, Jess and Kat.

I appreciate Thomas Semple, Meredith Cosgrove, Huw Ogilvie, Josh Penalba, Jason Bragg, Pat Blackwell, and Simon Ho for their helpful discussion and/or feedback on my work.

Jeremy Bruhl, Dianne Cook, Nunzio Knerr and Liam Bailey for their help with data analysis, Nunzio Knerr for mapping.

I am grateful to Professor Suk-Pyo Hong and my colleagues and friends at KHUS for their support and encouragements.

Marco Duretto for his encouragements, support and feedback on my research.

Friends from home: Colin, Jina, Suwan, Jihyun, Suye, Minhee, Lijing, Jinsu and Hwamyung for being there.

I thank the Australian Embassy in Seoul, Korea and the Australian government for making it possible for me to come here to study with Kapyong commemorative scholarship and for their support.

I thank my mum, dad and sisters in Korea for their love and support.

Abstract

Melaleuca sensu lato is the second biggest genus (ca. 380 spp.) that is mainly distributed in Australia, but some species are found in Southeast Asia, Papua New Guinea and New Caledonia. *Melaleuca* occupy a range of habitats and are found in all biomes, which makes this hyper-diverse genus an excellent group to carry out evolutionary studies.

The taxonomy of *Melaleuca sensu lato* (tribe Melaleuceae) has been controversial. The tribe originally included eight other genera: *Beaufortia*, *Calothamnus*, *Conothamnus*, *Eremaea*, *Lamarchea*, *Petraeomyrtus*, *Phymatocarpus*, and *Regelia*. However, with taxonomic changes all of the genera in the tribe were transferred to *Melaleuca.* To date, the phylogenetic studies based on molecular data have shown that *Melaleuca* is paraphyletic with respect to the other genera in the tribe. However, the taxonomic boundary of *Melaleuca* is still debated. Existing molecular phylogenies of Melelauceae have used only a few genetic markers, and the relationships within the genus remain unresolved with large polytomies.

Taxonomy of eucalypts has been problematic. Eucalypts include the following seven genera (*Eucalyptus, Corymbia, Angophora, Arillastrum*, *Allosyncarpia*, *Stockwellia*, *Eucalyptosis*). They are mainly Australian but some genera/species extend outside Australia. The taxonomic relationship of *Corymbia,* and *Angophora* is still debated. Some molecular phylogenies have shown monophyly of *Corymbia* and *Angophora* while *Corymbia* was paraphyletic with respect to *Angophora* in other studies.

In Chapter 2, the genetic diversity of *Melaleuca argentea* (*M. leucadendra* complex) was explored. Previous research suggested that populations in the Pilbara region of Western Australia had some genetic distinctions. The Pilbara populations were recorded as *M. argentea* "Ashburton biotype" (AB) in older literature, but no further information was available. Morphological characters and molecular data to assess the taxonomic status of the Pilbara population were used. We found that AB is more similar and closely related to *M. leucadendra* than to *M. argentea.* The results did not have conclusive evidence to support that AB is a distinct species from *M. leucadendra.* In order to test the taxonomic status, presence of gene flow between AB and *M. leucadendra* need to be further tested.

With the ultimate aim of estimating a comprephensive phylogeny of *Melaleuca* and eucalypts using more samples and many more loci compared to previous studies, we developed genetic markers for exon capture in Chapter 3. A workflow to locate orthologous and low copy number nuclear loci is introduced along with the method that was employed to identify the chloroplast markers. 209 chloroplast and nuclear loci that might be useful for Myrtaceae were identified by and 43 Myrtaceae taxa were successfully sequenced.

In Chapter 4, a gene tree approach for each individual locus was undertaken to remove potentially paralogous loci. We have found 144 loci that might be useful for *Melaleuca* and 174 loci for eucalypts. The present study contributes towards more robust estimations of phylogenetic relationships in the genus *Melaleuca,* eucalypts as well as other genera in Myrtaceae. Further work is required to verify the markers and to study phylogenetic relationships of the taxa at different taxonomic levels using the loci.

Thesis plan

This thesis is comprised of the following five chapters:

Chapter 1. Introduction: an overview of* Melaleuca *and eucalypts and their phylogenetic and taxonomic background.

I was responsible for writing and Celeste Linde and Mike Crisp provided editorial comments.

Chapter 2. Taxonomic status of* Melaleuca argentea *"ashburton biotype" of broad-leaf paperbarks (Myrtaceae,* Melaleuca leucadendra *species complex) from the Pilbara region of Western Australia.

Mike Crisp, Lyn Cook, Robert Edwards and I designed the study. Lyn Cook, her lab and Mike Crisp contributed to the genetic data. Robert Edwards provided a morphological character list that he had envisioned for this study. I collected morphological data, carried out analyses with guidance and assistance from Lyn Cook. I was also responsible writing. Lyn Cook, Mike Crisp and Celeste Linde provided editorial comments.

Chapter 3. Identification of genetic markers for genus to species level phylogenies of* Melaleuca *and other genera in Myrtaceae.

Carsten Külheim, Michael Crisp and Lyn Cook designed the study and CK developed the identification workflow. I was the main contributor to the genetic marker identification process and CK, Alicia Toon and Robert Edwards contributed to loci discovery during our workshops. CK designed the lab protocol and the details were modified by technicians, CK and I. I performed lab work with technicians and CK. AT and LC contributed to the whole genome shotgun sequencing of *Melaleuca leucadendra* and *M. bracteata*. Sarah Hsieh provided transcriptomes of *M. quinquenervia* and *M. alternifolia*. CK processed reads from the sequencing of target-captured libraries. Karen Musemann sorted and renamed headers to the sequences. I was responsible for the rest of analyses and writing. CK, MC, LC, and Celeste Linde provided editorial comments.

Chapter 4. Evaluation and comparison of 196 exon capture nuclear and chloroplast loci for Myrtaceae.

Celeste Linde, Mike Crisp, Lyn Cook and I designed the study. Carsten Külheim contributed to the data collection and processing as indicated above. Karen Musemann sorted and renamed headers to the sequences and aligned sequences. I was responsible for the rest of analyses and writing. CL, MC, and LC provided editorial comments.

Chapter 5. Discussion and concluding remarks.

I was responsible for writing and Celeste Linde and Mike Crisp provided editorial comments.

Table of contents

Acknowledgements……3

Abstract……5

Thesis plan……...7

Chapter 1……...……...……..10

Chapter 2…..……..…….....25

Chapter 3…….....63

Chapter 4…….....139

Chapter 5…….....175

Chapter 1

Introduction: an overview of *Melaleuca* and eucalypts and their phylogenetic and taxonomic background.

Melaleuceae (*Melaleuca sensu lato*): taxonomy, phylogenies

The Myrtaceae Juss. tribe Melaleuceae *sensu* Wilson *et al.* (2005; *Melaleuca sensu lato* Craven *et al.* 2014) includes approximately 380 species. Together with *Melaleuca* (290 spp.; *Melaleuca sensu stricto*), the following eight genera were included in the tribe: *Beaufortia* R.Br. (21 spp.), *Calothamnus* Labill. (40 spp.), *Conothamnus* Lindl. (3 spp.), *Eremaea* Lindl. (15 spp.), *Lamarchea* Gaudich. (2 spp.), *Petraeomyrtus* Craven (1 sp.), *Phymatocarpus* F.Muell. (3 spp.) and *Regelia* Schauer (5 spp.). Recently, Craven *et al.* (2014) transferred all the eight genera in *Melaleuca sensu lato* to *Melaleuca* (Table 1). *Melaleuca* is the second largest genus of Myrtaceae, next to *Eucalyptus*, in Australia.

Melaleuca s.l. is found in all biomes and diverse habitats across Australia (Fig. 1). It occupies various habitats ranging from desert, sclerophyll forest, clay flats, savannah and wetlands. Today, Australia can be divided into five biomes: Southwestern temperate sclerophyll biome (SWAFR: Southwest Australian Floristic Region), southeastern temperate sclerophyll, monsoonal, arid biome and everwet biome (Crisp *et al.* 2004, 2013; Kershaw *et al*. 1994; Fig. 1). The majority of the Australian flora is fire-dependent, sclerophyllous, and xeromorphic whereas little of the wet biome remains and a few families such as Myrtaceae, Proteaceae and genera such as *Eucalyptus* and *Acacia* dominate the present landscape (Crisp and Cook, 2013). *Melaleuca s.l.* is well represented in diversity hotspots such as southwest Western Australia (Fig. 1). In SWAFR, Myrtaceae are the most diverse angiosperm family and *Melaleuca s.l.* is one of the largest genera (Paczkowska & Chapman 2000; Beard *et al*. 2000). The high biodiversity of *Melaleuca s.l.* in the southwest temperate Australia contrasts to its relatively low diversity in the southeast temperate biome. Fine-scale patterns of soils might allow geographic overlap and it might result in high biodiversity (Crisp & Cook 2013). *Melaleuca* is a good example to study adaptation and evolution of Australian flora because it is species rich and occurs in such diverse range of habitats.

Melaleuceae includes ecologically significant species such as those in the *Melaleuca leucadendra* complex that dominate wetlands in the monsoon tropics. Many species are of high conservation importance, some are threatened and *M. wimmerensis* is critically endangered. The tribe also includes species of economic importance. Some species are used for essential oils (e.g. *Melaleuca alternifolia*) and others are popular ornamentals, such as bottlebrushes, formerly known as *Callistemon* (Craven 2006).

Some of the informal morphological characters shared by the group include persistent woody fruit and inconspicuous petals, however, no morphological synapomorphy (*Melaleuca sensu lato*) has been formally described for the tribe Melaleuceae (Edwards *et al.* 2010). The genus *Melaleuca* is based on the type specimen of *Melaleuca leucadendron* L. (L.), which was collected from Ambon, Indonesia (Blake 1968). *Melaleuca s. s.* was diagnosed by characters such as stamens that are united into five bundles, stamen bundles being opposite to petal, inflorescence type (bottle brush/head/spike), three locular capsules, sessile flowers with long stamens compared to petals (Bentham 1867; Blake 1968; Craven & Lepschi 1999). However, no character or combination of charaters can differentiate the genus from the tribe Leptospermeae (Edwards *et al.* 2010). Other genera in the tribe were erected based on few autapomorphies, which were largely based on ovule and anther characters (Edwards *et al.* 2010).

Brown *et al.* (2001) conducted a phylogenetic analysis on *Melaleuca* and related genera and showed that *Melaleuca* is paraphyletic. They used sequences of 5S rDNA and the ITS-1 region of the nuclear rDNA internal transcribed spacers. Edwards *et al.* (2010) conducted a phylogenetic analysis with a larger sample of species using sequences of cpDNA (*ndhF*) and morphological data, and verified that other members of the tribe were nested in *Melaleuca*. The separate generic status of *Melaleuca* from other genera in Melaleuceae was therefore not supported and the authors suggested that all genera in Melaleuceae (including *Callistemon*) should be placed into *Melaleuca* (Edwards *et al.* 2010; Table 1). However, other authors have expressed conflicting opinions on the classification of *Melaleuca* and *Callistemon* (Udovicic & Spencer 2012). Craven (2006) suggested *Callistemon* to be included in *Melaleuca* as the former was not distinct enough to be recognised as a genus and transferred *Callistemon* into *Melaleuca*. Udovicic and Spencer (2012) recognised *Callistemon* as a separate genus from *Melaleuca* because they considered that further study was needed to confirm phylogenetic relationships. Australian herbaria do not have a consensus about the classification of relationships between *Melaleuca* and *Callistemon* (APC 2012). State herbaria in the ACT, NT, QLD, and TAS include *Callistemon* within *Melaleuca*, but state herbaria in NSW, VIC, SA, and WA treat *Callistemon* as a separate genus (Udovicic & Spencer 2012). Currently, all of the state herbaria are neutral on the classification except that NSW is against including *Callistemon* in *Melaleuca* (K. Thiele pers. comm.).

Within *Melaleuca sensu stricto*, 12 informal species-groups have been recognised

(Barlow 1988; Barlow & Cowley 1988; Cowley *et al.* 1990; Quinn *et al.* 1992). The *Melaleuca leucadendra* complex (16 spp.) is one of those groups and is dominant in some communities of the Australian tropics and subtropics, thus making it ecologically significant (Barlow 1988; Boland *et al.* 1984; Cook *et al.* 2008; Craven & Cowie 2013; Franklin *et al*. 2007). Species in this group frequently occupy riverine gallery forests, coastal wetlands and flood-plain habitats (Boland *et al*. 1984; Franklin *et al*. 2007). The group extends beyond Australia to Indonesia, Malaysia, New Caledonia and New Guinea (Barlow 1988; Boland *et al*. 1984). From this group, *Melaleuca quinquenervia* is a problematic weed in the Americas (Wheeler 2006). It is also one of the most susceptible species within the Australian Myrtaceae to the recently introduced pathogen, myrtle rust, and thus potentially threatened in its native range (Carnegie & Lidbetter 2012).

Eucalypts: taxonomy, phylogenies

Eucalypts (Tribe Eucalypteae *sensu* Wilson *et al*. 2005; Myrtaceae) is a symbolic group of plants to the Australian vegetation and they include seven genera (Bayly *et al.* 2013; Bayly 2016). The largest genus, *Eucalyptus* L'Hér. (ca. 700 spp.), is mainly distributed in Australia with few species extending to New Guinea, Timor, and Southeast Asia (Bayly 2016). The second largest genus in the group is *Corymbia* K.D.Hill & L.A.S.Johnson (ca. 100 spp.), which is native to Australia and New Guinea. *Angophora* Cav. (ca. 10 spp.) is distributed in eastern Australia (Bayly *et al.* 2013; Ladiges *et al*. 2003). The rest of the four genera are monotypic to ditypic: *Arillastrum* Pancher ex Baill (1 sp.; New Caledonia), *Allosyncarpia* S.T.Blake (1 sp.; Australia), *Stockwellia* Carr, Carr & Hyland (1 sp.; Australia), *Eucalyptosis* C.T.White (2 spp.; New Guinea, Indonesia; Ladiges *et al.* 2003).

Taxonomic relationships among the genera *Eucalytus, Corymbia* and *Angophora* were controversial. The taxonomic position of *Angophora* and *Corymbia* were treated as separate genera from *Eucalyptus* (Hill & Johnson 1995). In contrast, Brooker (2000) classified *Angophora* and *Corymbia* as subgenera of *Eucalyptus.* However, this classification system (Brooker 2000) was not employed in later studies (Ladiges & Udovicic 2000; Slee *et al.* 2006). Steane *et al.* (2002) used ITS sequences and showed that *Corymbia* and *Angophora* were forming a clade with a good support and that they were well differentiated from *Eucalyptus sensu stricto*.

The phylogenetic relationship between the genera *Angophora* and *Corymbia* is still

debated. In earlier studies, monophyly of *Corymbia* was based on morphology and anatomy (Hill & Johnson 1995, Ladiges *et al.* 1999). Parra-O *et al.* (2010) used both nuclear molecular (ITS, ETS) and morphological data and showed that both *Corymbia* and *Angophora* form monophyletic sister groups. Bayly *et al.* (2013) carried out phylogenetic analyses using the chloroplast genome. In the study, *Corymbia* was paraphyletic as regards to *Angophora* with one clade of *Corymbia* being sister to *Angophora* and the other clade of *Corymbia* being sister to the rest: ((Cor,Ang),Cor). This result was consistent with Steane *et al*. (2002).

Eucalyptus includes ten subgenera (*Symphyomyrtus*, *Eucalyptus*, *Eudesmia, Alveolata, Acerosae, Cruciformes, Cuboidea, Idiogens, Minutifructus, Primitiva*; Bayly *et al.* 2016) and the taxonomy within the genus *Eucalyptus* is also debated. Whether subgenera *Eudesmia, Eucalyptus, Minutifructus*, and *Symphyomyrtus* are monophyletic or not has been questioned (Steane *et al.* 2002). In a phylogenetic analysis of whole chloroplast genomes, Bayly *et al.* (2013) showed that subgenus *Minutifructus* was nested within subgenus *Symphyomyrtus* while subgenus *Idiogenes* was nested within subgenus *Eucalyptus* (monocalypts).

Molecular markers and their usefulness

In Myrtaceae, universal nuclear and chloroplast markers have been utilised to explore subfamily, tribal to below generic level relationships (Grattapaglia *et al.* 2012). Wilson *et al*. (2001; 2005) utilised *matK* sequences and/or 5′ spacer regions and recognised the tribal classification system and the circumscriptions were well supported by Biffin *et al.* (2010; *matK*, ITS). Molecular phylogenies at intergeneric or intrageneric level within Melaleuceae and eucaltyps have also been informative tools to study the relationships among/within the genera, often challenging the traditional taxonomic classification system (e.g. Edwards *et al.* 2010; Steane *et al.* 2002).

To date, molecular phylogenies in Melaleuceae utilised a limited number of universal chloroplast/nuclear markers at different taxonomic levels (e.g. Edwards *et al.* 2010, 2013; Cook *et al.* 2008) and microsatellite or Simple Sequence Repeats (SSR) markers were also developed for or utilised at population level (e.g. Nevill *et al.* 2013). In *Melaleuca sensu lato,* chloroplast and/or nuclear markers have been useful to explore the generic relationship within the tribe Melelauceae (e.g. Brown *et al.* 2001, Edwards *et al.* 2010, Ladiges *et al*. 1999). Studies have, regardless of whether nuclear or chloroplast

marker was utilised, consistently showed non-monophyly of *Melaleuca sensu stricto* with rest of the genera in the Melaleuceae being nested within *Melaleuca sensu stricto* (Brown *et al.* 2001, Ladiges *et al*. 1999; Edwards *et al.* 2010).

Phylogenetic analyses by Edwards *et al.* (2010) utilised chloroplast *ndhF* sequences and morphology and showed strong monophyly of *Melaleuca sensu lato.* Species examined in the study were grouped into three main clades (A, B, C). However, the relationship among the A,B and C clades remained as polytomies and they could not find any diagnostic morphological character for each clade. Edwards *et al.* (2010) used samples from 10 informal species-groups in *Melaleuca sensu stricto* as well as taxa that represent all the genera in the tribe Melaleuceae. Clade A included samples of *M. leucadendra* and *M. acacioides* species-groups and the eight genera. Clade B included taxa from Australian *Callistemon*, *M. fulgens, M. laxiflora, M. ordinifolia, M. lanceolata* and *M cuticularis* species-groups. Clade C contained *M. foliolosa, M. huegelii group, M. lanceolata* and *M. cuticularis* species-groups. In the study, the following species-groups (*M. pungens, M. uncinata, M. cuticularis, M. ordinifolia, M. fulgens, M. laxiflora, M. lanceolata, M. huegelii* groups) were not monophyletic and species from *M. cuticularis* and *M. lanceolata* groups were found in both of the clades B and C (Edwards *et al*. 2010).

In eucalypts, a wide range of molecular markers have been utilised or developed for the group (Grattapaglia *et al.* 2012). Chloroplast/nuclear markers such as ITS, ETS, *psbA-trnH*, *trnL-trn*F and DArT (Diversity Arrays Technology) were useful to elucidate some of the generic, intrageneric/intergeneric boundaries or relationships (e.g. Gibbs *et al*. 2009; Steane *et al.* 2002; Steane *et al*. 2011). However, the relationship between the eucalypt genera *Angophora* and *Corymbia* still remains controversial and thus elucidating some of the generic/below generic level relationship in eucalypts still remain unresolved using molecular markers to date (Bayly 2016). Additionaly, RAPD (Random Amplified Polymorphism), AFLP (Amplified Fragment Length Polymorphism), SNP (Single Nucleotide Polymorphism) and SSR have also been used for phylogenetic estimations or they have been widely utilised to explore genetic variations at the population level (Grattapaglia *et al.* 2012).

Molecular Marker Identification

Development of Next-generation sequencing (NGS) has provided opportunities to collect large and high quality sequence data at a lower cost (Bayly 2016; Hörandl &

Appelhans 2015). With the new technology target sequencing allows enrichment of particular exons and/or regions of interest (Kozarewa *et al.* 2015) and it is now possible to target hundreds to thousands of loci (e.g. Blom *et al.* 2016; Budenhagen *et al.* 2016).

Identification of orthologous and low copy number genetic markers is important for phylogenetic studies (Valderrama *et al.* 2018). Commonly used methods for target loci identication includes target loci similarity approaches (e.g. reciprocal best hit, clustering methods) and phylogenetic approaches (e.g. Teasdale *et al.* 2016; Tekaia *et al*. 2016).

To date, phylogenetic studies of both *Melaleuca* and eucalypt groups were limited by the availability of limited marker sets and thus identification of molecular markers for the groups is crucial to elucidate the phylogenetic relationships. In *Melalueca sensu lato*, phylogenies at different taxonomic level (e.g. infraspecies level, species-group, species level) remain unresolved and the same problem can be found in eucalypts. Many more nuclear markers in eucalypts will help elucidate the species relationships in eucalypts (Bayly *et al.* 2016) and the same applies to the *Melaleuca* group.

Chapter outlines and aims

In Chapter 2, the problematic taxonomy of the *Melaleuca leucadendra* complex is investigated. Most of the species in the *M. leucadendra* species-group have broad leaves and papery bark, however, these characteristics can also be found in other species of *Melaleuca*. Instead of having autapomorphies, species in the complex are characterised by a combination of characters, and each of the character states occurs in more than one species of the complex (2008). Over the years, different species concepts have been applied to the group (Bentham 1868; Blake 1968; Byrnes 1986), and currently 16 species are recognised (Craven & Cowie 2013). Molecular data produced to date have not been sufficient to resolve the taxonomy of the *Melaleuca leucadendra* complex (Cook *et al*. 2008; Edwards *et al*. 2010). The described species are not reciprocally monophyletic using either cpDNA or nuclear loci (Cook *et al*. 2008; Edwards *et al*. 2013), and several species have considerable genetic variation across their distributions (Edwards *et al*. 2013). *Melaleuca argentea* is distributed across the tropics and subtropics of Australia (the northern areas of WA, NT, and QLD) and shows some regional clustering of cpDNA haplotypes and ITS alleles (Edwards *et al*. 2013). Of particular note in the study of Edwards *et al*. (2013) is the differentiation of populations in the Pilbara region of WA.

In the *psbA-rpl*2 (cpDNA) locus, the Pilbara populations all have a unique haplotype that is not found elsewhere in *M. argentea* (Edwards *et al*. 2013). In ITS (nrDNA), the Pilbara populations have a private allele, but share other alleles with populations outside the Pilbara to the north (Edwards *et al*. 2013).

Barlow (1988) recognised the Pilbara populations as an informal taxon, *M. argentea* "Ashburton biotype" but it was not described as being different from other populations of *M. argentea* and no formal taxonomic status was assigned to it. Based on their molecular data, Edwards *et al*. (2013) doubted the inclusion of the *M. argentea* "Ashburton biotype" in *M. argentea* and stated that further investigation was needed to determine its taxonomic status. In Chapter 2, we assessed both morphological and genetic diversity of *M. argentea* with a focus on the Pilbara populations.

In Chapter 3, we introduce a method used to identify orthologous loci with low copy number from both chloroplast and nuclear loci, with a focus on the target groups *Eucalyptus* and *Melaleuca* in Myrtaceae.

In Chapter 4, after excluding 13 loci with low sequence depth of coverage, using sequences of 47 samples across Myrtaceae, we assessed gene trees of 196 loci to investigate conflicts and whether those loci will potentially be useful for estimations of species to genus level phylogenies of *Melaleuca* and *Eucalyptus*, in addition to deeper relationships across Myrtaceae more broadly.

Lastly, Chapter 5 is a discussion and conclusion of all the chapters.

References

APC – Australian Plant Census (2012) *Australian Plant Census, IBIS database*. Centre for Plant Biodiversity Research, Council of Heads of Australian Herbaria. Published on the internet: http://www.anbg.gov.au/chah/apc (Accessed 6 August 2012).

ALA – Atlas of Living Australia (2017) http://spatial.ala.org.au/ (Accessed 10 December 2017).

Barlow, B.A. (1988) Patterns of differentiation in tropical species of *Melaleuca* L. (Myrtaceae). *Proceedings of the Ecological Society of Australia* 15: 239-247.

Bayly, M.J. (2016) Phylogenetic studies of eucalypts: fossils, morphology and genomes. *Proceedings of the Royal Society of Victoria* 128: 12-24.

Bayly, M.J., Rigault, P., Spokevicius, A., Ladiges, P.Y., Ades, P.K., Anderson, C., Bossinger, G., Merchant, A., Udovicic, F., Woodrow, I.E. & Tibbits, J. (2013) Chloroplast genome analysis of Australian eucalypts–*Eucalyptus, Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* 69: 704-716.

Beard, J.S., Chapman, A.R. & Gioia, P. (2000) Species richness and endemism in the Western Australian flora. *Journal of Biogeography* 27: 1257-1268.

Bentham, G. (1867) *Melaleuca*. in: *Flora australiensis*, vol.3, Myrtaceae to Compositae. London: Reeve.

Bentham, G. (1868) Notes on Myrtaceae. *The Journal of the Linnean Society of London, Botany* 10: 101-166.

Blake, S.T. (1968) A revision of *Melaleuca leucadendron* and its allies (Myrtaceae). *Contrributions from the Queensland Herbarium* 1: 1-114.

Blom, M.P., Bragg, J.G., Potter, S. & Moritz, C. (2016) Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Systematic biology* 66: 352-366.

Boland, D.J., Brooker, M.I.H., Chippendale, G.M., Hall, N., Hyland, B.P.M, Johnston, R.D., Kleinig, D.A. & Turner J.D. (1984) Forest trees of Australia. Thomas Nelson Australia, Melbourne.

Brown, G.K., Udovicic F. & Ladiges P.Y. (2001) Molecular phylogeny and biogeography of *Melaleuca, Callistemon* and related genera (Myrtaceae). *Australian Systematic Botany* 14: 565-585.

Budenhagen, C., Lemmon, A.R., Lemmon, E.M., Bruhl, J., Cappa, J., Clement, W.L., Donoghue, M., Edwards, E.J., Hipp, A.L., Kortyna, M. & Mitchell, N. (2016) Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. *bioRxiv* 086298.

Burbidge, A.A. (2016) A taxonomic revision of *Beaufortia* (Myrtaceae: Melaleuceae). *Nuytsia* 27: 165-202.

Byrnes, N.B. (1986) A revision of *Melaleuca* L. (Myrtaceae) in northern and eastern Australia, 3. *Austrobaileya* 2: 254-273.

Carnegie, A.J. & Lidbetter, J.R. (2012) Rapidly expanding host range for *Puccinia psidii sensu lato* in Australia. *Australasian Plant Pathology* 41: 13-29.

Cook, L.G., Morris, D.C., Edwards, R.D. & Crisp, M.D. (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* 47: 506-522.

Craven, L.A. (2006) New combinations in *Melaleuca* for Australian species of *Callistemon* (Myrtaceae). *Novon: A Journal for Botanical Nomenclature* 16: 468-475.

Craven, L.A. & Cowie, I.D. (2013) Taxonomic notes on the broad-leaved paperbarks (Myrtaceae, *Melaleuca*), including the description of one new species from northern Australia and a key to all taxa. *Blumea-Biodiversity, Evolution and Biogeography of Plants* 57: 207-209.

Craven, L.A., Edwards, R.D. & Cowley, K.J. (2014) New combinations and names in *Melaleuca* (Myrtaceae). *Taxon* 63: 663-670.

Craven, L.A. & Lepschi, B.J. (1999) Enumeration of the species and infraspecific taxa of *Melaleuca* (Myrtaceae) occurring in Australia and Tasmania. *Australian Systematic Botany* 12: 819-927.

Crisp, M.D., Cook, L.G. & Steane, D. (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 359: 1551-1571.

Crisp, M.D. & Cook, L.G. (2013) How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. *Annual Review of Ecology, Evolution, and Systematics* 44: 303-324.

Edwards, R.D., Craven, L.A., Crisp, M.D. & Cook, L.G. (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L. (Myrtaceae) is not monophyletic. *Taxon* 59: 744-754.

Franklin, D.C., Brocklehurst, P.S., Lynch, D. & Bowman, D.M. (2007) Niche differentiation and regeneration in the seasonally flooded *Melaleuca* forests of northern Australia. *Journal of Tropical Ecology* 23: 457-467.

Edwards, R.D., Crisp, M.D. & Cook, L.G. (2013) Niche differentiation and spatial partitioning in the evolution of two Australian monsoon tropical tee species. *Journal of Biogeography* 40: 559-569.

Gibbs, A.K., Udovicic, F., Drinnan, A.N. & Ladiges, P.Y. (2009) Phylogeny and classification of *Eucalyptus* subgenus *Eudesmia* (Myrtaceae) based on nuclear ribosomal DNA, chloroplast DNA and morphology. *Australian Systematic Botany* 22: 158-179.

Grattapaglia, D., Vaillancourt, R.E., Shepherd, M., Thumma, B.R., Foley, W., Külheim, C., Potts, B.M. & Myburg, A.A. (2012) Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics & Genomes* 8: 463-508.

Hill, K.D. & Johnson, L.A. (1995) Systematic studies in the eucalypts 7. A revision of the bloodwoods, genus Corymbia (Myrtaceae). *Telopea*: 185-504.

Hörandl, E. & Appelhans M.S. eds. (2015) Next generation sequencing in plant systematics. Koeltz Scientific Books.

Kershaw, A.P., Martin, H.A. & McEwen Mason, J.R.C. (1994) The Neogene: a period of transition. In: Hill, R.S. (Ed.), History of the Australian vegetation: Cretaceous to Recent. Cambridge Press, Cambridge, 299-327.

Ladiges, P.Y., McFadden, G.I., Middleton, N., Orlovich, D.A., Treloar, N. & Udovicic, F. (1999) Phylogeny of *Melaleuca, Callistemon*, and related genera of the *Beaufortia* suballiance (Myrtaceae) based on 5S and ITS-1 spacer regions of nrDNA. *Cladistics* 15: 151-172.

Ladiges, P.Y., Udovicic, F. & Drinnan, A.N. (1995) Eucalypt phylogeny—molecules and morphology. *Australian Systematic Botany* 8: 483-497.

Ladiges, P.Y., Udovicic, F. & Nelson, G. (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography* 30: 989-998.

Nevill, P.G., Williams, A., Krauss, S., Bradbury, D., Samaraweera, S. & Gardner, M.G. (2013) Development of microsatellite loci for the riparian tree species *Melaleuca argentea* (Myrtaceae) using 454 sequencing. *Applications in Plant Sciences* 1: 1200401.

Ochieng, J.W., Steane, D.A., Ladiges, P.Y., Baverstock, P.R., Henry, R.J. & Shepherd, M. (2007) Microsatellites retain phylogenetic signals across genera in eucalypts (Myrtaceae). *Genetics and Molecular Biology* 30: 1125-1134.

Kozarewa, I., Armisen, J., Gardner, A.F., Slatko, B.E. & Hendrickson, C.L. (2015) Overview of target enrichment strategies. *Current protocols in molecular biology* 7-21.

Paczkowska, G. & Chapman, A.R. (2000) The Western Australian flora: a descriptive catalogue. Wildflower Society of Western Australia, the West Austern Australian Herbarium, CALM and the Botanic Gardens & Parks Authority, Perth.

Parra-O, C., Bayly, M.J., Drinnan, A., Udovicic, F. & Ladiges, P. (2009) Corrigendum to: Phylogeny, major clades and infrageneric classification of *Corymbia* (Myrtaceae), based on nuclear ribosomal DNA and morphology. *Australian Systematic Botany* 22: 384-399.

Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E. & Potts, B.M. (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* 15: 49-62.

Teasdale, L.C., Koehler, F., Murray, K.D., O'Hara, T. & Moussalli, A. (2016) Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources* 16: 1107-1123.

Udovicic, F. & Spencer, R.D. (2012) New combinations in *Callistemon* (Myrtaceae). *Muelleria* 30: 23-25.

Wheeler, G.S. (2006) Chemotype variation of the weed *Melaleuca quinquenervia* influences the biomass and fecundity of the biological control agent *Oxyops vitiosa*. *Biological Control* 51: 689-702.

Wilson, P.G., O'Brien, M.M., Gadek, P.A. & Quinn, C.J. (2001) Myrtaceae revisited: a reassessment of infrafamilial groups. *American Journal of Botany* 88: 2013-2025.

Wilson, P.G., O'Brien, M.M., Heslewood, M.M. & Quinn, C.J. (2005) Relationships within Myrtaceae *sensu lato* based on *matK* phylogeny. *Plant Systematics and Evolution* 251: 3-19.

Tables

Table 1. Recent taxonomic changes in *Melaleuca sensu lato* (tribe Melaleuceae). Craven (2006) transferred *Callistemon* into *Melaleuca.* The rest of the genera in the tribe, which are marked with (*), were not treated in the study. In 2014, Craven *et al*. trasnferred all the seven genera into *Melaleuca*.

Wilson *et al.* (2005)	Craven (2006)	Craven *et al.* (2014)
Melaleuca sensu stricto	*Melaleuca*	*Melaleuca*
Callistemon		
Beaufortia	*Beaufortia**	
Conothamnus	*Conothamnus**	
Eremaea	*Eremaea**	
Lamarchea	*Lamarchea**	
Petraeomyrtus	*Petraeomyrtus**	
Phymatocarpus	*Phymatocarpus**	
Regelia	*Regelia**	

Figures.

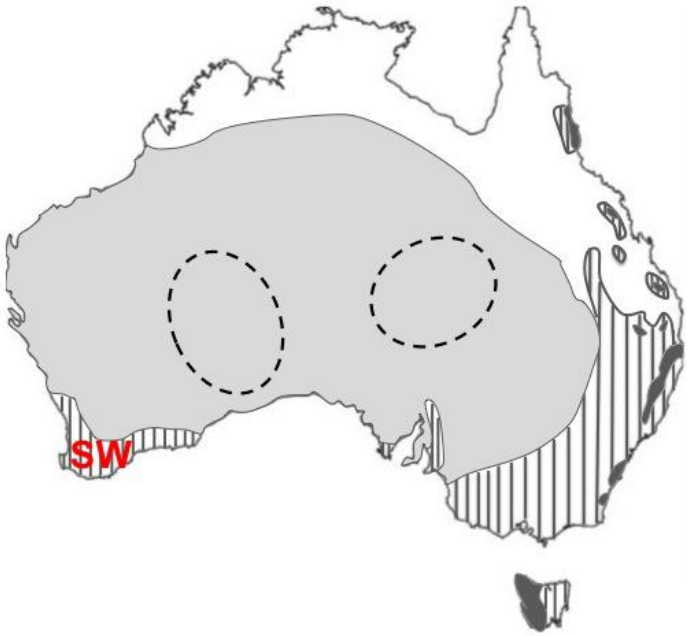

FIGURE 1. A simplified map of concurrent biomes of Australia (modified from Crisp & Cook 2013; Crisp *et al.* 2004; Kershaw *et al*. 1994). Different shades represent the following: white (monsoonal), light grey (arid), dark grey (everwet), vertical stripes (temperate sclerophyll). *Melaleuca* is distributed across Australia in all biomes but it is sparsely distributed to absent from some of the arid biomes including the regions highlighted in dashed circles (ALA 2017). The center of the diversity for *Melaleuca* is the Southwest Autralian Floristic Region which is marked as SW.

Chapter 2

Taxonomic status of *Melaleuca argentea* “ashburton biotype” of broad-leaf paperbarks (Myrtaceae, *Melaleuca leucadendra* species complex) from the Pilbara region of Western Australia

Abstract

Melaleuca leucadendra complex comprises ecologically significant tropical and subtropical trees that form a dominant component of the Australian monsoon tropics. Taxonomy of the *Melaleuca leucadendra* complex continues to be problematic. Most species in the complex cannot be defined by morphological autapomorphies and recent molecular analyses have also not been able to resolve the taxonomy of the group. Here we explore morphological and genetic diversity in *M. argentea* - a species distributed across northern Australia, with potentially isolated populations in the Pilbara region of WA. According to a recent study (Edwards *et al*. 2012), the populations of *M. argentea* from the Pilbara have a unique chloroplast haplotype in all samples and some private alleles of ITS. Furthermore, the niche space occupied by the Pilbara populations differs significantly from elsewhere across the range of *M. argentea*. Barlow (1988) recognised the Pilbara populations as *M. argentea* "Ashburton biotype" but did not provide any further information. We used morphological characters from *M. argentea* as well as five other species across the *M. leucadendra* complex to assess whether morphology can differentiate *M. argentea* "Ashburton biotype". We then assessed the two morphologically similar populations of *M. argentea* under a biological species concept, in order to determine its taxonomic status. Although *M. argentea* "Ashburton biotype" represented some distinctive genetic and morphological characters from the other taxa, there was not sufficient evidence to support its species distinctiveness from *M. leucadendra*. Both the chloroplast and nuclear haplotypes of *M. argentea* "Ashburton biotype" were nested within *M. leucadendra*. Although the Pilbara populations have been previously recognised as unusual for *M. argentea* "Ashburton biotype", we find that the populations are more similar to, and more closely related to, *M. leucadendra* than to *M. argentea*. The taxonomic status of *M. argentea* "Ashburton biotype" remains unresolved and we need a further test to decide whether there is gene flow between *M. leucadendra* and *M. argentea* "Ashburton biotype".

Introduction

Melaleuca L. (Myrtaceae) is an economically and ecologically important genus in Australia: a source of essential oil and ornamental plants (Cox *et al.* 2000; Brophy *et al.* 2013) and as the dominant shrubs and trees of wetlands (Franklin *et al.* 2007). One of the most readily recognised groups, the *M. leucadendra* (L.) L. complex (broad-leaved paperbarks), comprises ecologically significant tropical and subtropical trees that form a dominant component of the Australian monsoon tropics. They occupy riverine gallery forest, coastal wetlands and flood-plain habitats (Boland *et al.* 1984; Franklin *et al.* 2007), and are of high conservational importance (Barlow 1988). The group naturally extends beyond Australia to Indonesia, Malaysia, New Caledonia and New Guinea (Barlow 1988; Boland *et al.* 1984) and one species, *M. quinquenervia* (Cav.) S.T.Blake, has become an invasive weed in the Americas (Dray *et al.* 2006; Pratt *et al.* 2005; Serbesoff-King 2003). *Melaleuca leucadendra* complex contains the type species of the genus, *Melaleuca leucadendron* L. (L.) (Blake 1968).

The taxonomy of the broad-leaved paperbarks continues to be problematic: most species are not defined by autapomorphies (Cook *et al.* 2008). In other words, the majority of species in the complex are characterised by a combination of morphological characters with most of the character states occurring in more than one species of the complex. Different species concepts have been applied to the group at different times, further complicating the taxonomy. For example, Bentham (1868) recognised the group as a single polymorphic species but there are currently 16 species recognised (Blake 1968; Byrnes 1986; Craven & Barlow 1997; Brophy *et al.* 2013; Craven & Cowie 2013).

Recent molecular studies have also cast doubt on the current taxonomy of the broad-leaved paperbarks. The described species are not reciprocally monophyletic using either cpDNA (*psbA-rpl2*) or nrDNA (ITS) loci (Cook *et al.* 2008; Edwards *et al.* 2013), although this alone does not rule out species status (e.g., Avise *et al.* 1987), and several species have considerable genetic variation across their distributions to the extent that each might represent more than one species (Edwards *et al.* 2013). One of these species, *M. argentea*, has considerable genetic variation in cpDNA and nrDNA across northern Australia, some of which is geographically structured (Edwards *et al.* 2013). In particular, populations from the Pilbara region of Western Australia (Fig. 1, 3) were found to be fixed for a unique cpDNA haplotype, and to have some private alleles for

the nuclear ITS region (Edwards *et al.* 2013). Furthermore, the niche space occupied by specimens in the Pilbara was determined to be significantly different from elsewhere across the range of *M. argentea* (Edwards *et al.* 2013).

The Pilbara region of northwest Australia is a geologically old craton with very different soils from the sandy deserts of the surrounding regions (Wasson 1982; see Fig. 1). It is bounded to the north by the Great Sandy Desert, which separates it from the savannah biome of the Kimberley region of the Australian Monsoon Tropics (AMT). Previously unrecognised species have recently been discovered in the region (e.g., spiders, Harms & Framenau 2013; and plants, Maslin & Leeuwen 2008; Dillon 2014), and some taxa appear to have been long isolated from relatives nearby (e.g., geckos, Pepper *et al.* 2006).

In parts, the Pilbara is deeply cut by gorges with permanent water: these moister areas in an otherwise arid landscape are habitat for mesic-adapted taxa (Byrne *et al.* 2008). It is in these deep gorges and along intermittent riverbeds that *M. argentea* occurs, the only member of the broad-leaved paperbarks whose range in Western Australia extends south of the Great Sandy Desert (Fig. 1, 3). Current and planned mining in the Pilbara has the potential to reduce the water tables on which the populations of broad-leaf paperbarks rely, potentially threatening their long-term persistence in some areas (Nevill *et al.* 2013).

Melaleuca argentea was named and described by Fitzgerald (1918) from specimens from the Kimberley Region of Western Australia, whereas the first record of specimens of broad-leaf paperbarks from the Pilbara was identified as *M. leucadendra* (Blake 1968). Subsequently, Barlow (1988) diagnosed the Pilbara populations as *M. argentea* and called them *M. argentea* "Ashburton biotype", but provided no further information about morphological or ecological differences between them and typical *M. argentea*. Based on the molecular data and ecological tests of Edwards *et al.* (2013), further investigation is needed to determine biological and taxonomic status of the specimens from the Pilbara. Here we assess the species status of *M. argentea* "Ashburton biotype" using DNA sequences and morphological characters from specimens across the range of *M. argentea*, as well as from five other species of the *M. leucadendra* complex that occur in northwest Australia. We apply a biological species concept (Mayr 1942) in which gene flow between populations maintains a unit shaped by selection, and which

is differentiated from other entities (species) by low, or lack of, gene flow. Such units can be recognised using multiple different types of data (Cook *et al.* 2010). We tested if morphology can differentiate the Pilbara populations from the rest of the taxa.

Materials and methods

Sampling

Given the lack of autapomorphies and the non-monophyly of many recognised species of the *M. leucadendra* complex, testing species boundaries for even one species in the group requires broad sampling of co-distributed taxa in case morphology alone does not accurately reflect species membership. Seven species of the *M. leucadendra* species complex are known to occur in Western Australia: *M. argentea* (Fig. 1), *M. cajuputi* (Fig. 2), *M. leucadendra* (Fig. 2), *M. nervosa*, *M. sericea*, *M. viridiflora* and *M. lasiandra*. We sampled each of these (Table 1) using herbarium specimens identified by experts (Brendan Lepschi, Bryan Barlow) or by authors of this study, except *M. lasiandra*, which is a shrub rather than a tree and has a clear autapomorphy that separates it from all other members of the complex (hairy staminal filaments; Craven & Cowie 2013). There are several subspecies recognised for *M. cajuputi* but we used only *M. cajuputi* ssp. *cajuputi* because it occurs in Western Australia whereas the other two subspecies do not (Craven & Lepschi 1999; Holliday 2004).

Where possible, we tried to incorporate samples used for DNA sequencing in the scoring of morphological characters. However, this was frequently not possible because many specimens sampled for DNA had only vegetative vouchers and we needed to be able to score floral characters as well. Additional specimens were sourced from the Australian National Herbarium (CANB) and the Western Australia Herbarium (PERTH) to allow scoring of both floral and vegetation characters from the one specimen. Only five flowering specimens of the *M. argentea* "Ashburton biotype" were available: one from each of CANB and PERTH, and two collected by Vicki Long (VLPilb013-01, VLPilb013-02) and one by Leon Deschamps (Deschamps 1) specifically for this study (now housed at CANB, see Table 1). Abbreviations of herbaria are as in Index Herbariorum http://sciweb.nybg.org/science2/IndexHerbariorum.asp.

Molecular phylogenetics

New DNA for this study was extracted from dried leaf samples using the

CTAB/chloroform method. We amplified two regions (ITS, the internal transcribed spacers of the nuclear ribosomal cistrons, and the chloroplast region *psbA-rpl2*) to match those used by Edwards *et al*. (2013), and using the same primers and amplification conditions. Successfully amplified products were cleaned using Antarctic Phosphatase and Exonuclease I (New England Biolabs, Australia) before being sequenced by Macrogen (Republic of Korea) using the Sanger sequencing method. Sequence traces were edited in Geneious R7 (Kearse *et al.* 2012) and aligned using MAFFT (Katoh & Standley 2013) with the G-INS-1 progressive method algorithm (Katoh *et al.* 2005), then adjusted by eye where needed. Unambiguous indels were coded as present or absent.

Statistical parsimony networks (haplotype networks) of sequences from the four closely related taxa, *M. argentea* "Ashburton biotype", *M. argentea*, *M. leucadendra* and *M. cajuputi* from northwest Australia (WA, NT) were estimated using TCS (Clement *et al.* 2002) in PopART (Leigh & Bryant 2015). We excluded morphologically distinct taxa for the phylogenetic analyses. We used the option of gaps as missing data. In statistical parsimony networks, each circle/node indicates a different haplotype and how many nucleotides away each haplotype is from one another is presented (Mardulyn 2012). We used haplotype networks because we wanted to determine whether there is gene flow between *M. argentea* "Ashburton biotype" and the rest of four taxa.

Morphometrics

Morphometrics is a commonly used tool for finding and testing morphological disjunctions among taxa using quantitative data (Bateman 2001; Henderson 2006), and for finding characters to use in describing and circumscribing taxa (e.g., Chandler & Crisp 1998; Goldman *et al.* 2004). We measured 25 quantitative morphological characters on each specimen (Table 2), including leaf and flower characters. Where possible, all leaf measurements were made on the second most basal leaf on a latest season's branchlet (Fig. 4). We used the average of two measurements for most characters except that we measured an average of five oil gland diameters for each leaf and an average of three filaments per flower. We used a dissecting microscope (Wild Heerbrugg M7A) with an ocular micrometer to measure size characters of leaf hairs, petals, calyces, filaments, styles and oil glands. Larger structures, such as leaves and flowers, were measured using rulers. Hair density was measured on first year leaves (i.e. in the most recent seasonal growth).

We collated three datasets: "25LF_ALL" included all 25 leaf and flower characters for 39 specimens representing the six species listed above and *M. argentea* "Ashburton biotype"; "25LF_4SPP" comprised the same dataset but 21 specimens from four taxa: *M. argentea*, *M. argentea* "Ashburton biotype", *M. cajuputi* ssp. *cajuputi* and *M. leucadendra*; and "7LEAF" comprised seven leaf characters for 38 specimens of the same four taxa (Table 1). We collected seven leaf characters, "7LEAF", as a partial dataset, because in this way we could measure the characters from most specimens.

We used ordination and clustering methods as implemented in NTSYSpc v.2.20 (Rohlf 2008). Ordination summarises large multivariate datasets into a few dimensions under no assumption of hierarchical structure in the data (Pimentel 1981). In this study, we used non-metric multidimensional scaling (NMDS) because of its weaker assumption of a rank-order relationship between ordination space and the data matrix (Austin 1985). NMDS analyses of all datasets were carried out using options of standardisation (STAND: performs a linear transformation of a given data matrix), double centering (DCENTER) and eigen vector (EIGEN). Cluster analysis was performed using SIMINT to compute dissimilarity indices (distances) of continuous data, followed by UPGMA based on Euclidean distances in the SAHN module. We generated phenograms using the TREE module. The cophenetic correlation coefficient was computed using COPH and MXCOMP to test the goodness of fit of the cluster analysis to the similarity matrix.

In order to determine whether there were significant differences in measured traits between *M. argentea* "Ashburton biotype" and the morphologically similar broad-leaf paperbarks from Western Australia, we conducted t-tests for the morphological characters included in the 25LF_4SPP and 7LEAF datasets (Table 3).

Isolation-by-distance

Because populations of *M. argentea* "Ashburton biotype" occur at the southwestern end of the range of its nominal species, we wanted to know if it might appear to be different only because it is at the end of a morphological or genetic cline. We tested for genetic and morphological isolation-by-distance (IBD) within species and between species, including *M. argentea* "Ashburton biotype" as a taxon, and for morphological clines with latitude.

IBD was tested using linear regression in GraphPad Prism version 6 (GraphPad Software, San Diego California USA, www.graphpad.com) and its significance was assessed using randomisations implemented in the program (equivalent to a Mantel test). Genetic IBD of ITS and *psbA-rpl2* was assessed using the proportion of base-pair differences between sequences (P-distance) generated using Mega 5.2.2 (2011). Morphological IBD was assessed for the 25LF_4SPP and 7LEAF datasets using morphological distance (Euclidean distance) between taxa, which was computed using NTSYS. Pairwise geographic distances, expressed as kilometres (km), were calculated using the Geographic Distance Matrix Generator (2015, biodiversityinformatics.amnh.org/open_source/gdmg/). Latitudes were scored in decimal degrees (Table 1).

Results

Morphometrics

Melaleuca nervosa, *M. sericea* and *M. viridiflora* are clearly distinguished morphologically from the other species examined in this study (Figs. 5, 6). *Melaleuca sericea* is distinguished by having narrower leaves (<6.3 mm) and shorter petioles (<3.8 mm) than other species. *Melaleuca nervosa* and *M. viridiflora* have longer floral length (>17mm) than the other species. *Melaleuca viridiflora* and *M. nervosa* are distinguished by distance between open triad and calyx lobe length characters. *Melaleuca viridiflora* has longer distance between open triads (4.5-6.7 mm) than *M. nervosa* (2.5-3.8 mm). *Melaleuca viridiflora* has longer calyx lobe length (1.9-2.5 mm) than *M. nervosa* (0.7-1.3 mm). These three species, *M. sericea*, *M. nervosa* and *M. viridiflora*, also clustered distantly from *M. argentea* "Ashburton biotype" and the other broad-leaf paperbarks in the cluster analyses and NMDS of dataset 25LF_ALL (Figs. 5, 6). We then excluded these three differentiated species from further analyses.

Sixteen leaf and flower morphological characters were significantly different (P<0.05) between *M. argentea* "Ashburton biotype" and *M. argentea*, *M. leucadendra* and *M. cajuputi* (Table 3). *Melaleuca argentea* "Ashburton biotype" is significantly different from *M. argentea* in one leaf and six flower characters (ll5_1st, f1, f3, f5, f10, f20 and f21), from *M. cajuputi* in one leaf and three flower characters (ll13, f5, f8 and f21), and from *M. leucadendra* in six leaf and six flower characters (ll10, ll11, ll16, ll30, lr, lx, f1, f2, f10, f20, and f21; Table 3). Despite the differences, most of the mearement ranges of the *Melaleuca argentea* "Ashburton biotype" and the three taxa overlapped.

Melaleuca argentea "Ashburton biotype" is clustered near *M. cajuputi*, *M. argentea* and *M. leucadendra* in all ordination and cluster analyses (Figs. 7, 8). Although the species form discrete clusters in NMDS analyses of the 25LF_4SPP dataset (flower and leaf characters; Fig. 7), there is overlap of clusters in all NMDS analyses using leaf characters alone (7LEAF; Fig. 8). *Melaleuca cajuputi*, although not overlapping *M. argentea* "Ashburton biotype" in any analyses, was morphometrically close in all (Figs. 7, 8). *Melaleuca argentea* "Ashburton biotype" is distinguished from *Melaleuca argentea* by the first year hair density and from *M. cajputi* by leaf hair types on juvenile leaves (spreading in *M. cajputi* – specimens examined: Deschams 1, VLPilb013-01, VLPilb013-02; appressed in *M. argentea* "Ashburton biotype" – specimens examined: Craven 6455, Craven 4536). In cluster analysis, sample L5 was distantly clustered from the rest of *M. leucadendra* samples. This sample represented an extreme end of the leaf size range (long and wide leaf size). Therefore, the sample was excluded in further analyses.

Morphological IBD was detected only in *M. cajuputi* (25LF_4SPP and 7LEAF datasets, $P < 0.01$; Table 4), for which there were only three samples, and not within *M. argentea, M. leucadendra* or *M. argentea* "Ashburton biotype". There was significant IBD when *M. argentea* "Ashburton biotype" was analysed with *M. argentea* and *M. leucadendra* but not when analysed with *M. cajuputi* (Table 4).

There was no correlation between leaf morphology and latitude within *M. argentea* "Ashburton biotype", *M. cajuputi* or *M. argentea*, but there was a significant correlation in *M. leucadendra* ($P = 0.0254$; Table 4).

DNA sequences

Twelve haplotypes of *psbA-rpl2* were recovered from the four species (Fig. 9). All 16 samples of *M. argentea* "Ashburton biotype" shared a unique haplotype whereas there were multiple haplotypes recovered from each of the other three species: seven in *M. cajuputi*, six in *M. argentea* and five in *M. leucadendra*. Four of the haplotypes were shared among species (Fig. 9). A haplotype from *M. cajuputi* specimen H35 (haplotype 2) was only one step (substitution) from the *M. argentea* "Ashburton biotype" haplotype (Fig. 9).

Forty-one alleles of ITS were amplified from the four species, only two of which were shared across species (Fig. 10). The majority of specimens of *M. argentea* "Ashburton biotype" shared a unique allele (haplotype 1, Fig. 10). Other ITS alleles from *M. argentea* "Ashburton biotype" either clustered nearby (within four steps), one of which (haplotype 2) was shared with *M. argentea* specimen H26, or clustered separately in the allelic network (Fig. 10). Alleles from *M. cajuputi* specimen H35 were distant in the allelic network from those from *M. argentea* "Ashburton biotype". Alleles of none of species fell into a single cluster and there is no rooting of the network that would result in monophyly for any species. ITS alleles of other species that are a couple of substitutions away from those of *M. argentea* "Ashburton biotype" are from specimens geographically far away, and two or more substitutions away in *psbA-rpl2* (e.g., H26 and RDE307).

There was significant isolation-by-distance in ITS for each of the four taxa (Table 5): negatively correlated in *M. argentea* "Ashburton biotype" and positively correlated in the others, but there was no IBD in comparisons using *psbA-rpl2* (Table 5).

Discussion

The populations previously recognised as *Melaleuca argentea* "Ashburton biotype", from the Pilbara region of northwest Western Australia (Fig. 3), represent some morphological and genetic characters within the *M. leucadendra* species complex. Nevertheless, no one piece of evidence provides definite evidence for species distinctiveness on its own.

We extended sampling of *M. argentea* "Ashburton biotype" over that which Edwards *et al.* (2013) used, five samples in chloroplast and three in nuclear, did not recover any additional cpDNA haplotypes despite much greater coverage of the geographic range of the taxon, confirming a diagnostic chloroplast *psbA-rpl2* haplotype common to all specimens. The lack of genetic variability in cpDNA might indicate that the Pilbara populations have not been abundant or isolated for a long period of time. There is diversity in ITS though, with alleles falling in two separate clusters in the network (Fig. 10): a result that adds two more alleles to the previous findings for *M. argentea* "Ashburton biotype" (Edwards *et al.* 2013) and our results are consistent with the lack of monophyly found for other species of broad-leaved paperbarks (Cook *et al.* 2008; Edwards *et al.* 2013). In general, chloroplast and nuclear markers provide maternal and

bipaternal evolutionary histories, respectively, as well as different mutation rates (in general, lower in the chloroplast; Gaudeul *et al.* 2011). In the chloroplast sequences, no genetic differences was found in the *M. argentea* "Ashburton biotype" while several different haplotyps as well as a shared haplotype were found in the nuclear haplotype network. Conflicting phylogenetic patterns of the chloroplast and nuclear marker could potentially explain pollen and seed dispersal and if hybridisation has occurred then disagreeing patterns among nuclear markers could be found (Cook *et al.* 2008).

The ITS alleles of *M. argentea* "Ashburton biotype" of in both cpDNA and nuclear are nested among those derived from *M. leucadendra* (Fig. 10). However, the lack of sharing of cpDNA haplotypes and ITS alleles between *M. argentea* "Ashburton biotype" and nearby populations of other species of the *M. leucadendra* complex, despite non-monophyly, indicates that there has been no recent gene flow. If *M. argentea* "Ashburton biotype" were genetically distinct only because it represented a bottlenecked population at the end of a cline, we would expect to find the most closely related (most similar) haplotypes and alleles in adjacent populations. However, this is not the case. Even though *M. cajuputi* (H35; haplotype 2) from one of the closest geographic localities was only one substitution different at the cpDNA locus (Fig. 9), it was not close in the allelic network for ITS (Fig. 10). When ITS alleles were shared, it was between specimens that are geographically distant. This indicates that sharing of ITS alleles is more likely due to incomplete lineage sorting than recent gene flow, as suggested by Cook *et al*. (2008): if there were gene flow it should be between closely located populations – those that are close enough for the dispersal of pollen by insects or vertebrates such as fruit bats.

This study is the most extensive morphometric analysis of the *M. leucadendra* complex to date. *Melaleuca sericea*, *M. viridiflora* and *M. nervosa* were readily distinguished from the rest examined in this study in the morphological analyses (Figs. 5, 6; also see Table 1). Although we found morphological characters that separate *M. argentea* "Ashburton biotype" from *M. argentea* and *M. cajuputi* by hair density and hair type respectively *M. argentea* "Ashburton biotype" was not so easily differentiated from *M. leucadendra*. Although twelve leaf size and flower characters were significantly different ($P<0.05$) between *Melaleuca argentea* "Ashburton biotype" and *M. leucadendra* (Table 3), there is mostly an overlap (or nearly so) in the range of those measurements such that no characters are diagnostic for the two taxa. Because the number of the *Melaleuca argentea*

"Ashburton biotype" specimens that were measured in this study was based on five specimens only, further specimens need to be included. Importantly, however, we found that the significant differences in morphological characters such as leaf length and hair density between *M. argentea* "Ashburton biotype" and the other species are not because it represents the end of a morphological cline of one of the other species. Either there is no cline in the other species (*M. argentea* and *M. leucadendra*) or there is no significant IBD when *M. argentea* "Ashburton biotype" is included in the same analysis (*M. cajuputi*). Leaf morphology (7LEAF) is not correlated with latitude except in *M. leucadendra*.

Melaleuca argentea "Ashburton biotype" specimens have been assigned as either *M. argentea* or *M. leucadendra*. *Melaleuca argentea* "Ashburton biotype" exhibits almost the same range of leaf measurements and characteristics as *M. argentea*, except leaf hairiness. This probably explains why Barlow (1988) assigned the Pilbara populations of broad-leaved paperbark to this species. On the other hand, the low density of leaf hairs in *M. argentea* "Ashburton biotype" might explain why Blake (1968) originally assigned the Pilbara specimens to *M. leucadendra* rather than to *M. argentea* because the former has low leaf hair densities while the latter has high densities (giving the leaves the characteristic silver appearance which is the basis of its species epithet).

How and when *M. argentea* "Ashburton biotype" came to be distributed through the Pilbara is currently an enigma. Despite the long-term occurrence of the Great Sandy Desert, which is a region of high aridity with very little habitat suitable for riverine or gallery forest trees, the Great Sandy Desert does not separate the Pilbara populations from those in the Kimberley to the north (Fig. 1, 3). There is no corresponding deep divergence between *M. argentea* "Ashburton biotype" and other broad-leaved paperbarks. The lack of deep divergence might indicate that *M. argentea* "Ashburton biotype" dispersed into the region only relatively recently. It is only long enough ago to have diverged by a single nucleotide change in the *psbA-rpl2* region sequenced. On the other hand, diversity in ITS is likely a result of ancestral polymorphism and indicates that the colonising population was not completely bottlenecked or that there was some ongoing gene flow via pollen after colonisation.

Our results do not provide definitive evidence that *M. argentea* "Ashburton biotype" is a distinct species from *M. leucadendra*, which was the main question in this chapter. In

order to answer this question, it needs further testing with more nuclear sequences to assess whether there has been gene flow. Regardless of the taxonomic status of *M. argentea* "Ashburton biotype", some of its genetic distinctions could be taken into consideration for the conservation of the taxa when mining in the area is taken into account. In chapter 3, we identified marker sets for phylogenetic studies of *Melaleuca* and *Eucalyptus* that could prove useful. For the future dirction, to answer the question, sequencing some of the most variable nuclear loci for *M. leucadendra* group out of the marker sets will be useful.

References

Austin, M.P. (1985) Continuum concept, ordination methods, and niche theory. *Annual Review of Ecology and Systematics* 16: 39-61.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. & Saunders, N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18: 489-522.

Barlow, B.A. (1988) Patterns of differentiation in tropical species of *Melaleuca* L. (Myrtaceae). *Proceedings of the Ecological Society of Australia* 15: 239-247.

Bateman, R.M. (2001) Evolution and classification of European orchids: insights from molecular and morphological characters. *Journal Europäischer Orchideen* 33: 33-119.

Bentham, G. (1868) Notes on Myrtaceae. *The Journal of the Linnean Society. Botany*. 10: 101-166.

Blake, S.T. (1968) A revision of *Melaleuca leucadendron* and its allies (Myrtaceae). *Contr. Queensland Herbarium* 1: 1-114.

Boland, D.J., Brooker, M.I.H., Chippendale, G.M., Hall, N., Hyland, B.P.M, Johnston, R.D., Kleinig, D.A. & Turner J.D. (1984) Forest trees of Australia. Thomas Nelson Australia, Melbourne.

Brophy, J.J., Craven, L.A. & Doran, J.C. (2013) Melaleucas: their botany, essential oils and uses. Australian Centre for International Agriculture Research, Canberra.

Byrne, M., Yeates, D.K., Joseph, L., Kearney, M., Bowler, J., Williams, M.A.J., Cooper, S., Donnellan, S.C., Keogh, J.S., Leys, R., Melville, J., Murphy, D.J., Porch, N. & Wyrwoll, K.H. (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology* 17: 4398-4417.

Byrnes, N.B. (1986) A revision of *Melaleuca* L. (Myrtaceae) in northern and eastern Australia, 3. *Austrobaileya* 2: 254-273.

Chandler, G.T. & Crisp, M.D. (1998) Morphometric and phylogenetic analysis of the *Daviesia ulicifolia* complex (Fabaceae, Mirbelieae). *Plant Systematics and Evolution* 209: 93-122.

Clement, M., Snell, Q., Walke, P., Posada, D. & Crandall, K. (2002) TCS: estimating gene genealogies. *Proceedings 16th International Parallel and Distribution Processing Symposium* 2: 184.

Cook, L.G., Morris, D.C., Edwards, R.D. & Crisp, M.D. (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* 47: 506-522.

Cook, L.G., Edwards, R.D., Crisp, M.D. & Hardy, N.B. (2010) Need morphology always be required for new species descriptions? *Invertebrate Systematics* 24: 322-326.

Cox, S.D., Mann, C.M., Markham, J.L., Bell, H.C., Gustafson J.E., Warmington J.R. & Wylie, S.G. (2000) The mode of antimicrobial action of the essential oil of *Melaleuca alternifolia* (tea tree oil). *Journal of applied Microbiology* 88: 170-175.

Craven, L.A. & Barlow, B.A. (1997) New taxa and new combinations in *Melaleuca* (Myrtaceae). *Novon* 7: 113-119.

Craven, L.A. & Cowie, I.D. (2013) Taxonomic notes on the broad-leaved paperbarks (Myrtaceae, *Melaleuca*), including the description of one new species from northern Australia and a key to all taxa. *Blumea* 3: 207-209.

Craven, L.A. & Lepschi, B.J. (1999) Enumeration of the species and infraspecific taxa of *Melaleuca* (Myrtaceae) occurring in Australia and Tasmania. *Australian Systematic Botany* 12: 819-927.

Dillon, S.J. (2014) *Grevillea saxicola* (Proteaceae), a new species from the Pilbara of Western Auistralia. *Nuytsia* 24: 103-108.

Donnelly M.J., Girod R., Bsansky N.J. & Lehmann T. (2004) Revisiting the role of introgression vs shared ancestral polymorphism as key processes shaping genetic diversity in the recently separated sibling species of *Anopheles gambiae* complex. *Heridity* 92: 61-68.

Dray, F.A. Jr, Benett, B.C. & Center, T.D. (2006) Invasion history of *Melaleuca quinquenervia* (Cav.) S.T. Blake in Florida. *Castanea* 71: 210-225.

Dugard, P., Todman, J. & Staines, H. (2010) Approaching multivariate analysis: a practical introduction. Routledge, Sussex.

Edwards, R.D., Craven, L.A., Crisp, M.D. & Cook, L.G. (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L. (Myrtaceae) is not monophyletic. *Taxon* 59: 744-754.

Edwards, R.D., Crisp, M.D. & Cook, L.G. (2013) Niche differentiation and spatial partitioning in the evolution of two Australian monsoon tropical tree species. *Journal of Biogeography* 40: 559-569.

Ersts, P.J. (2015) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg [Accessed 14 Jun. 2015].

Fitzgerald, W.V. (1918) The botany of the Kimberleys, north-west Australia. *Journal and Proceedings of the Royal Society of Western Australia* 3: 102-224.

Franklin, D.C., Brocklehurst, P.S., Lynch, D. & Bowman, D.M. (2007) Niche differentiation and regeneration in the seasonally flooded *Melaleuca* forests of northern Australia. *Journal of Tropical Ecology* 23: 457-467.

Gaudeul, M., Giraud, T., Kiss, L. & Shykoff, J.A. (2011) Nuclear and chloroplast microsatellites show multiple introductions in the worldwide invasion history of common ragweed, *Ambrosia artemisiifolia*. *PloS one* 6: e17658.

Goldman, D.H., van den Berg, C. & Griffith, M.P. (2004) Morphometric circumscription of species and infraspecific taxa in *Calopogon* R.Br. (Orchidaceae). *Plant Systematics and Evolution* 247: 37-60.

Harms, D. & Framenau, V.W. (2013) New species of mouse spiders (Araneae: Mygalomorphae: Actinopodidae: Missulena) from the Pilbara region, Western Australia. *Zootaxa* 3637: 521-540.

Henderson, A. (2006) Traditional morphometrics in plant systematics and its role in palm systematics. *Botanical Journal of the Linnean Society* 151: 103-111.

Holliday, I. (2004) *Melaleucas: a field and garden guide*. Reed New Holland Publishers, Australia.

Katoh, K., Kuma, K.I., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511-518.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organisation and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Leigh, J.W. & Bryant, D. (2015) Popart: full‐feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110-1116.

Mayr, E. (1942) Systematics and the origin of species. Columbia University Press, New York.

Mardulyn, P. (2012) Trees and/or networks to display intraspecific DNA sequence variation? *Molecular ecology* 21: 3385-3390.

Maslin, B.R. & Leeuwen, S. (2008) New taxa of *Acacia* (Leguminosae: Mimosoideae) and notes on other species from the Pilbara and adjacent desert regions of Western Australia. *Nuytsia* 18: 139-188.

Nevill, P.G., Williams, A., Krauss, S., Bradbury, D., Samaraweera, S. & Gardner, M.G. (2013) Development of microsatellite loci for the riparian tree species *Melaleuca argentea* (Myrtaceae) using 454 sequencing. *Applications in Plant Sciences* 1: 1200401.

Pepper, M., Doughty, P. & Keogh, J.S. (2006) Molecular phylogeny and phylogeography of the Australian *Diplodactylus stenodactylus* (Gekkota; Reptilia) species-group based on mitochondrial and nuclear genes reveals an ancient split between Pilbara and non-Pilbara *D. stenodactylus*. *Molecular Phylogenetics and Evolution* 41: 539–555.

Pimentel, R.A. (1981) A comparative study of data and ordination techniques based on a hybrid swarm of sand verbenas (*Abronia* Juss.). *Systematic Biology* 30: 250–267.

Pratt, P.D., Rayamajhi, M.B., Van, T.K., Center, T.D. & Tipping, P.W. (2005) Herbivory alters resource allocation and compensation in the invasive tree *Melaleuca quinquenervia*. *Ecological Entomology* 30: 316–326.

Rohlf, F.J. (2008) NTSYSpc, Numeric Taxonomy System and Multivariate Analysis System. Version 2.2.

Serbesoff-King, K. (2003) *Melaleuca* in Florida: A literature review on taxonomy, distribution, biology, ecology, economic importance and control measures. *Journal of Aquatic Plant Management* 41: 98-112.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28: 2731-2739.

The New York Botanic Garden. (2014) *Index Herbariorum*. Available at: http://sciweb.nybg.org/science2/IndexHerbariorum.asp [Accessed 11 Aug. 2014].

Wasson, R.J. (1982) Landform development in Australia. In: Evolution of the Flora and Fauna of Arid Australia (eds Barker WR, Greenslade PJM), pp. 23–33. Peacock Publications, Adelaide.

Tables

Table 1. Information on the samples used in the morphological and molecular studies.

Sample name	Taxon name	Morphological data		Molecular data		Herbarium	GPS coordinates	
		LF25	7LEAF	*psbA-rpl2*	ITS		Latitude	Longitude
A1_W (Lazarides 6526)	*Melaleuca argentea*	✓	✓			CANB	-18.18	125.58
A2_W (Speck 5017)	*M. argentea*	✓	✓			CANB	-15.96	127.45
A3_W (Speck 4903)	*M. argentea*	✓	✓			CANB	-14.31	126.63
A4_Q (Story 8044)	*M. argentea*	✓	✓			CANB	-18.21	141.41
A5_N (Blake 16695)	*M. argentea*	✓	✓			CANB	-13.23	131.11
A6_N (Speck 1662)	*M. argentea*	✓	✓			CANB	-14.7	132.05
AR1 (H32; Morse 1769)	*M. argentea*		✓	✓	✓	CANB	-15.93	127.21
AR2 (H31; Mitchell 3721)	*M. argentea*		✓		✓	CANB	-15.9166	127.9972
AR3 (H79; Must 1095)	*M. argentea*		✓	✓		CANB	-16.204784	136.882782
arg_GB4 (Brown 4)	*M. argentea*			✓	✓	-	-15.752981	127.74765
arg_Gbarg	*M. argentea*			✓	✓	-	-12.900691	130.592337
arg_H21 (Puttock 14429)	*M. argentea*			✓		CANB	-15.48	135.4
arg_H23 (Lea 844)	*M. argentea*			✓	✓	CANB	-14.3	126.63
arg_H76 (Thomson 828)	*M. argentea*			✓		CANB	-17.16	137.76
arg_LGC02210	*M. argentea*			✓		-	-18.105833	125.701944
arg_RDE268	*M. argentea*				✓	-	-12.582764	132.874843
arg_RDE270	*M. argentea*			✓	✓	BRI	-12.582764	132.874843
arg_RDE304	*M. argentea*			✓	✓	-	-15.678948	126.404132
arg_Sgherza17	*M. argentea*			✓	✓	PERTH	-16.604	122.788
bil_H102_a1 (Newbey 10602)	*M. argentea* "Ashburton biotype"			✓	✓	CANB	-22.4666	116.6166
bil_LGC01886	*M. argentea* "Ashburton biotype"			✓	✓	-	-21.570471	117.054616
bil_LGC01888	*M. argentea* "Ashburton biotype"			✓		-	-22.477627	118.558126
bil_LGC01890	*M. argentea* "Ashburton biotype"			✓	✓	-	-22.022741	115.598390

bil_LGC01891	*M. argentea* "Ashburton biotype"			✓	✓	-	-24.704871	115.316110
bil_LGC01892	*M. argentea* "Ashburton biotype"			✓	✓	-	-22.544480	115.497828
bil_LGC01893	*M. argentea* "Ashburton biotype"			✓		-	-24.859121	113.688269
H1_W (Young 146)	*M. argentea* "Ashburton biotype"	✓	✓			CANB	-20.3333	119.25
H2_W (Deschamps 1)	*M. argentea* "Ashburton biotype"	✓	✓	✓	✓	CANB	-21.7775	114.9822
H3_W (VLPilb013-01)	*M. argentea* "Ashburton biotype"	✓	✓	✓	✓	CANB	-21.578	117.0872
H4_W (George 3398)	*M. argentea* "Ashburton biotype"	✓	✓			PERTH	-20.516	118.066
HB1 (H29; Morse 1843)	*M. argentea* "Ashburton biotype"		✓	✓	✓	CANB	-21.6	118.86
HB11 (H28; Mitchell 1496)	*M. argentea* "Ashburton biotype"		✓	✓	✓	PERTH	-21.722	116.213
HB14 (VLPilb013-02)	*M. argentea* "Ashburton biotype"		✓	✓	✓	CANB	-20.4577	118.8169
HB15 (Smith88)	*M. argentea* "Ashburton biotype"		✓	✓		PERTH	-25.05	115.217
HB2 (H100; Kalotas 1319)	*M. argentea* "Ashburton biotype"		✓	✓	✓	CANB	-20.53	119.08
HB3 (Gardner s.n.)	*M. argentea* "Ashburton biotype"		✓			PERTH	-25.155	116.941
HB4 (Morse 1949)	*M. argentea* "Ashburton biotype"		✓			CANB	-21.6	118.86
HB5 (Carr 4560)	*M. argentea* "Ashburton biotype"		✓			PERTH	-20.667	119.233
HB6 (Duero15)	*M. argentea* "Ashburton biotype"		✓	✓	✓	PERTH	-19.768	121.402
HB7 (H25; Mitchell1444)	*M. argentea* "Ashburton biotype"		✓	✓	✓	PERTH	-21.864	121.007
HB8 (H9; Newbey 10622)	*M. argentea* "Ashburton biotype"		✓			CANB	-22.4666	116.6166
C1_W (H35; Keneally 6849)	*M. cajuputi*	✓	✓	✓	✓	CANB, PERTH	-16.5666	122.7833
C2_N (Craven 4536)	*M. cajuputi*	✓	✓			CANB	-12.61	132.3
C3_N (Blake 17511)	*M. cajuputi*	✓	✓			CANB	-14.9333	133.06660
caj_Gbcaj	*M. cajuputi*				✓		-12.5598	132.3048333
caj_H113 (Cowie 6520)	*M. cajuputi*				✓	CANB	-11.9261	136.5458
caj_H114 (Lea783)	*M. cajuputi*				✓	CANB	-14.93	133.13
caj_H47 (Reeve s.n)	*M. cajuputi*			✓		-	-12.104139	134.884644
caj_LGC00768	*M. cajuputi*				✓	-	-12.73777778	132.7663889
caj_RDE262	*M. cajuputi*			✓		-	-12.6715	131.371833

caj_RDE265	*M. cajuputi*			✓		-	-12.736	132.7598333
L1_Q (Paijmans 1904)	*M. leucadendra*	✓	✓			CANB	-18.75	146.2
L2_N (Lazarides 322)	*M. leucadendra*	✓	✓			CANB	-12.75	132.6333
L3_N (Cardale s.n.)	*M. leucadendra*	✓	✓			CANB	-16.78	135.75
L4_N (Mitchell 8606)	*M. leucadendra*	✓	✓			CANB	-14.2341	135.6033
L5_N (Craven 4537)	*M. leucadendra*	✓				CANB	-12.61	132.3
L6_Q (Muller s.n.)	*M. leucadendra*		✓			CANB	-16.8903	141.0594
L7_W (Kenneally 10717)	*M. leucadendra*	✓	✓			PERTH, CANB	-15.61	128.28
L8_W (Kenneally 2110)	*M. leucadendra*	✓	✓			PERTH, CANB	-15.467	125.483
LE10 (H26; Evans 2972)	*M. leucadendra*		✓	✓	✓	CANB	-14.95	132.15
LE12 (Kenneally 10710)	*M. leucadendra*		✓			PERTH	-14.897	128.661
LE14 (Crisp 10515)	*M. leucadendra*		✓			CANB	-12.4463	130.88
LE6 (Bailey s.n.)	*M. leucadendra*		✓			CANB	-12.61	132.86
LE8 (O' Keefe s.n.)	*M. leucadendra*		✓			CANB	-18.7	138.48
LE9 (Mitchell 2779)	*M. leucadendra*		✓	✓		CANB	-16.9794	123.9508
leu_H118 (Purdie 4768)	*M. leucadendra*			✓	✓	CANB	-16.6583	125.9288
leu_H15 (Lea 865)	*M. leucadendra*			✓	✓	CANB	-15.65	130.46
leu_Kenneally9665	*M. leucadendra*			✓		PERTH	-16.033333	124.766667
leu_LGC00264	*M. leucadendra*			✓	✓	-	-17.424564	124.972916
leu_LGC00265	*M. leucadendra*			✓		-	-17.424564	124.972916
leu_LGC00765	*M. leucadendra*			✓	✓	-	-14.18111	132.193611
leu_LGC02209	*M. leucadendra*			✓	✓	-	-17.412222	124.942778
leu_LGC02211	*M. leucadendra*			✓	✓	-	-18.105833	125.701944
leu_RDE269	*M. leucadendra*				✓	-	-12.582764	132.874843
leu_RDE291	*M. leucadendra*				✓	-	-16.024	130.7721667
leu_RDE303	*M. leucadendra*			✓		BRI	-15.678948	126.404132
leu_RDE307	*M. leucadendra*				✓	-	-14.671013	125.734642
leu_RDE308	*M. leucadendra*				✓	-	-14.671013	125.734642
N1_W (Lazarides 6401)	*M. nervosa*	✓				CANB	-17.25	127.05

N2_W (Lazarides 6560)	*M. nervosa*	✓	CANB	-16.93	122.53
N3_W (Lazarides 6564)	*M. nervosa*	✓	CANB	-17.1333	122.6
N4_W (Wilson 887)	*M. nervosa*	✓	CANB	-19.685	121.2561
N5_N (Perry s.n.)	*M. nervosa*	✓	CANB	-17.5833	133.5833
N6_Q (Perry 3566)	*M. nervosa*	✓	CANB	-20.5333	145.3666
N7_Q (Lazarides 6878)	*M. nervosa*	✓	CANB	-22.7666	149.8833
S1_W (Lazarides 5139)	*M. sericea*	✓	CANB	-17.2	126.4
S2_W (Fryxell 3989)	*M. sericea*	✓	CANB	-15.96	126.8
S3_W (Forbes 2273)	*M. sericea*	✓	CANB	-15.91	127.3
S4_W (Aplin 4996)	*M. sericea*	✓	CANB	-17.2166	126.2666
S5_N (Craven 9148)	*M. sericea*	✓	CANB	-15.61	130.36
S6_N (Perry 2683)	*M. sericea*	✓	CANB	-15.71	129.88
V1_N (Symon 6930)	*M. viridiflora*	✓	CANB	-18.66	130.16
V2_Q (Crome 696)	*M. viridiflora*	✓	CANB	-17	145.5
V3_W (Beard 7043)	*M. viridiflora*	✓	PERTH	-15.762	128.735
V4_W (Kenneally 7644)	*M. viridiflora*	✓	CANB	-16.9666	123.2333
V5_W (Fryxwell 4604)	*M. viridiflora*	✓	CANB	-16.11	123.76

Table 2. A list of leaf and flower characters used in this study for 25LF_All and 7LEAF datasets. All the character measurements are in millimeters (mm) unless otherwise stated. We made two measurements for all the characters except that we measured five oil glands from each leaf and an average of three filaments per flower. Characters included in the 7LEAF dataset are marked as an asterisk (*).

	Code	Character
Leaf	ll5_1st*	Hair density on a mature 1st year leaf (number of hairs/0.026 mm^2)
	ll10*	Length of leaf
	ll11*	Width of leaf
	ll13*	Length of petiole
	ll15_ctr	Number of primary veins at the central part of a leaf
	ll16	Oil gland diameter
	ll17	Oil gland density (number of glands/0.026 mm^2)
	ll30*	Length between midrib tip and axil axis
	ll34	Leaf thickness (length between adaxial leaf surface to abaxial leaf surface)
	lr*	Curved leaf length excluding petiole
	lx*	Length between widest point of a leaf to the leaf tip
Flower	f1	Length between the base of a style and the tip of a stigma
	f2	Length between a tip of an anther to the base of a filament after dissection of a flower
	f3	Number of filaments per bundle (total number of stamens/bundle number)
	f4	Length between a tip to the base of a petal
	f5	Width at the widest point of a petal
	f7	Length between calyx tip and base
	f8	Width at the widest point of a calyx lobe
	f10	Width at the widest point of a of flowering inflorescence
	f11	Length of filament (between a tip of an anther to the base of a filament)
	f13	Distance between open triads
	f14	Number of flowers per one inflorescence
	f20	Flower length between the tip of an anther to the base of floral cup
	f21	Hair density of a floral tube
	f22	Width of a rachis base

Table 3. Averaged (± SD standard deviation; where applicable) floral and leaf character measurements of *Melaleuca argentea* "Ashburton biotype", *M. argentea* (arg), *M. leucadendra* (leu) and *M. cajuputi* (caj) in 25LF_ALL and 7LEAF dataset. T-test results of all the characterss between *M. argentea* "Ashburton biotype" (AB) and *M. argentea* (arg), *M. leucadendra* (leu) and *M. cajuputi* (caj).

Code	Averaged morphological measurements of each character (± SD; where applicable)				T-test of two taxa		
	AB	*M. argentea*	*M. leucadendra*	*M. cajuputi*	AB-arg	AB-leu	AB-caj
f1	10.35 (0.3851)	13.36 (1.224)	13.29 (2.119)	8.683 (1.674)	**0.0016**	**0.0228**	0.104
f10	21.88 (1.652)	28.42 (1.348)	28.23 (3.805)	21.83 (5.923)	**0.0001**	**0.0106**	0.9895
f11	70.6875 (18.7943)	65.0417 (9.9755)	85.375 (22.8254)	65.0833 (37.4185)	0.5482	0.2948	0.8023
f13	4.792 (0.5023)	5.3889 (1.3225)	6.4892 (1.7487)	3.6778 (0.8829)	0.4206	0.0922	0.085
f14*	51.75 (n=4)	42.75 (n=6)	36 (n=8)	39 (n=3)	0.1381	0.0727	0.1429
f2	7.9819 (0.9868)	10.0204 (1.7542)	10.8496 (2.1905)	8.7222 (2.425)	0.0703	**0.0342**	0.5961
f20	10.61 (1.098)	15.42 (0.9958)	14.46 (2.083)	10.12 (1.881)	**<0.0001**	**0.0067**	0.6794
f21*	5.75 (n=4)	13.75 (n=6)	0.25 (n=8)	11 (n=3)	**0.0095**	**0.0101**	**0.0286**
f22	1.4 (0.1633)	1.4375 (0.2349)	1.4125 (0.3044)	1.55 (0.3905)	0.79	0.9412	0.5107
f3*	7.55 (0.4123)	6.765 (0.1344)	8.056 (1.124)	9.304 (1.637)	0.0095	0.1778	0.2
f4	2.9063 (0.2221)	3.1517 (0.4135)	3.4625 (0.714)	2.2667 (0.6506)	0.3143	0.1671	0.1193
f5	2.656 (0.2904)	3.113 (0.2479)	3.072 (0.3337)	1.883 (0.3403)	0.0283	0.0608	**0.0227**
f7	1.6281 (0.4905)	1.3208 (0.1913)	1.7795 (0.5547)	1.5359 (0.2355)	0.1946	0.6548	0.7792
f8	1.956 (0.2085)	1.843 (0.1677)	2.303 (0.3382)	1.267 (0.2517)	0.3665	0.0931	**0.0105**
ll10	88.225 (20.105)	93.1389 (8.7911)	138.726 (27.2984)	89.49 (22.10)	0.4992	**<0.0001**	0.924
ll11	12.489 (3.042)	14.2611 (3.1025)	18.876 (6.868)	14.88 (1.087)	0.1905	**0.004**	0.208
ll13	9.428 (1.713)	9.6944 (1.1165)	10.919 (1.900)	5.75 (0.75)	0.6853	**0.042**	**0.0028**
ll15_ctr*	8.25 (n=4)	7 (n=6)	7.5 (n=8)	8 (n=3)	0.1048	0.4909	>0.9999
ll16	0.006432 (0.001127)	0.007084 (0.0005536)	0.007708 (0.0007508)	0.006429 (0.0004348)	0.2515	**0.0396**	0.9971
ll17*	1.5 (n=4)	1.75 (n=6)	1 (n=8)	1.5 (n=3)	0.7143	>0.9999	>0.9999
ll30	5.912 (2.827)	4.8833 (2.3301)	11.842 (5.970)	2.792 (1.573)	0.3735	**0.0026**	0.0887
ll34	0.3875 (0.052)	0.4292 (0.0679)	0.4031 (0.0674)	0.375 (0.0433)	0.3311	0.6948	0.7506
ll5_1st*	1.153 (1.525)	13.8333 (6.6568)	1.00 (1.202)	1 (n=3)	**<0.0001**	0.8099	0.9471
lr	89.367 (20.323)	94.1222 (8.7513)	141.603 (28.645)	90.25 (22.11)	0.5169	**<0.0001**	0.9471

lx	46.728 (14.063)	47.0167 (4.9622)	84.711 (22.535)	48.95 (11.93)	0.9537	**<0.0001**	0.8037

Level of confidence $P<0.05$ for the T-test result is in bold.

Legend for morphological characters of dataset 25LF_4SPP (measurements are in mm unless otherwise stated):

ll5_1st* (hair density on a mature 1st year leaf; number of hairs/0.026 mm^2, ll10* (length of leaf), ll11* (width of leaf), ll13* (length of petiole), ll15_ctr (number of primary veins at the central part of a leaf), ll16 (oil gland diameter), ll17 (oil gland density; number of glands/0.026 mm^2), ll30* (length between midrib tip and axil axis), ll34 (leaf thickness: length between adaxial leaf surface to abaxial leaf surface), lr* (curved leaf length excluding petiole length), lx* (length between widest point of a leaf to the leaf tip), f1 (length between the base of a style and the tip of a stigma), f2 (length between a tip of an anther to the base of a filament after dissection of a flower), f3 (number of filaments per bundle; total number of stamens/bundle number), f4 (length between a tip to the base of a petal), f5 (width at the widest point of a petal), f7 (length between calyx tip and base), f8 (width at the widest point of a calyx lobe), f10 (width at the widest point of a of flowering inflorescence), f11 (length of filament: between a tip of an anther to the base of a filament), f13 (distance between open triads), f14 (number of flowers per one inflorescence), f20 (flower length between the tip of an anther to the base of floral cup), f21 (hair density of a floral tube), f22 (width of a rachis base). 7LEAF datasets include seven leaf characters that are marked with an asterisk (*).

Table 4. Mantel test between morphological distances (Euclidean distance of 25LF_All and 7LEAF datasets) and geographic distance of *Melaleuca argentea* "Ashburton biotype (AB), *M. argentea*, *M. cajuputi* and *M. leucadendra*. Mantel test between morphological distance and a latitude matrix of all the above mentioned *Melaleuca* are included (*P*: significance test using random permutation). 25LF_All dataset includes 25 leaf and flower characters as in the footnote (all the measurements are in mm unless otherwise stated).

	25LF_All		7LEAF		Latitude vs. 7LEAF	
Population	R^2	*P*	R^2	*P*	R^2	*P*
AB & *M. argentea*	0.2438	0.0006	0.008827	0.1044	0.007435	0.6819
AB & *M. cajuputi*	0.07642	0.2251	0.00351	0.4932	0.001085	0.9001
AB & *M. leucadendra*	0.0834	0.0187	0.04154	0.0004	0.5152	<0.0001
M. argentea	0.04555	0.445	0.009454	0.4801	0.03571	0.5779
AB	0.01076	0.845	0.004649	0.5207	0.04026	0.4916
M. cajuputi	0.9997	0.0103	0.9982	0.0271	0.6836	0.3803
M. leucadendra	0.004882	0.7329	0.002041	0.7433	0.4431	0.0254

Legend for morphological characters:
ll5_1st* (hair density on a mature 1st year leaf; number of hairs/0.026 mm^2, ll10* (length of leaf), ll11* (width of leaf), ll13* (length of petiole), ll15_ctr (number of primary veins at the central part of a leaf), ll16 (oil gland diameter), ll17 (oil gland density; number of glands/0.026 mm^2), ll30* (length between midrib tip and axil axis), ll34 (leaf thickness: length between adaxial leaf surface to abaxial leaf surface), lr* (curved leaf length excluding petiole length), lx* (length between widest point of a leaf to the leaf tip), f1 (length between the base of a style and the tip of a stigma), f2 (length between a tip of an anther to the base of a filament after dissection of a flower), f3 (number of filaments per bundle; total number of stamens/bundle number), f4 (length between a tip to the base of a petal), f5 (width at the widest point of a petal), f7 (length between calyx tip and base), f8 (width at the widest point of a calyx lobe), f10 (width at the widest point of a of flowering inflorescence), f11 (length of filament: between a tip of an anther to the base of a filament), f13 (distance between open triads), f14 (number of flowers per one inflorescence), f20 (flower length between the tip of an anther to the base of floral cup), f21 (hair density of a floral tube), f22 (width of a rachis base). 7LEAF datasets include seven leaf characters that are marked with an asterisk (*).

Table 5. Mantel test results for genetic distance (p-distance: the proportion of different sites) of the nuclear (ITS) and chloroplast (*psbA-rpl2*) sequences and geographic distance matrices (*P*: significance test using random permutation) of four *Melaleuca* taxa: *M. argentea* "Ashburton biotype", *M. argentea*, *M. cajuputi, M. leucadendra.* "na" refers to not available.

	ITS		*psbA-rpl2*	
Taxa	R^2	*P*	R^2	*P*
AB & *M. argentea*	0.3029	<0.0001	0.2488	<0.0001
AB & *M. cajuputi*	0.6197	<0.0001	0.5876	<0.0001
AB & *M. leucadendra*	0.1872	<0.0001	0.1289	<0.0001
M. argentea	0.06904	<0.0001	0.01084	0.2905
AB	0.01686	0.0149	na	na
M. cajuputi	0.06268	0.0167	0.04731	0.3436
M. leucadendra	0.127	<0.0001	0.04264	0.0346

Figures.

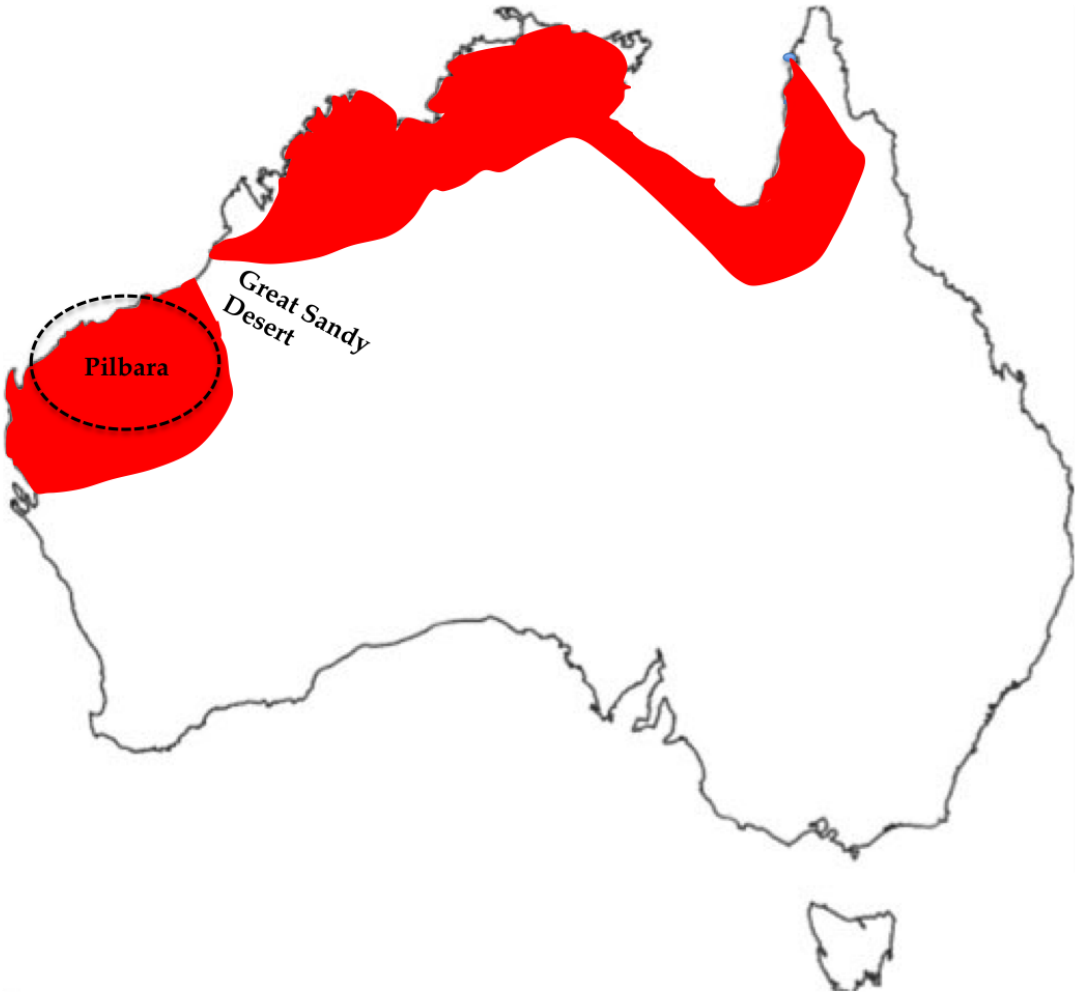

FIGURE 1. A distribution map of *Melaleuca argentea* in Australia (shaded in red).

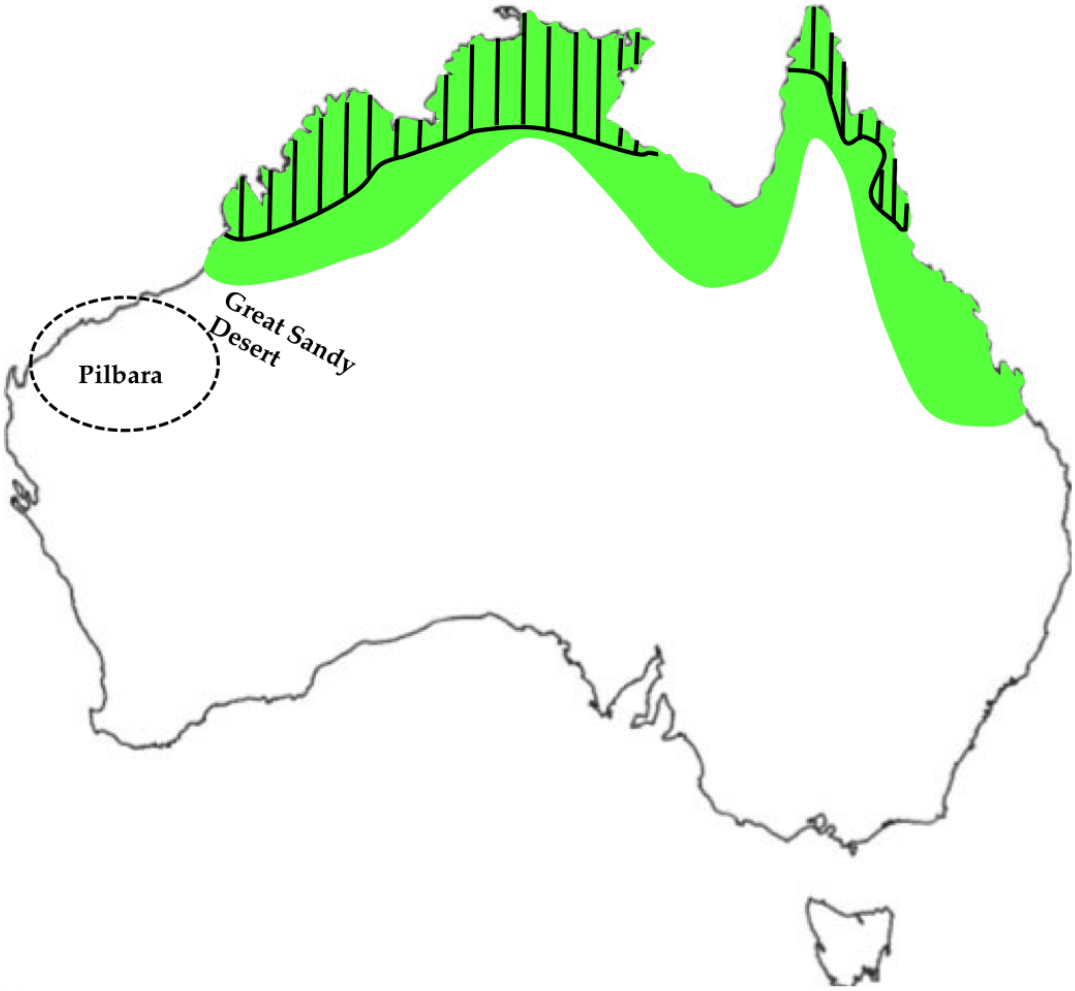

FIGURE 2. A distribution map of *Melaleuca leucadendra* (shaded in green) and *M. cajuputi* (shaded with vertical stripes).

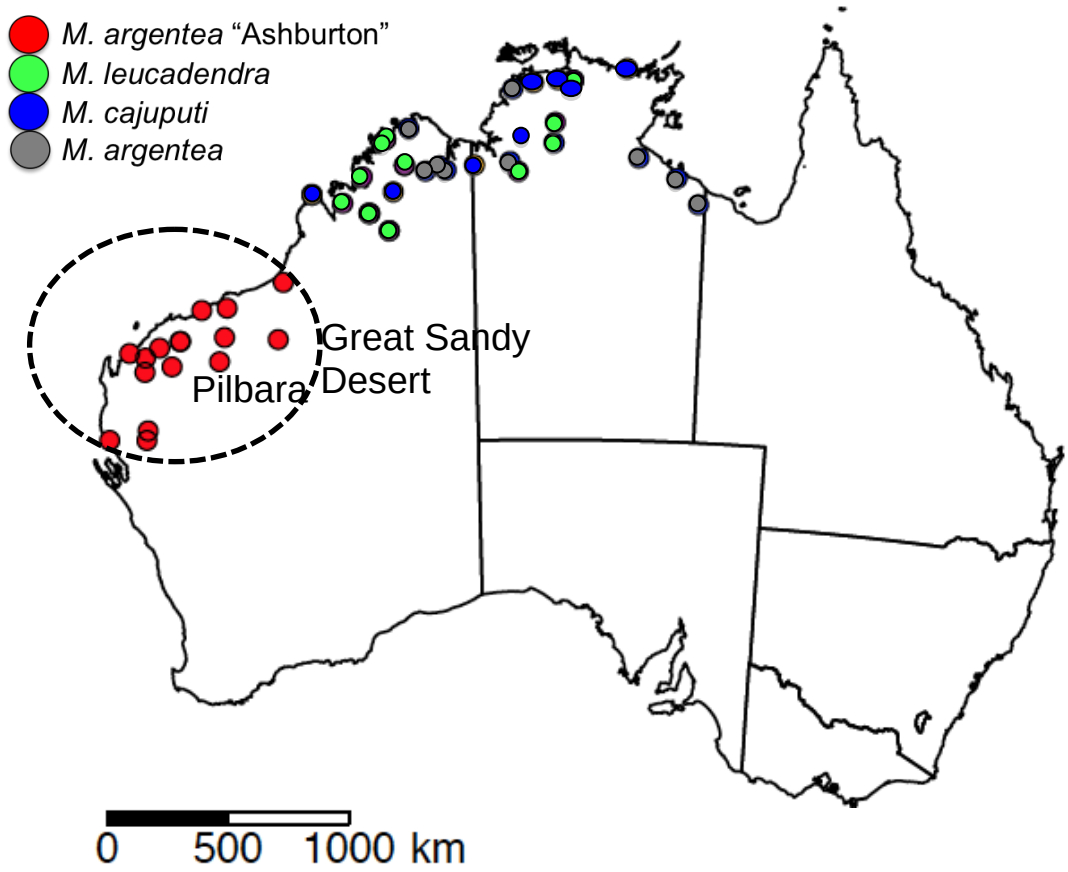

FIGURE 3. A map of Australia with collection localities for DNA samples (nrDNA ITS and cpDNA *psbA-rpl2*) of *Melaleuca argentea* (grey), *M. argentea* "Ashburton biotype" (red), *M. cajuputi* (blue) and *M. leucadendra* (green). Circles indicate collection sites.

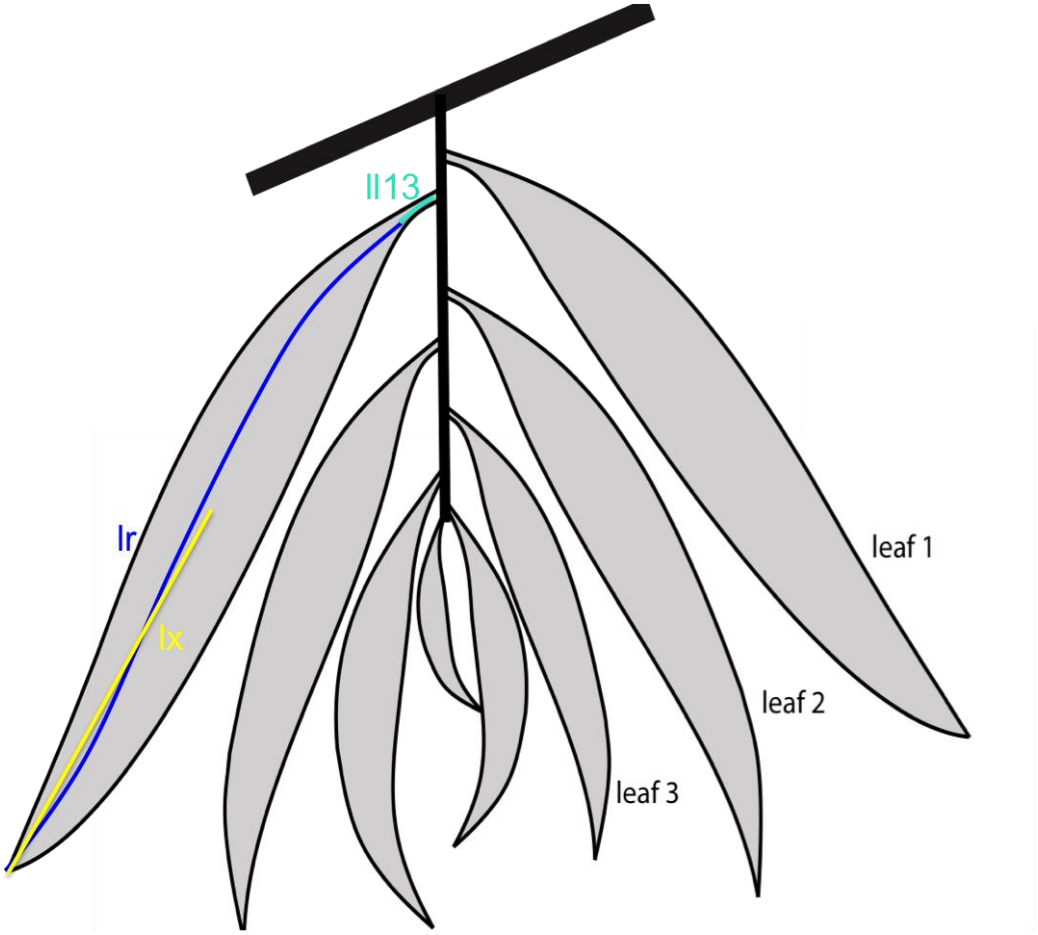

FIGURE 4. An illustration of three leaf morphological measurements (ll13, lr, lx). ll13: length of petiole; lr: curved leaf length excluding petiole length; lx: length between widest point of a leaf to the tip.

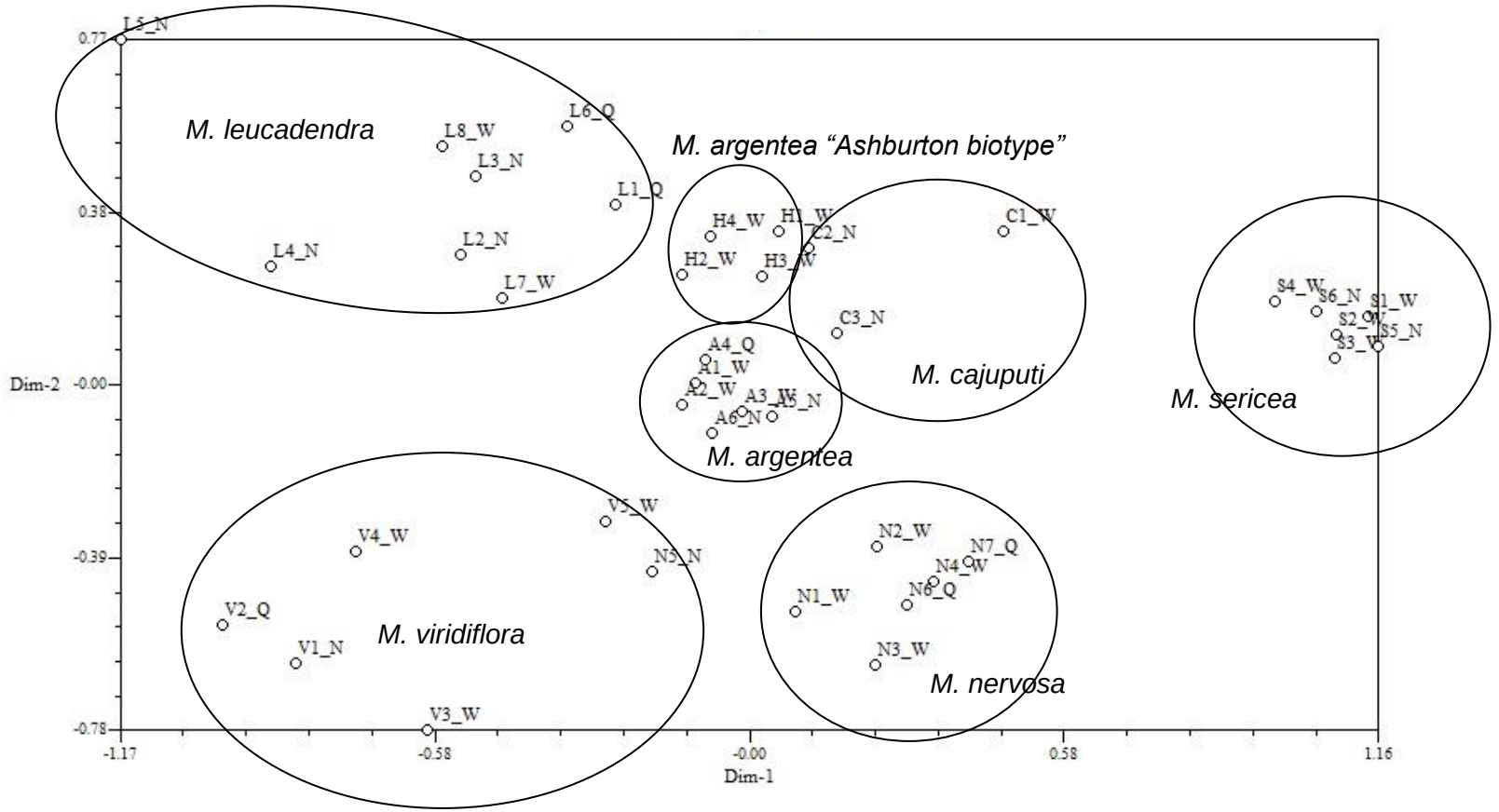

FIGURE 5. Non-metric multidimensional scaling (NMDS) analysis of dataset 25LF_ALL, which is consisted of 25 flower and leaf morphological character measurements of seven *Melaleuca* species as below. The first two digits of the specimen code represent species as follows. H1-H4: *M. argentea* "Ashburton biotype"; A1-A6: *M. argentea*. C1-C3: *M. cajuputi*; L1-L8: *M. leucadendra*; N1-N7: *M. nervosa*; V1-V5: *M. viridiflora*; S1-S6: *M. sericea*. The last digit indicate the states as follows: N: Northern Territory; Q: Queensland; W: Western Australia. Ellipses indicate boundaries of *M. leucadendra, M. argentea, M. argentea* "Ashburton biotype samples and *M. cajuputi* respectively.

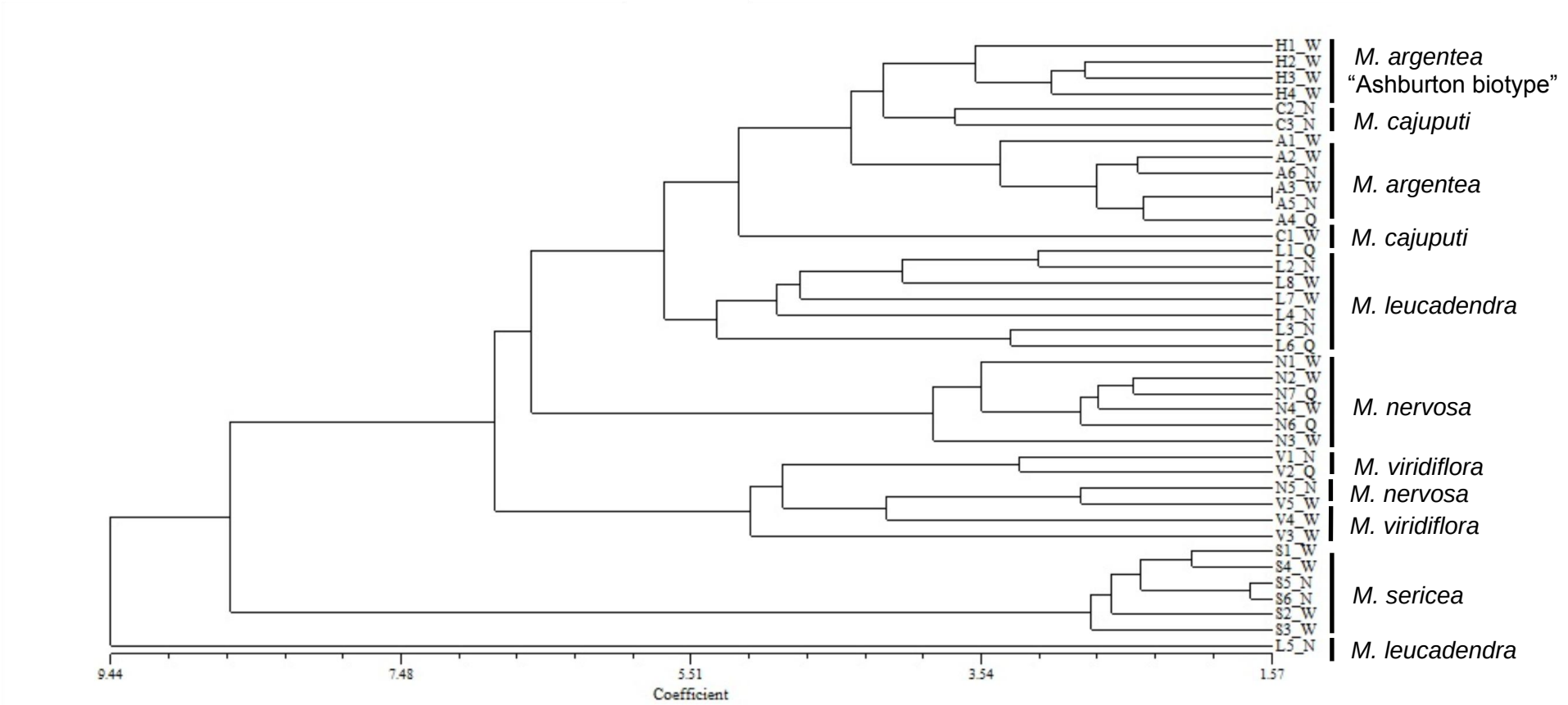

FIGURE 6. Cluster analysis of the leaf and morphological dataset of 25LF_ALL. The morphological characters measured in this study are stated as below. Seven taxa in *Melaleuca leucadendra* complex were used in the study and the first two digits of the specimen code represent species as follows. H1-H4: *M. argentea* "Ashburton biotype"; A1-A6: *M. argentea*. C1-C3: *M. cajuputi*; L1-L8: *M. leucadendra*; N1-N7: *M. nervosa*; V1-V5: *M. viridiflora*; S1-S6: *M. sericea*. The last digit indicate the states as follows: N: Northern Territory; Q: Queensland; W: Western Australia.

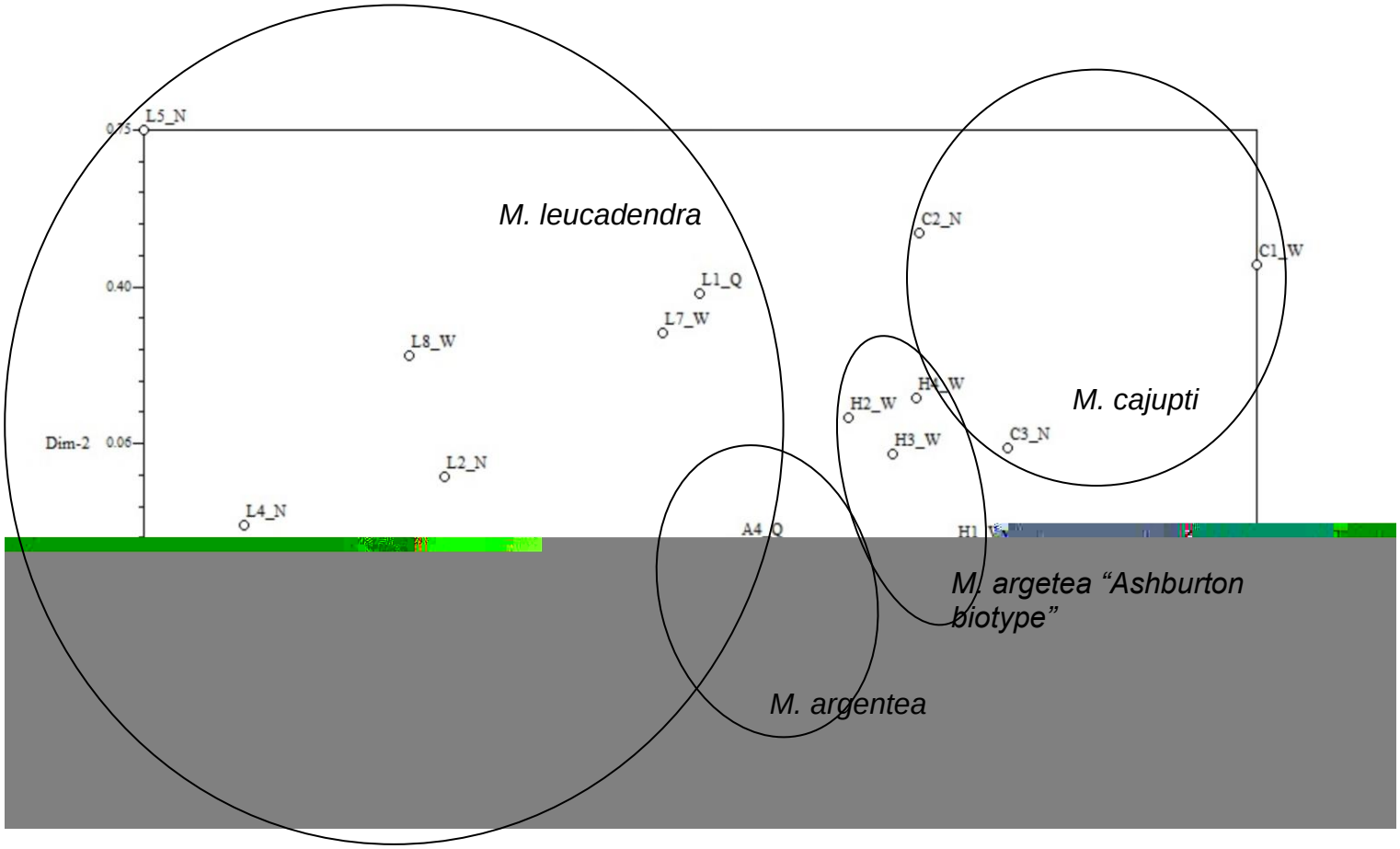

FIGURE 7. NMDS analysis of dataset 25LF_4SPP, which is consisted of 25 flower and leaf morphological characters in four taxa as below. See the legend for the characters used for the dataset. The first two digits of the specimen code represent species as follows. H1-H4: *M. argentea* "Ashburton biotype"; A1-A6: *M. argentea*. C1-C3: *M. cajuputi*; L1-L8: *M. leucadendra*. The last digit indicates the states as follows: N: Northern Territory; Q: Queensland; W: Western Australia. Ellipses indicate boundaries of *M. leucadendra, M. argentea, M. argentea* "Ashburton biotype samples and *M. cajuputi* respectively.

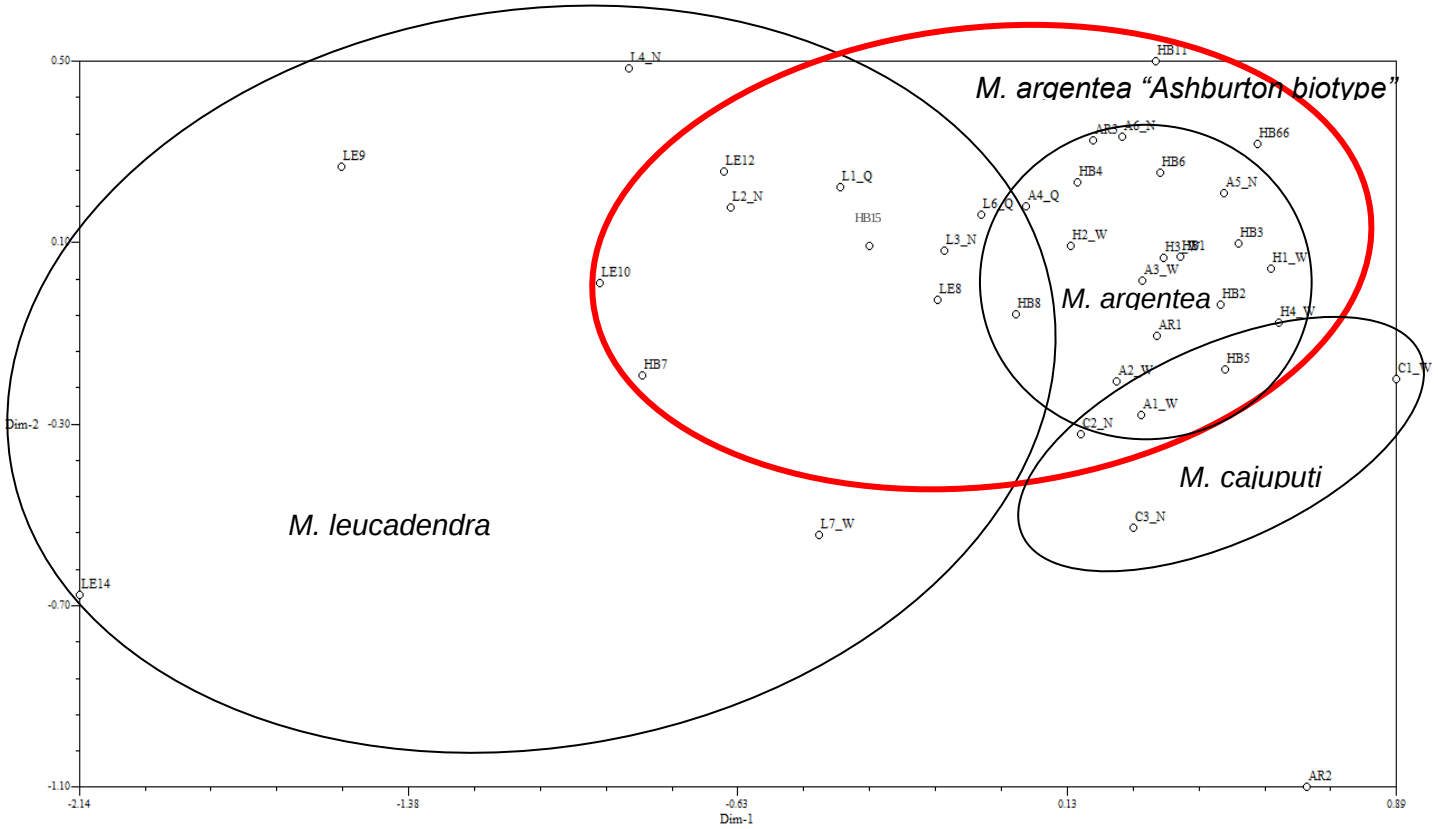

FIGURE 8. Non-metric multidimensional scaling (NMDS) of dataset 7LEAF, which is comprised of leaf characters of the four taxa from *M. leucadendra* complex. The first digit of the specimen code represents the following species. H: *M. argentea* "Ashburton biotype"; A: *M. argentea*; C: *M. cajuputi*; L: *M. leucadendra*. Different ellipses represent the boundary of samples in the analyses. The thick ellipse represents the boundary of *M. argentea* "Ashburton biotype samples and thin ellipses indicate boundaries of *M. leucadendra, M. argentea* and *M. cajuputi* smaples respecitvely.

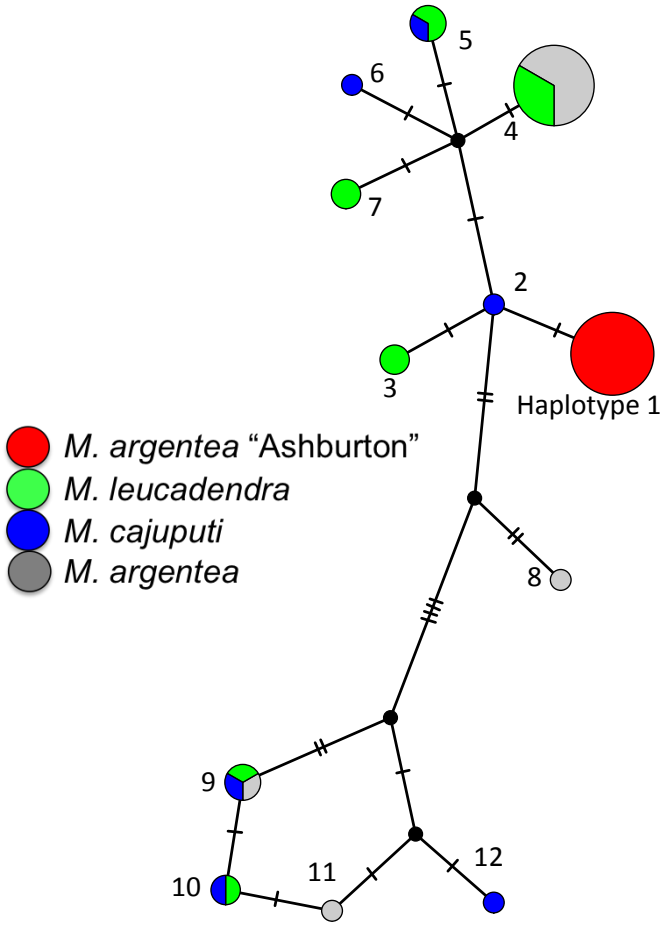

FIGURE 9. A parsimony haplotype network for chloroplast sequences (*psbA-rpl2*). Different colours indicate different taxa as follows: *Melaleuca argentea* (grey), *M. argentea* "Ashburton biotype" (red), *M. cajuputi* (blue) and *M. leucadendra* (green). Each coloured circles represent haplotypes that were present in this study except the smallest for black circles (unsampled haplotypes). The size of each circle is proportional to the number of samples used.

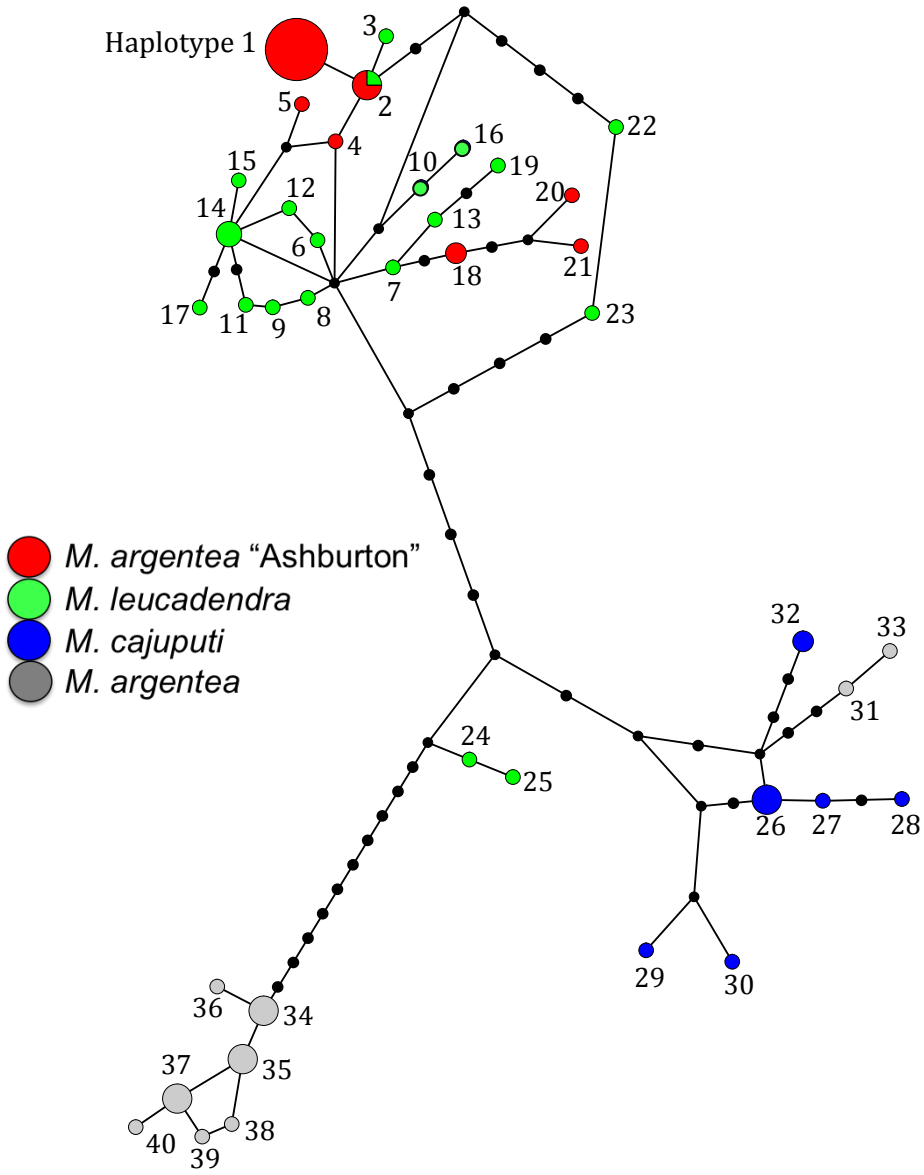

FIGURE 10. A parsimony haplotype network for nuclear sequences (ITS). Different colours indicate different taxa as follows: *Melaleuca argentea* (grey), *M. argentea* "Ashburton biotype" (red), *M. cajuputi* (blue) and *M. leucadendra* (green). Each circle represents different haplotype. Coloured haplotypes were present in this study and black dots indicate unsampled haplotypes. The size of each circle is proportional to the number of samples used.

Chapter 3

Identification of genetic markers for genus to species level phylogenies of *Melaleuca* and other genera in Myrtaceae

Abstract

It is challenging to resolve the phylogenetic relationships of closely related species using a small set of loci, often because of incomplete lineage sorting and/or ongoing hybridisation. In *Eucalyptus* and *Melaleuca* (Myrtaceae), most diversification occurred in the last 20 million years and resolving the species phylogenies has been problematic. Thus, we aimed to identify multiple chloroplast and nuclear loci to estimate the phylogenies. We introduce methods used for the genetic marker identification for species to generic level phylogenies. Resources used for the identification include annotated genome, transcriptomes, and whole genome shotgun sequences of *Melaleuca* and *Eucalyptus*. To identify nuclear loci to target, we first selected reference contigs from *Melaleuca* transcriptomes. Second, we assembled our contigs against the whole genome *Melaleuca* shotgun sequences to select loci that are unique to the genome using average coverage from the assembly. Third, we blasted the selected contigs against the *E. grandis* genome to obtain matching coding DNA sequences and to compare copy number of the matching loci in *E. grandis* and other species in the plant kingdom to filter out potential high copy loci. Finally, using a gene tree estimation approach we selected loci that meets the taxonomic priors and removed the rest. For the chloroplast loci, we aligned three genomes and selected loci by number of single nucleotide polymorphisms and indels. In total, we identified 263 kbp of 212 loci. We verified 209 loci by successfully sequencing all the loci in 47 samples of five genera in Myrtaceae. However, a lower average coverage for the nuclear loci of *Heteropyxis*, which is an outgroup of *Melaleuca* and *Eucalyptus,* was achieved. The average percentage of reads mapped back to the *E. grandis* targets was 57.6%. 83.5% of the data matrix had an average coverage of more than 20 reads. We identified 209 loci that are potentially useful for phylogenies in *Melaleuca* as well as other genera, including *Eucalyptus*, in Myrtaceae. Furthermore, this chapter provides a useful method that is applicable to plants and other groups of organisms with unresolved phylogenies.

Introduction

For many years, nuclear, mitochondrial and/or chloroplast loci have been used to estimate phylogenetic relationships among taxa. In plants, chloroplast and nuclear genes are often employed. Chloroplast genomic regions have been widely used to estimate phylogenetic relationships because of their simple and stable genetic structure (Dong *et al*. 2012). Chloroplast loci are transmitted uniparentally and evolve more conservatively than nuclear genes (Gielly & Taberlet 1994; Dong *et al*. 2012; Akhani *et al*. 2013; Zhu *et al*. 2013). In contrast, the nuclear genome provides many independent, and unlinked gene markers that evolve at different rates (Townsend *et al*. 2008). Nuclear loci are generally considered as fast evolving compared to the chloroplast in plants (Smith 2015). Using both nuclear and chloroplast loci in phylogenetic studies can be informative because they represent different evolutionary histories (biparental and maternal respectively; e.g. Xiang *et al*. 2015).

Depending on the research question, different parts of the nuclear genome can be targeted. When targeting highly diverged taxa we expect exons to be more reasonable candidates compared to intronic or noncoding regions, as they have less variation and can be aligned and compared more readily. On the other hand, intronic or noncoding regions can have excessive length variations and higher substitution rates, often making them difficult or impossible to align, especially in distantly related taxa (Mattee *et al*. 2001; Boekhorst & Snel 2007).

Different approaches have been taken to identify candidate loci for phylogenetic inference, mostly in the form of reduced representation methods (Bragg *et al.* 2015). These approaches often use restriction enzymes to find and target in a subset of the genome, and examples include Reduced Representation Library (RRL), Restriction-site Associated DNA Sequencing (RAD) and Genotyping By Sequencing (GBS). These methods have the advantage in that they do not require reference data, yet they do not work well for distantly related taxa as restriction sites tend to be less conserved as taxa become more distant (Rubin *et al*. 2012).

Another type of reduced representation method is target enrichment, which is more targeted than those reduced representation methods using restriction enzymes. These approaches include PCR generated probes (SCPP, Penalba *et al.* 2014), developing markers that mainly target 3'UTR and coding regions (Tonnabel *et al.* 2014), or

transcriptome based exon capture. Bragg *et al*. (2015) used *Anolis* lizard exons to identify holomologs across *Carlia, Lampropholis* and *Saproscincus* transcriptomes using reciprocal best BLAST (basic local alignment search tool). Li *et al*. (2007) took a genome comparison approach: exons were compared within and between the zebra fish and pufferfish genomes (from Ensembl) to filter presumably single-copy orthologous exons of the two fish followed by checking sequence similarity and coverage parameters using BLAST. These methods can be useful for phylogenetic estimation across more highly diverged taxa (Bragg *et al*. 2015; Portik *et al*. 2016).

Different sets of phylogenetic markers are used depending on the taxonomic level of interest. Targeting Ultra-Conserved Elements (UCEs), with ultraconserved element enrichment, can be useful for intermediate- to deep-level phylogenies (Lemmon & Lemmon 2013). UCEs are not difficult to locate and align because of their conserved sequences (Stephen *et al*. 2008). However, UCEs do not always provide enough phylogenetic signal to resolve shallow nodes in phylogenies (e.g. Faircloth *et al.* 2012). Marker sets used for anchored phylogenomics (Lemmon *et al.* 2012), which is another variation of hybrid enrichment approach, can be useful across multiple phylogenetic levels, as both conserved loci and flanking regions containing variation can be targeted. Likewise, Exon capture has been commonly used for multiple level phylogenetics (Lemmon & Lemmon 2013), however, one often needs prior knowledge such as a reference genome to begin with.

For phylogenetic estimation, it is important to have genetic markers that are orthologous and single or low copy. Orthologs are genes that are homologous due to speciation (Tekaia *et al*. 2016), not due to duplication events. Careful identification of orthologs is crucial, because an accurate estimation of species relationships assumes orthology of the genes used (Tekaia *et al*. 2016). In particular, low copy nuclear genes are useful to resolve closely related species relationships (Sang 2002).

For orthology searches, similarity approaches are commonly used, which depend on comparing genomes and clustering genes by similarity to identify orthologous groups (Tekaia *et al*. 2016). Similarity approaches that rely on reciprocal best hits work well in comparative accuracy (Tekaia *et al*. 2016). However, the reciprocal best hit search is limited to one-to-one ortholog searches and cannot detect one-to-many or many-to-many orthologs (Salichos & Rokas 2011). Additionally, in the case where there is a gene

loss, these methods might result in overinclusion (Salichos & Rokas 2011). Another commonly used method is tree-based approach that uses gene trees to locate orthologous relationships (Teasdale *et al*. 2016). The phylogenetic approaches are regarded as more dependable for orthology assessments compared to similarity based methods (Gabaldón 2008; Kcot *et al*. 2013 and references therein). However, one of the biggest challenges in phylogeney-based search is the uncertainty in known or inferred species relationship (Gabaldón 2008).

To date, a small number of loci from the chloroplast and nuclear regions have been used to estimate phylogenies of *Melaleuca* and *Eucalyptus* (Myrtaceae) by themselves or in a combination of both (Steane *et al.* 2002; Lucas *et al.* 2007; Cook *et al*. 2008; Biffin *et al*. 2010; Edwards *et al*. 2010; Crisp *et al*. 2011; Thornhill *et al.* 2015). Bayly *et al.* (2013) recently carried out phylogenetic analyses using the chloroplast genome of 39 species across five genera (*Eucalyptus, Corymbia, Angophora, Allosyncarpia*, *Stockwellia*). It has often been challenging to resolve genus- or species-level relationships using only a few loci, as in the cases of *Eucalyptus* and *Melaleuca* (e.g. Steane *et al*. 2002; Cook *et al*. 2008; Edwards *et al*. 2010; Crisp *et al*. 2012). This is likely because the small number of loci provided could not offer clear phylogenetic signal to resolve the relationships. Given that much of the diversity of the groups arose in the last approximately 20 million years (Thornhill *et al*. 2015), incomplete lineage sorting and/or ongoing hybridisation might be present, as seen in *Melaleuca* (Cook *et al*. 2008).

In this study, we aimed to identify multiple loci from both the nuclear and chloroplast genomes that are potentially useful for resolving species/generic relationships within *Melaleuca* and *Eucalyptus,* and across the Myrtaceae. Because these two genera are estimated to last have shared a common ancestor approximately 63-72.8 Ma, and the crown age of Myrtaceae is estimated to be 75-93.5 Ma (Thornhill *et al*. 2015), we expect our target loci to be useful for the rest of the genera in Myrtaceae as well. For the nuclear loci we targeted exonic (coding) regions because they provide a useful amount of variation, yet are similar enough to be aligned and compared, even between these two deeply diverged genera. Specifically, we looked for orthologous and low-copy loci for the phylogenetic studies. Ultimately, with these loci we aim to infer phylogenies of *Melaleuca* and *Eucalyptus,* as well as other genera across the family Myrtaceae.

We used transcriptomes and shotgun sequences of *Melaleuca* species, and an annotated genome of *Eucalyptus grandis* to identify orthologous and low-copy nuclear loci. We also describe how we identified chloroplast loci to target by aligning three chloroplast genomes from two *Melaleuca* species and *Eucalyptus grandis*. Finally, we verified our target loci using exon capture and next generation sequencing (NGS) methods.

Material and Methods

Resources used to identify nuclear target loci

Melaleuca *RNA Sequencing and read processing.* Transcriptomes of *Melaleuca quinquenervia* were provided by Sarah Hsieh (Hsieh *et al*., unpublished data). For the transcriptomes, RNA sequencing was performed on eight *M. quinquenervia* as well as eight *Melaleuca alternifolia* individuals. In brief, RNA was extracted from leaf tissue of each of the 16 individuals at two different time points in a plant-fungal interaction experiment. Libraries were prepared from total RNA using the TruSeq™ RNA sample preparation kit v2 (Illumina®, San Diego, USA) and sequenced on the Illumina HiSeq2500 platform with a 150 bp paired-end read protocol. After assessing the read quality, trimming of adaptors and low quality reads Trinity was used to assemble the transcriptome *de novo* (Gabherr *et al*. 2011) for each species. We randomly selected up to a few hundred nuclear contigs per iteration as references to start the target locus identification workflow.

***Whole genome shotgun sequencing* of *two distantly related* Melaleuca *individuals*.** To ensure target loci are present in other *Melaleuca* species, one distantly related (*M. bracteata*) and a closely related species (*M. leucadendra*) to *M. quinquenervia* were shotgun sequenced to confirm presence of target loci. For the whole genome sequencing, genomic DNA was extracted from *M. bracteata* and *M. leucadendra* (See Table S1) using a modified CTAB extraction protocol (Sahu *et al*. 2012) combined with a Qiagen spin column protocol (DNeasy plant mini kit, Qiagen) to maximize quantity and quality of DNA. The library was prepared using a TruSeq DNA LT Sample Preparation Kit (Illumina®, San Diego, USA) following the manufacturer's instructions. The samples were sequenced using the Illumina HiSeq2500 platform with a 150 bp paired-end read protocol. The read counts from the run were 42,756,290 in *M. bracteata* and 44,538,892 in *M. leucadendra*.

The* Eucalyptus grandis *annotated genome. To identify target loci, we used a BLAST search that is implemented in Phytozome (www.phytozome.com), to query selected *M. quinquenervia* transcriptomic contigs against the annotated *Eucalyptus grandis* genome. Phytozome is an online resource that provides access to 68 sequenced and annotated green plant genomes (Goodstein *et al*. 2012): thus, the gene ancestry function in Phytozome was used to check copy numbers of matching loci in *Eucalyptus grandis* as well as across other species in the plant kingdom. Coding DNA Sequences (CDS) of *E. grandis* corresponding to the selected *M. quinquenervia* contigs (targets) were obtained.

Identification of target genetic loci and probe design

Target nuclear loci (Fig. 1). The shotgun sequences of *M. leucadendra* and *M. bracteata,* as well as transcriptomes of *M. alternifolia,* were assembled against the selected *M. quinquenervia* references using CLC genomics workbench (CLC bio, Aarhus, Denmark). From the assembly of the *Melaleuca* shotgun genome sequences of the *M. quinqunervia* reference sequences, the median read coverage of *M. leucadendra* and *M. bracteata* shotgun sequences were 18.4 and 15.7 reads per nucleotide position respectively. A series of filtration steps were developed to remove potentially repetitive, paralogous or high copy loci. The steps were mainly based on 1) average coverage, 2) copy number comparison of loci of interest, 3) alignment and gene tree estimation of target loci. First, we selected contigs for downstream analysis when the median coverage fell within 1 standard deviation of *M. leucadendra* and *M. bracteata*. Contigs outside the median range were excluded because we assumed that the loci might not be unique in the genome. A BLAST search of the selected contigs using the default setting (Expect threshold: -1) was used against the closest fully sequenced and annotated genome (*E. grandis* in this study) from Phytozome. After selecting the highest BLAST hit we used the gene ancestry function to compare putative copy numbers of the queried loci across other species in the plant kingdom.

In the second filtration step we chose contigs that had less than or equal to six homologs per species with an emphasis on single or low copy genes. We excluded those loci with more than six homologs to remove potential high copy loci. We then aligned targeted loci with transcriptomes, shotgun sequences of *Melaleuca* species and the coding sequences of the *E. grandis* genome using Geneious alignment with standard parameters in Geneious v7.1.4 (http://www.geneious.com, Kearse *et al*. 2012).

A tree for each locus was estimated using a Tamura-Nei Genetic Distance Model and Neighbor-Joining tree build method in Geneious v7.1.4.

Thirdly, we filtered out potential paralogs by excluding loci where the gene tree did not show phylogenetic relationships as previously inferred using standard phylogenetic loci (Edwards *et al*. 2010; Thornhill *et al*. 2015). Loci were excluded in case where all the *Melaleuca* species did not cluster together or where *E. grandis* was not clustered outside of *Melaleuca* (as expected, given Thornhill *et al*. 2015). We also excluded loci if *M. leucadendra* and *M. quinquenervia* were not sister species (as expected, given Edwards *et al*. 2010). We repeated the whole procedures by randomly selecting M. *quinquenervia* references until we found over 240 kbp of loci.

Target chloroplast loci. We used the *E. grandis* chloroplast genome from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and shotgun sequences of *Melaleuca leucadendra* and *M. bracteata* to locate target chloroplast loci. First, three chloroplast genomes were aligned using the Mauve alignment in Geneious v7.1.4 (http://www.geneious.com, Kearse *et al*. 2012). The whole alignment was manually checked for insertions and deletions as well as the number of single nucleotide polymorphisms. All the loci with nine or more SNPs out of 500 bp were chosen in search for some of the most variable markers in the chloroplast genome. We also included the *trnL-trnF* region, the *psbA-trnH* intergenic space region, the *matK-trnK* region and the *ndhF* gene that were used in previous phylogenetic studies of Myrtaceae.

Probe design. Both the chloroplast and nuclear probes were designed by NimbleGen (Madison, USA) using their standard parameters of the SeqCap EZ System. We provided them a combination of the following two sets of sequences: 1) ortholog sequences from *E. grandis* and 2) a set of *Melaleuca* consensus sequences for each of the target locus. The number of probes used for chloroplast loci were approximately 10 times less than for nuclear loci. This was to avoid overrepresentation of chloroplast loci in subsequent sequencing because of the higher number of chloroplast genomes in plant cells.

DNA extraction and library preparation

DNA extraction and shearing of DNA. Genomic DNA (gDNA) was extracted from leaf tissues of 48 samples across five genera in Myrtaceae (Table. 1) using Qiagen

DNeasy 96 Plant Kit (Qiagen, Hilden, Germany). The samples in this study include 43 species as follows: *Arillastrum* (one species), *Corymbia* (four species), *Eucalyptus* (31 species), *Heteropyxis* (one species), and *Melaleuca* (six species). The six species from the genus *Melaleuca* used represent the three main monophyletic clades of the tribe of the phylogeny of Edwards *et al*. (2010; also see introduction of this chapter). We followed the generic name changes of Craven *et al*. (2014) for *Melaleuca* species. In short, genera in the tribe Melaleuceae (e.g. *Calothamnus* species) are now *Melaleuca*. The gDNA was quantified 1) using LabChip DS Spectrophotometer (Trinean, Gentbrugge, Belgium) or 2) by separating them by agarose gel electrophoresis. For library preparation the input DNA was diluted to 600 ng in 100 µL. The diluted gDNA was sheared for a target size ranging between 250-400 bp using the Bioruptor ultrasonicator (Diagenode, Liege, Belgium) for 30 sec at high intensity. This was repeated 10 times with a 30 sec waiting period between cycles.

Library preparation. The gDNA library was prepared using the Rohland & Reich protocol (2012), with minor modifications in reagents and incubation settings. After the enrichment PCR we randomly picked three samples for Qubit fluorometric quantification using dsDNA HS Assay kit (Qubit Invitrogen, Carlsbad, USA). All the samples were quantified by agarose gel electrophoresis. The average DNA concentrations of 2.8 ng/µL was estimated by comparing the Qubit quantification of the three samples with the DNA concentrations estimated by agarose gel electrophoreses. To achieve a minimum of 1 µg of total library DNA, 9.2 µL per sample were pooled after estimating the average DNA concentration.

Target hybridisation, recovery, wash, and sequencing

We hybridised the probes to the sample target DNA (library) using the SeqCap EZ Developer Library (NimbleGen, Madison, USA) following the manufacturer's instructions with minor modifications in the hybridisation mix preparation and incubation settings. The main modification to the hybridisation steps was to denature the DNA library and hybridisation mix for 10 minutes at 95°C, then gradually decrease the temperature from 95°C to 47°C followed by a 47°C incubation for 72 hours. This was to ensure that the DNA and probes are single stranded in the solution to maximise chances of probes hybridising to the target DNA. This step can be critical in case some of the target loci with high annealing temperature as the temperature drop down might be the only chance for the probes to hybridise back to the target DNA. Recovery

and wash of hybridised samples were carried out using the SeqCap Hybridisation and Wash Kit (NimbleGen, Mannheim, Germany) following the manufacturer's instructions with a slight modification in temperature settings (See Appendix S1 for detailed lab protocol).

Following recovery of the captured samples, we performed semi-qPCR to quantify the DNA in the hybridised library and estimate the number of cycles for the indexing PCR step. After the indexing PCR of captured libraries, we purified the samples with Sera-mag SPRI beads to remove primer dimers from the indexing PCR product. We quantified the concentration of the hybridised libraries using the Qubit fluorometer (Qubit Invitrogen, Carlsbad, USA). One capture of a library that consisted of 48 samples (see Table 1) was then sequenced on the Illumina Miseq platform (100 bp paired-end read protocol) at the Bio-molecular Research Facilities at the Australian National University.

Read filtration

The quality of the raw reads was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We used Flexbar (Dodt *et al.* 2012) to sort by barcodes, and to remove barcodes and low quality reads. The processed reads were double-checked using FastQC for the quality of the cleaned reads. The reads were mapped against the *E. grandis* targets using the CLC genomics workbench (CLC bio, Aarhus, Denmark). If a nucleotide was represented from more than 60% of the reads for a site we took that nucleotide for the final sequences. One sample (*Arillastrum*) was excluded from the subsequence analyses because of poor read quality.

Nucleotide diversity

Sequences of each target locus were aligned using L-INS-I (Katoh & Standley, 2013) in MAFFT v.7.215 (http://mafft.cbrc.jp/alignment/software/). Then the alignments of 209 loci were concatenated using FasConCat (https://www.zfmk.de/en/research/research-centres-and-groups/fasconcat). The nucleotide diversity across the 47 samples of 209 loci compared to *Eucalyptus globulus* was calculated using Geneious v7.1.4 (http://www.geneious.com, Kearse *et al*. 2012).

Nucleotide diversity versus percentage of reads mapped back to the *E. grandis* targets was compared using linear regression in Prism version 6 (GraphPad). This was to test whether the nucleotide diversity might have been a factor which contributed to the reduced capture sensitivity measure. The GC contents of *E. grandis* targets were calculated using Geneious (v7.1.4 http://www.geneious.com, Kearse *et al*. 2012). The GC content was compared with the average coverage to test if the GC content might have been an attribute for low average coverage of loci.

The number of parsimony informative sites (informative sites) for each locus was calculated in Mega v6 (Tamura *et al*. 2013). We assessed 1) the number of informative sites, and 2) the percentage of informative sites per target length. The number of informative sites is not necessarily a definite measure of phylogenetic informativeness as it would depend on which phylogenetic analyses one choose to conduct. Nevertheless, the number of informative sites can be used as a measure of phylogenetic usefulness for each locus (e.g. Portik *et al*. 2016).

Results

We successfully sequenced 43 species across five genera (*Melaleuca*, *Eucalyptus*, *Corymbia*, *Arillastrum*, *Heteropyxis*) (Table 1). We recovered 209 nuclear and chloroplast loci consisting of 263,164 bp out of 212 target loci (Supplementary Table S1). The target nuclear loci included 176 markers of 241,047 bp, which covers 0.39% of the *E. grandis* nuclear genome. In the chloroplast, 36 markers of 22,117 bp covering 13.8% of the *E. grandis* chloroplast genome were included. The length of chloroplast and nuclear target loci ranged from 216 bp to 6,555 bp with the majority of them between 500 bp to 2,000 bp in size (Table 2).

Read filtration

The total number of reads for the 209 loci across 47 samples was 47,707,942. The number of reads filtered by barcodes that is unique to each sample ranged from 17,766-4,366,802 (Table 1, Fig. 2). The average number of reads per sample was 991,582 and the total number of reads without valid barcodes was 387,400.

Performance of exon capture

The following three criteria were used to assess the capture performance using the filtered reads: 1) capture specificity, 2) capture sensitivity and 3) enrichment factor (EF). These measures showed us how successful the capture worked.

1. Capture specificity

Capture specificity refers to the percentage of filtered reads that were mapped back to the target sequences (Penalba *et al*. 2014). The percentages of reads that mapped back to the *E. grandis* targets were 31.9-66.7% for the different sample, with an average of 57.6% (Table 1). The average capture specificity was as follows for each genus: *Eucalyptus* (60.2%), *Arillastrum* (58.7%), *Corymbia* (57.6%), *Melaleuca* (46.3%), and *Heteropyxis* (31.9%).

2. Capture sensitivity

Capture sensitivity is the percentage of the targeted loci that is covered by at least one read (Penalba *et al*. 2014). Coverage of the 209 target loci of the 47 samples is provided in Supplementary Table S1. The average coverage is for target loci, excluding the flanking regions. The average coverage per sample ranged from 2.8 to 491.6, with an average of 140. The data matrix recovery of 47 samples across 209 loci was over 99.9% (Supplementary Table S1). Only 0.002% of the data matrix attained zero average coverage. We were able to attain the following percentage of data matrix for each genus: 99.52% in *Melaleuca*, 99.95% in *Eucalyptus*, 98.68% in *Corymbia*, 100% in *Arillastrum*, and lastly 97.6% in *Heteropyxis* (Supplementary Table S1). Average coverage of each genus varied from 41.4-48.9 in *Melaleuca,* 52.2-66.7 in *Eucalyptus,* 55.5-59.7 in *Corymbia* and 57.2-60.2 in *Arillastrum*. A heat map showing the average coverage across all the samples and loci is shown in Figure 3. Up to 83.8% of the data matrix had average coverage of 20 (or more), which was set as another threshold to assess capture sensitivity. In *Heteropyxis*, 88.24% of the data matrix was recovered in the chloroplast and 4% in the nuclear. The average coverage in the chloroplast loci (196.8 reads) was 1.5 times higher than in the nuclear target loci (129.5 reads; Table 1). The average coverage of chloroplast target loci (3.4-1209.8 reads) varied more greatly compared to the nuclear loci (2.1-449.1 reads). Samples such as *Heteropyxis*, which is an outgroup of both *Melaleuca* and *Eucalyptus* (see Thornhill *et al*. 2015) showed up to eight times as high average coverage in the chloroplast compared to nuclear loci.

Eucalyptus grandis target sequence length and percentage of GC contents were compared with the average coverage using a linear regression (Fig. 5). This was to test if GC contents attributed to capture specificity. The GC content showed a significant negative correlation with average coverage ($R^2 = 0.2072$, $P<0.0001$). Particularly, when GC content exceeded 70%, the average coverage decreased significantly.

3. Enrichment factor

An enrichment factor (EF) for both chloroplast and nuclear loci was assessed. Enrichment factor is calculated as follow (Mertes *et al*. 2011): EF = (Reads on Target/Total Number of Reads)/(Target size/Genome size). We used the average obtained for "(Reads on Target/Total Number of Reads)", which is 0.576 (see Table 1). The combined chloroplast and nuclear target size is 263,164 bp. We used *E. grandis* chloroplast genome size of 160,137 bp (Pavita *et al*. 2011) and the nuclear genome is estimated as 605 Mbp (Myburg *et al.* 2014). Enrichment factor (EF) in this study is 1,324.

Nucleotide diversity and coverage

The P-distance of concatenated chloroplast and nuclear data relative to *Eucalyptus globulus* ranged from 0.04 to 0.43, with an average of 0.11. The P-distance of *Corymbia calophylla, Eucalyptus goniantha* and *E. stenostoma* ranged from 0.29 to 0.45. Due to the low number of filtered reads, the nucleotide diversity of the three samples was multiple times higher compared to the average P-distance across all taxa. Therefore, we excluded those samples from further nucleotide diversity analyses. Another outlier was *Heteropyxis natalensis* (P-distance = 0.43), which is an out-group from the rest of the samples in Myrtaceae according to a previous phylogenetic study (Thornhill *et al.* 2015). We did not exclude *Heteropyxis natalensis* because the low coverage in the nuclear targets might have been due to phylogenetic distance rather than due to low number of filtered reads. *Heteropyxis* had over 149,000 filtered reads (47,754 filtered reads mapped back to target). The average coverage of chloroplast loci in *Heteropyxis* was 60, which is eight times higher coverage than the nuclear loci (Table 1).

We used linear regression to test if there is correlation between average coverage and P-distance. When all the samples were included, P-distance and the average coverage had a significant negative correlation ($R^2 = 0.158$, $P = 0.0057$). The linear regression of P-distance and the percentage of reads mapped against the targets excluding outliers showed a significant negative correlation ($R^2 = 0.36$, $P < 0.0001$).

Assessing phylogenetic informativeness

The number of informative sites of each locus was calculated as a measure of phylogenetic informativeness of each locus (Table 2). The average number of informative sites in the combined chloroplast and nuclear target regions was 207 per locus. In the chloroplast loci, the number of informative sites ranged from 12 to 193, with an average of 50.7. In the nuclear loci, the number of informative sites varied from 22 to 859, with an average of 236.3.

The average informative sites per target locus were 15.8%. The average percentage of informative sites per chloroplast locus ranged from 2.4% to 16.7%, with an average of 7.8%. In the nuclear loci, the percentage of informative sites per target locus ranged from 1.6% to 33.5% with an average of 17.4%. In the chloroplast target loci, we included four gene regions (*matK-trnK*, *ndhF*, *psbA-trnH* and *trnL-trnF*) previously used in Myrtaceae phylogenetic studies (e.g. Edwards *et al.* 2010; Wilson *et al.* 2001). The average number of informative sites in the previously used chloroplast loci was 128.3 and the percentage of informative sites was 9.9%. Compared to the other target chloroplast loci that had an average number of informative sites (40) and the percentage of informative sites (7.5%), variation in the previously used phylogenetic markers were higher.

Discussion

In this study, we identified 212 low copy and orthologous nuclear and chloroplast loci for phylogenetic estimations of *Melaleuca* and *Eucalyptus*. The sequencing results demonstrate that the target loci were mostly successfully captured in *Melaleuca* and *Eucalyptus* as well as in three other genera targeted in the family with a dropdown of capture performance in nuclear loci in *Heteropyxis*, which is an outgroup of *Melaleuca* and *Eucalyptus*. The average coverage for nuclear loci in *Heteropyxis* was 7.5, which allows sequence assembly, but with relatively low confidence. Conversely, the chloroplast loci still worked successfully for this genus, with an average coverage of 60. It is possible that because the nuclear loci are more variable compared to the chloroplast loci that the capture sensitivity dropped down significantly only in the nuclear genes for this genus, compared to the other genera. The capture performance based on specificity and sensitivity across 209 loci was as good or better than in many other published target capture studies (e.g. Bi *et al*. 2012; Penalba *et al.* 2014) based on

capture specificity and sensitivity. These targeted loci are expected to be useful across most other genera in Myrtaceae.

In target nuclear regions, 83.51 % of the data matrix had an average coverage of more than 20, which again shows that the exon capture worked successfully across the majority of samples and most of the loci. For the probe design, approximately 10 times fewer probes for chloroplast targets compared to the nuclear loci were used to designed the probes due to much higher abundance of chloroplast genomes per plant cell. The average coverage of chloroplast loci (196.8) was higher than the nuclear loci (129.5), but in some samples the nuclear loci had higher or similar average coverage to that of the chloroplast. This result shows that reduced number of probes for the chloroplast targets compared to nuclear is important to obtain relatively balanced average coverage.

Here we make two suggestions for target loci selection and the associated lab protocol. First, we suggest excluding target loci with extremely high GC contents. Some of the target loci that did not work well had high GC contents (around 70 percent or more). This is likely because sequences with very high GC contents might not hybridise well due to formation of secondary structures (Mamanova *et al*. 2010). Bi *et al*. (2012) also discovered that targets with either very high or low GC percentages resulted in low coverage.

Second, we suggest a modification to the lab protocol in our study. Varied capture efficiency within the same species from the same genera might have been because of the input concentration of the library before pooling samples. We pooled samples using averaged concentration from Qubit quantification and gel electrophoresis images for all the samples. If the input amount of the library had been more constant by checking the concentration of each sample individually, we might have observed even higher coverage across all samples, particularly those from the same species.

We assessed the number and percentage of parsimonious informative sites from each target as a proxy for phylogenetic informativeness. Based on the number and the percentage of informative sites, we have successfully identified many potentially informative nuclear and chloroplast loci for multiple taxonomic depths in Myrtaceae.

The average percentage of informative sites in chloroplast markers used in previous studies (9.9%) was slightly higher than chloroplast markers in this study (7.8%).

Our study used multiple genomic resources including an annotated genome of *Eucalyptus* as well as transcriptomes and whole genome shotgun sequences from *Melaleuca* species. Our method identified over 200 loci from nuclear and chloroplast loci. The novelty of this study is that we have used different combinations of workflow to find orthologous and low copy loci by 1) comparing average coverage of reference sequences across two shotgun sequences 2) comparing copy number of loci across *E. grandis* and other plant genera, and 3) assessing gene trees. As more annotated genomes, transcriptomes and other genomic resources become available across different groups of organisms, the methods outlined in this study will become more easily achievable.

References

Akhani, H., Malekhammadi, M., Mahdavi, P., Gharibiyan, A. & Chase, M.W. (2013) Phylogenetics of the Irano-Turanian taxa of *Limonium* (Plumbaginaceae) based on ITS nrDNA sequences and leaf anatomy provides evidence for species delimitation and relationships of lineages. *Botanical Journal of the Linnean Society* 171: 519–550.

Bayly, M.J., Rigault, P., Spokevicius, A., Ladiges, P.Y., Ades, P.K., Anderson, C., Bossinger, G., Merchant, A., Udovicic, F., Woodrow, I.E. & Tibbits, J. (2013) Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* 69: 704-716.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* 13: 403.

Biffin, E., Harrington, M.G., Crisp, M.D., Craven, L.A. & Gadek, P.A. (2007) Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and Myrtaceae. *Molecular Phylogenetics and Evolution* 43: 124-139.

Boekhorst, J. & Snel, B. (2007) Identification of homologs in insignificant blast hits by exploiting extrinsic gene properties. *BMC Bioinformatics* 8: 356.

Brown, G.K., Udovicic, F. & Ladiges, P.Y. (2001) Molecular phylogeny and biogeography of *Melaleuca, Callistemon* and related genera (Myrtaceae). *Australian Systematic Botany* 14: 565-585.

Bragg, J.G., Potter, S., Bi, K. & Moritz, C. (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources* 16: 1059–1068.

Cook, L.G., Morris, D.C., Edwards, R.D. & Crisp, M.D. (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* 47: 506–522.

Craven, L.A., Edwards, R.D. & Cowley, K.J. (2014) New combinations and names in *Melaleuca* (Myrtaceae). *Taxon* 63: 663-670.

Crisp, M.D., Burrows, G.E., Cook, L.G., Thornhill, A.H. & Bowman, D.M. (2011) Flammable biomes dominated by eucalypts originated at the Cretaceous-Palaeogene boundary. *Nature Communications* 2: 193.

Dodt, M., Roehr, J.T., Ahmed, R. & Dieterich, C. (2012) FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1: 895–905.

Dong, W., Liu, J., Yu, J., Wang, L. & Zhou, S. (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One* 7: e35071.

Edwards, R.D., Craven, L.A., Crisp, M.D. & Cook, L.G. (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L. (Myrtaceae) is not monophyletic. *Taxon* 59: 744-754.

Faircloth, B.C. & Glenn, T.C. (2014) Protocol: Preparation of an AMPure XP substitute (AKA Serapure). doi: 10.6079/J9MW2F26.

Gabaldón, T. (2008) Large-scale assignment of orthology: back to phylogenetics? *Genome Biology* 9: 235.

Gielly, L. & Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. *Molecular Biology and Evolution* 11: 769–777.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. & Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40: D1178-D1186.

Grabherr, M.G., Haas, B.J., Yassour M, Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. &

Chen, Z. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644-652.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kcot, K.M., Citarella, M.R., Moroz, L.L. & Halanych, K.M. (2013) PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics* 9: 429-435.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. & Thierer, T. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Lemmon, A.R. & Lemmon, E.M. (2012) High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology* 61: 745–761.

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99-121.

Li, C., Ortí, G., Zhang, G. & Lu, G. (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* 7: 44.

Lucas, E.J., Harris, S.A., Mazine, F.F., Belsham, S.R., Nic Lughadha, E.M., Telford, A., Gasson, P.E. & Chase, M.W. (2007) Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). *Taxon* 56: 1105–1128.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. & Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111–118.

Maricic, T., Whitten, M. & Pääbo, S. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* 5: e14004.

Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J.M., Van Der Zaag, P.J., Franke, A., Nilsson, M., Lehrach, H. & Brookes, A.J. (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics* 10: 374–386.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D. & Goodstein, D.M. (2014) The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.

Paiva, J.A., Prat, E., Vautrin, S., Santos, M.D., San-Clemente, H., Brommonschenkel, S., Fonseca, P.G., Grattapaglia, D., Song, X., Ammiraju, J.S. & Kudrna, D. (2011) Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC genomics* 12: 137.

Peñalba, J.V., Smith, L.L., Tonione, M.A., Sass, C., Hykin, S.M., Skipwith, P.L., McGuire, J.A., Bowie, R.C. & Moritz, C. (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources* 14: 1000–1010.

Portik, D.M, Smith, L.L. & Bi, K. (2016) An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources* 16: 1069–1083.

Potts, B.M. & Reid, J.B. (2003) Tasmania's eucalypts: their place in science. *Papers and Proceedings of the Royal Society of Tasmania* 137: 21–37.

Rohland, N & Reich, D. (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* 22: 939–946.

Rubin, B.E., Ree, R.H. & Moreau, C.S. (2012) Inferring phylogenies from RAD sequence data. *PLoS ONE* 7: e33394.

Sahu, S.K., Thangaraj, M. & Kathiresan, K. (2012) DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *ISRN Molecular Biology* 205049: 1–6.

Salichos, L. & Rokas, A. (2011) Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS ONE* 6: e18755.

Sang, T. (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Smith, D.R. (2015) Mutation rates in plastid genome: they are lower than you might think. *Genome Biology and Evolution* 7: 1227-1234.

Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E. & Potts, B.M. (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* 15: 49–62.

Tamura, K., Stecher, G., Peterson, D., Fillipski, A. & Kumur, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Teasdale, L.C., Koehler, F., Murray, K.D., O'Hara, T. & Moussalli, A. (2016) Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources* 16: 1107-1123.

Tekaia, F. (2016) Inferring orthologs: open questions and perspectives. *Genomics Insights* 9: 17–28.

Thornhill, A.H., Ho, S.Y., Külheim, C. & Crisp, M.D. (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Molecular Phylogenetics and Evolution* 93: 29–43.

Tonnabel, J., Olivieri, I., Mignot, A., Rebelo, A., Justy, F., Santoni, S., Caroli, S., Sauné, L., Bouchez, O. & Douzery, E.J. (2014) Developing nuclear DNA phylogenetic markers

in the angiosperm genus *Leucadendron* (Proteaceae): A next-generation sequencing transcriptomic approach. *Molecular Phylogenetics and Evolution* 70: 37–46.

Townsend, T.M., Alegre, R.E., Kelley, S.T., Wiens, J.J. & Reeder, T.W. (2008) Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution* 47: 129–142.

Wilson, P.G., O'Brien, M.M., Gadek, P.A. & Quinn, C.J. (2001) Myrtaceae revisited: a reassessment of infrafamilial groups. *American Journal of Botany* 88: 2013-2025.

Xiang, Q.P., Wei, R., Shao, Y.Z., Yang, Z.Y., Wang, X.Q. & Zhang, X.C. (2015) Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Molecular Phylogenetics and Evolution* 82: 1-14.

Zhu, W.-D., Nie, Z.-L., Wen, J. & Sun, H. (2013) Molecular phylogeny and biogeography of *Astilbe* (Saxifragaceae) in Asia and eastern North America. *Botanical Journal of the Linnean Society* 171: 377–394.

Tables.

Table 1. Number of filtered reads, percentage of reads mapped back to the *Eucalyputus grandis* targets and the average coverage of 47 samples. P-distance of each species compared to the *Eucalyputus globulus.* Cpl refers to chloroplast loci and ncl refers to nuclear loci.

Collector no.	Species	Number of filtered reads	Reads mapped to target (%)	Coverage (cpl)	Coverage (ncl)	Coverage (total)	P-dist
McCoy s.n.	*Arillastrum gummiferum*	1594438	57.2	230.0	219.6	221.3	0.10
McCoy s.n.	*A. gummiferum*	849038	60.2	105.1	126.0	122.7	0.10
Cs229	*Corymbia bunites*	2785808	56.9	389.3	383.0	384.0	0.11
Cs377	*C. calophylla*	18232	58.1	7.5	2.1	3.0	0.47
Cs380	*C. haematoxylon*	416744	59.7	187.0	49.4	71.2	0.12
Cs493	*C. peltata*	2237606	55.5	545.3	276.5	319.0	0.11
Cs474	*Eucalyptus brockwayi*	1213666	62.1	326.0	180.2	203.3	0.06
ANBG9404806	*E. coccifera*	1250566	55.2	235.4	169.2	179.7	0.04
CANB632680	*E. extrica*	1028128	64.9	451.3	141.9	190.8	0.06
Cs729	*E. globoidea*	679956	61.8	52.0	109.1	100.1	0.06
CANB688145	*E. globulus* ssp. *maidenii*	1005544	58.1	145.2	143.6	143.9	0.04
CANB494954	*E. globulus* ssp. *pseudoglobulus*	307928	52.2	18.0	41.1	37.4	0.00
Cs357	*E. gomphocephala*	1085564	58	188.9	153.1	158.7	0.05
Cs424	*E. goniantha* ssp. *kynoura*	17766	60.6	3.4	2.7	2.8	0.45
Cs496	*E. howittiana*	715782	58.8	185.6	100.0	113.5	0.05
Cs275	*E. insularis*	586980	60.1	100.4	86.0	88.3	0.06
CANB413807	*E. insularis*	351106	65.8	24.8	59.2	53.8	0.06
Cs716	*E. leucoxylon* ssp. *leucoxylon*	862498	58.2	163.1	123.4	129.7	0.05
Cs721	*E. ligulata* ssp. *ligulata*	100490	61.5	19.8	15.0	15.8	0.10
Cs574	*E. moorei* ssp. *serpentinicola*	1360608	58.5	220.7	191.8	196.3	0.06
Cs744	*E. nitida*	603824	55.4	108.8	79.3	83.9	0.06
Cs708	*E. optima*	285804	60.3	52.1	42.3	43.8	0.05
Cs305	*E. pachycalyx* ssp. *pachycalyx*	52466	59	16.9	6.9	8.5	0.14

Cs426	*E. pachyloma*	398460	63	84.3	60.4	64.1	0.06
Cs742	*E. perriniana*	369972	60.3	120.4	53.2	63.8	0.04
Cs482	*E. pilularis*	1413500	62	262.3	204.9	213.9	0.06
CANB638520	*E. platydisca*	1465552	62	228.5	220.8	222.0	0.06
Cs431	*E. pleurocarpa*	667054	62.2	176.4	96.0	108.7	0.06
	E. preissiana ssp. *lobata*	121636	58.7	34.1	16.7	19.5	0.08
Cs227	*E. pumila*	1299962	60.9	219.2	197.1	200.6	0.05
CANB632673	*E. selachiana*	344242	64.8	144.2	46.7	62.1	0.07
CANB693174	*E. selachiana*	1070662	59.5	171.8	154.2	157.0	0.07
Cs368	*E. spathulata*	1080026	56	132.8	151.3	148.4	0.05
Cs366	*E. staeri*	209512	60.2	18.8	32.2	30.1	0.07
Cs733	*E. stenostoma*	31016	59.9	5.1	4.7	4.7	0.29
Cs191	*E. stoatei*	1225546	59.3	261.9	176.6	190.1	0.05
Cs769	*E. tereticornis* ssp. *tereticornis*	1691082	60.4	104.5	268.6	242.7	0.05
Cs715	*E. verrucata*	438384	60.7	98.0	62.7	68.3	0.06
Cs727	*E. virginea*	313464	66.7	33.4	53.5	50.4	0.06
Cs492	*E. willisii* ssp. *falciformis*	369260	63.2	121.6	52.3	63.2	0.06
CANB576168	*Heteropyxis natalensis*	149702	31.9	60.0	7.5	15.8	0.43
RDE176	*Melaleuca capitata*	3455370	48.9	1209.8	356.9	491.6	0.14
Harwood1546	*M. cornucopiae*	4366802	44.3	710.7	449.1	490.4	0.14
AF4308	*M. foliolosa*	1434504	48.1	287.9	158.6	179.0	0.15
RDE75	*M. squarrosa*	1971326	47.8	308.2	227.5	240.2	0.14
AF4286	*M. sylvana*	1671790	47.3	212.3	193.8	196.7	0.14
RDE154	*M. gracilis*	1634982	41.4	469.0	139.3	191.4	0.14
	AVERAGE	991582	57.6	196.8	129.5	140.1	0.11

Table 2. Information on the target chloroplast and nuclear loci that were identified for *Melaleuca* and *Eucalyptus* phylogenies. 47 samples across the family Myrtaceae. Na in the Annotation column refers to not available. C and n in the Locus Name column refers to chloroplast and nuclear respectively. Average coverage refers to the average number of reads mapped back to the target *Eucalyptus grandis* sequences. Length of sequence refers to the length of each target locus. GC (%) indicates the percentage of the nucleotide guanine-cytosine content from the aligned sequences of the 47 Myrtaceae samples. Informative sites (number/percentage) refer to parsimony-informative sites. Annotations of each locus from the Phytozome website is in the last column.

Locus Name	Average Coverage	Length of Sequence (bp)	GC (%)	Informative Sites (number)	Informative Sites (%)	Annotation
c_20012500	188.7	494	31	37	7.4	na
c_25013000	192.4	500	35	49	10.1	na
c_30013500	258.1	500	32.2	45	8.7	na
c_35014000	345.1	500	31.8	37	7.3	na
c_1740117900	129.5	499	40.3	21	4.2	na
c_1980120300	203.3	499	38.3	17	3.4	na
c_2080121300	334.3	500	39.6	16	3.2	na
c_4510145600	152.0	500	40.8	12	2.4	na
c_4902049519	176.0	499	34.7	29	5.6	na
c_5020150700	201.0	498	38.2	16	3.2	na
c_5230152800	366.2	500	34.4	30	5.9	na
c_5550156000	160.5	495	32.1	32	6.3	na
c_6490165400	166.9	500	31.6	40	8.0	na
c_6760168100	285.6	499	34.1	34	6.8	na
c_6810168600	143.7	500	38.6	44	8.6	na
c_6919169690	142.3	500	32	36	7.1	na
c_7290173400	451.9	495	37.2	19	3.8	na
c_8330183800	78.8	499	29.9	41	8	na
c_9002690525	352.9	500	38.8	35	7.0	na
c_9220192700	259.2	500	36.2	38	7.5	na
c_150761151760	235.1	994	30.4	116	11.5	na

c_152701153200	57.1	498	27.5	40	7.9	na
c_154601155100	94.0	499	33.3	41	8.0	na
c_157101157600	132.1	498	36.6	40	8.0	na
c_158701159200	134.2	500	31	36	7.1	na
c_164501165000	153.2	498	29.5	51	10.1	na
c_165301165800	205.2	499	30.7	80	15.8	na
c_166301166800	108.3	494	25.5	68	13.1	na
c_167001167500	75.4	500	26.8	59	11.7	na
c_*matK-trnK*	324.6	2285	32.6	193	8.3	na
c_*ndhF*	191.9	2060	31.5	173	8.7	na
c_*psbA-trnH*	79.2	553	29.7	101	16.7	na
c_*trnL-trnF*	117.2	764	35.7	46	5.8	na
n_2	120.9	1956	45.7	332	16.8	Peptidase family C78 family not nominated
n_13	171.5	225	48	26	11.6	Guanine nucleotide exchange domain
n_208	145.1	2007	43.7	441	21.9	Trimethylguanosine synthase
n_1873	124.4	1284	47	161	12.4	Adenylate cyclase associated protein
n_1931	133.1	1326	48	207	15.5	Uncharacterised conserved protein contains RCC1 domain
n_3075	147.3	996	47.3	148	14.8	CDS Apolipoprotein D Lipocalin
n_3214	160.5	2607	43.7	379	14.4	RNA polymerase III transcription factor TFIIIC
n_3659	210.2	1332	47.7	22	1.6	CDS ACT domain TRNANUCLEOTIDYLTRANSFERASE 1
n_3712	123.5	1110	45.2	179	16.0	Adenine deaminase adenosine deaminase
n_4167	273.8	1662	45.2	239	14.3	na
n_4178	141.9	2373	43.6	297	12.4	GRIP domain MYOSIN HEAVY CHAINRELATED protein binding CDS
n_4206	135.0	456	46.9	72	15.7	Membrane transport protein
n_4460	144.4	1380	49.1	189	13.6	Cyclophilin type peptidylprolyl cistrans isomerase CDS
n_4589	85.9	1242	62.6	234	18.5	No functional annotations CDS
n_4706	84.1	513	53	89	17.3	PROTEASOME REGULATORY SUBUNITS
n_4744	259.5	1746	46	184	10.5	Microtubuleassociated protein essential for anaphase spindle elongation CDS
n_5190	170.1	1518	50.2	265	17.1	Uncharacterised gene CDS
n_5930	157.8	1863	44	322	17.1	Serine threonine protein kinase CDS

n_6015	159.9	2313	44.5	259	11.1	DNA excision repair protein ERCC 3
n_6199	147.5	1683	43.9	278	16.4	C3 sterol dehydrogenase 3betahydroxysteroid dehydrogenase and related dehydrogenases
n_22854	147.9	1637	48.1	310	18.7	No functional annotations
n_23033	133.7	1320	45.1	195	14.8	SEC14 RELATED PROTEIN
n_25721	153.7	567	38.3	98	16.8	Guanylyl cyclase Uncharacterized conserved protein
n_27210	178.1	1236	46.4	222	17.7	Family not named CDS
n_30177	133.4	990	53.6	169	17.0	Myblike DNAbinding domain
n_30296	80.5	1650	65.1	392	23.4	Pectinesterase and lant invertase pectin methylesterase inhibitor
n_30327	55.5	321	72.3	47	13.9	Protein of unknown function DUF3743
n_30355	198.3	651	50.8	100	15.3	Peroxisomal membrane protein MPV17 and related proteins
n_30410	112.5	600	55.2	115	18.5	Predicted protein tyrosine phosphatase
n_30431	92.4	1479	49.2	250	16.8	Proliferationassociated nucleolar protein NOL1
n_30471	105.7	687	60.1	133	19.2	Singlestranded DNA-binding protein
n_30519	120.2	1902	44.1	371	19.3	XS zinc finger domain XS domain XH domain
n_30562	145.8	1290	51.6	256	19.7	Arabidopsis proteins of unknown function
n_30839	176.8	5070	48.8	775	15.2	Ubiquitin activating enzyme
n_30906	124.8	822	48.7	130	15.8	Mitochondrial chloroplast ribosomal protein L22
n_31003	126.6	1692	50.5	270	15.7	Predicted GTPases
n_31043	150.8	1821	50.5	236	12.8	Cytosine deaminase FCY1 and related enzymes
n_31075	133.6	1128	51.3	157	13.6	Cell cycle arrest protein BUB3
n_31209	162.2	1788	53.2	232	12.8	Darabinono1 4lactone oxidase
n_31220	194.4	1635	49.1	216	13.0	Pyrophosphatedependent phosphofructo1kinase
n_31275	140.5	1056	50.3	235	22.1	NAD dependent epimerase dehydratase related
n_31348	101.6	729	50.5	122	16.2	Membrane protein involved in ER to Golgi transport
n_31501	147.1	1977	49.6	342	17.2	na
n_31545	218.1	687	53.7	105	15.3	na
n_31598	148.4	4044	47.5	535	13.1	Intracellular membranebound Ca^{2+}independent phospholipase A2
n_31662	161.0	1029	51.5	200	19.4	na
n_31750	134.4	1425	49.1	241	16.7	na
n_31806	102.8	1104	43	187	17.3	Carbon catabolite repressor protein 4
n_32080	85.6	705	47.9	140	19.6	Uridylate kinase adenylate kinase

n_32112	127.1	675	51.4	128	18.9	Predicted glycine cleavage system H protein
n_32269	127.2	2664	44.7	473	17.5	ATP dependent helicase smarca
n_32478	156.8	1002	52.9	203	20.0	Zincbinding oxidoreductase
n_32479	134.7	1458	51.2	363	24.7	Serine carboxypeptidases lysosomal cathepsin A
n_32563	198.2	1056	45.5	185	17.5	Serine threonine protein kinase
n_32592	142.1	972	58.2	188	19.1	Uncharacterized enzymes related to aldose 1epimerase
n_32608	169.4	621	46.1	112	17.8	GTPase Rab18 small G protein superfamily
n_32660	45.7	1341	65.2	333	24.6	UDPglucuronosyl and UDPglucosyl transferase
n_32708	113.9	2127	50.1	414	19.3	Formin Homology 2 Domain
n_32747	149.1	1281	48.9	224	17.4	Serine threonine protein kinase
n_32861	102.5	1443	48.9	247	17.2	Sulfite exporter TauE SafE ATPDEPENDENT PROTEASE
n_32864	121.1	567	45.5	108	18.7	LUPUS LA RIBONUCLEOPROTEIN
n_32883	20.4	1509	70.8	307	20.1	Hs1pro1 Nterminus and Hs1pro1 protein Cterminus
n_33080	145.5	1425	46.2	222	15.5	Serine protease
n_33128	159.4	1710	46.1	256	15.2	Predicted ATPase
n_33152	146.1	3429	45.8	522	15.1	Chromatin remodeling protein HARP SMARCAL1 DEADbox superfamily
n_33353	175.6	846	46.2	109	12.8	Protein of unknown function DUF124
n_33384	215.7	1209	44.3	153	12.7	Tumor differentially expressed TDE protein
n_33409	80.0	411	51.6	69	16.5	na
n_33440	85.5	1542	48.6	450	28.6	EUKARYOTE SPECIFIC DSRNA BINDING PROTEIN
n_33442	122.7	2535	46.5	382	15.0	Betagalactosidase
n_33459	46.1	474	43	116	24.3	na
n_33482	91.1	2163	63.2	407	18.5	Serine threonine protein kinase
n_33521	150.8	1650	44.9	222	13.3	Branched chain aminotransferase BCAT1
n_33548	111.8	1575	48.3	221	13.9	Xanthine uracil transporters
n_33580	154.9	1668	52	390	23.1	Putative adipose regulatory protein
n_33688	132.4	1629	50.9	281	17.2	Deoxyribodipyrimidine photolyase cryptochrome
n_33877	81.7	1953	51.6	497	25.2	SPX domain and EXS family
n_33888	96.9	1203	61.7	232	19.2	Arylacetamide deacetylase
n_33912	133.8	1482	46	186	12.5	na
n_34024	188.9	1719	52.3	306	17.5	Phosphomethylpyrimidine kinase

n_34029	154.3	1596	50.8	346	21.2	Predicted hydrolase involved in interstrand crosslink repair
n_34100	127.6	1644	45.7	274	16.6	Endocytosis signaling protein EHD1
n_34206	103.8	600	63.7	123	20.1	Late embryogenesis abundant protein
n_34262	181.0	597	42.5	74	12.3	na
n_34341	118.4	1074	59.9	242	21.8	na
n_34416	153.1	543	44.6	88	16.2	BASIC HELIXLOOPHELIX LEUCINE ZIPPER
n_34473	122.2	1476	46.5	272	18.3	Peptidase family M28 and Transferrin receptorlike dimerisation domain
n_34636	160.4	1218	57.4	214	17.5	EamAlike transporter family
n_34730	106.5	903	50.4	151	16.5	Peroxisomal membrane protein MPV17 and related proteins
n_34751	115.0	2202	47.9	359	16.1	Betagalactosidase
n_34792	169.5	795	52.7	186	23.3	Protein of unknown function DUF688
n_34840	131.3	1200	44.6	190	15.8	na
n_34877	174.2	942	52	122	12.9	Protein kinase PCTAIRE and related kinases
n_35801	70.9	612	49	65	11.7	U4 U6 U5 tri snRNP component SNU23
n_39171	158.4	2076	45.7	487	23.2	Nucleic acid binding
n_40086	83.0	798	51.6	174	21.8	Predicted haloacid dehalogenaselike hydrolase
n_40152	125.3	1236	48.9	249	20.3	CENTROMEREBINDING PROTEIN 1 CBP1
n_40184	92.4	225	52.9	38	16.8	na
n_40263	133.3	999	49.5	162	15.8	Adenylylsulfate kinase
n_40267	120.0	1074	44.7	178	16.2	Tumor differentially expressed TDE protein
n_40333	90.9	561	45.1	68	12.0	Peroxisomal membrane protein MPV17 and related proteins
n_40488	77.3	216	35.6	49	21.9	na
n_40593	148.7	666	50.2	119	17.8	SECC motif
n_40617	215.2	1551	49.2	289	18.6	Predicted membrane protein
n_40685	119.4	4215	42.5	791	18.7	E3 ubiquitin ligase involved in syntaxin degradation
n_40707	107.9	858	57.2	149	17.3	Putative ABC transport system permease protein
n_40711	119.3	1488	60.8	260	17.2	Uncharacterized membrane protein predicted efflux pump
n_40715	117.8	2268	42	356	15.6	Uncharacterized conserved protein
n_40746	230.8	1191	40.6	183	15.3	na
n_40755	238.2	1035	46.6	136	13.1	Meiotic recombination protein Dmc1
n_40796	99.6	810	52.2	153	18.8	Protein of unknown function DUF3675 Zinc finger C3HC4 type

						RING finger
n_40821	115.3	1845	42.6	439	23.6	na
n_40832	145.6	819	41.8	136	16.5	na
n_40849	111.6	447	47.4	75	16.6	Uncharacterized conserved protein contains ML domain
n_40909	17.2	477	76.3	81	16.5	na
n_40919	142.0	1779	47.6	260	14.5	Zinc transporter and related ZIP domaincontaining proteins
n_40951	200.3	1749	45.2	305	17.3	Glucose6phosphate 1dehydrogenase
n_41002	147.2	1755	49	257	14.6	H+ oligopeptide symporter
n_41012	99.7	1842	44.1	274	14.8	longchain acylCoA synthetase
n_41115	126.1	2241	51.4	394	17.4	Transketolase
n_41116	203.1	921	49	133	14.4	Cobyrinic acid acdiamide synthase
n_41143	100.8	1641	52	389	23.3	Uncharacterized conserved protein
n_41171	109.1	618	50.5	67	10.8	na
n_41174	139.6	1611	45	295	18.2	PPR repeat
n_41186	122.5	1458	46.6	210	14.4	ZINC FINGER PROTEIN TRANSCRIPTION FACTOR
n_41200	88.9	981	47.1	163	16.5	RIBOSOMAL PSEUDOURIDINE SYNTHASE
n_41215	101.5	1410	45	156	11.0	Serine threonine protein kinase
n_41348	45.4	291	52.9	58	18.9	na
n_41364	47.0	1449	62.7	335	22.4	MEKK and related serine threonine protein kinases
n_41372	117.4	1380	48	162	11.7	Polypyrimidine tractbinding protein
n_41395	138.2	1038	48.2	192	18.4	Predicted kinase
n_41451	148.9	3306	45	547	16.6	Glycosyl hydrolases family 28 and PPR repeat
n_41455	127.5	1926	37.3	327	16.8	D3phosphoglycerate dehydrogenase
n_41455	65.2	252	56.9	26	10.1	na
n_41541	100.5	1323	55.6	226	16.7	Alcohol dehydrogenase class III
n_41625	90.0	396	50.3	137	33.5	na
n_41627	115.6	1506	46	187	12.6	PH domain
n_41647	108.1	999	60.5	181	17.6	G1 Sspecific cyclin D
n_41657	175.5	813	52.9	161	19.7	Fbox domain
n_41712	111.2	1575	48.3	222	14.0	Xanthine uracil transporters
n_41818	141.1	258	48.8	56	21.2	na
n_41838	65.8	396	48.7	67	16.8	Predicted Yippeetype zincbinding protein

n_41845	116.3	267	43.4	33	12.3	na
n_41892	93.2	312	44.2	58	17.2	na
n_41898	152.7	2766	43.9	474	17.0	Uncharacterized conserved protein HEN1 CORYMBOSA2
n_41911	98.2	1005	63.5	214	21.0	SMP30 Gluconolaconase LRElike region
n_41930	88.0	1971	46.4	247	12.8	Phosphatidylserine decarboxylase
n_41999	118.8	1254	52.6	301	23.9	Zinc finger C3HC4 type RING finger PROTEIN 6 12 38
n_42036	141.6	3711	44.3	596	16.1	Znfinger protein
n_42060	84.2	1152	56.9	331	28.5	Transferase family
n_42126	99.2	405	48.1	68	16.6	Snf7 protein transport
n_42166	147.3	1569	51.1	411	26.0	Molecular chaperones GRP78 BiP KAR2 HSP70 superfamily
n_42230	92.0	1098	60.5	196	17.6	Protein of unknown function DUF793
n_42256	120.4	1317	48.7	228	17.1	BetaTrCP transducin repeats containing Slimb proteins
n_42353	123.1	1419	57.6	259	19.3	Cysteine desulfurase NFS1
n_42393	44.7	639	72	131	19.7	Plant invertase pectin methylesterase inhibitor
n_42490	72.5	810	52.3	170	20.5	VAMPassociated protein involved in inositol metabolism
n_42511	235.3	1416	47.4	233	16.6	Cytochrome P450 CYP4 CYP19 CYP26 subfamilies
n_42561	88.3	1155	55.2	258	21.8	na
n_42576	98.2	1236	53.2	264	21.2	EamAlike transporter family
n_42584	38.2	1191	69.8	352	28.9	Kelch motif Fbox domain
n_42717	153.4	6555	42.5	859	13.0	DNA polymerase epsilon catalytic subunit A
n_42721	72.1	819	67.2	118	14.6	Leucine Rich Repeat protein binding
n_42723	193.7	1686	57.9	352	20.7	AcylCoA synthetase
n_42754	197.4	1533	48.9	324	20.9	Serine carboxypeptidases lysosomal cathepsin A
n_44831	117.2	1941	52.3	476	23.8	SWI SNF COMPLEX RELATED
n_49139	144.8	1752	48.9	324	18.3	PeptideOfucosyltransferase
n_BI096866	111.4	1416	60.5	169	11.9	Glycine serine hydroxymethyltransferase
n_BI096871	121.6	579	58.5	114	19.2	Ribulosebisphosphate carboxylase small chain
n_BI096903	94.1	582	51.2	118	19.6	Photosystem II 10kDa protein
n_BI096904	93.0	1209	59.8	172	14.1	Glyceraldehyde 3phosphate dehydrogenase
n_BI097026	126.1	798	48.6	124	15.5	na
n_dxs2	116.5	2190	52.9	410	18.5	Transketolase C terminal domain, pyrimidine binding domain, Dehydrogenase E1 component

n_hdr	158.1	1404	48.1	209	14.7	na
n_idi1	166.3	882	52.4	148	16.7	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase
n_idi2	149.6	1863	46	366	19.4	PPR repeat

Figures.

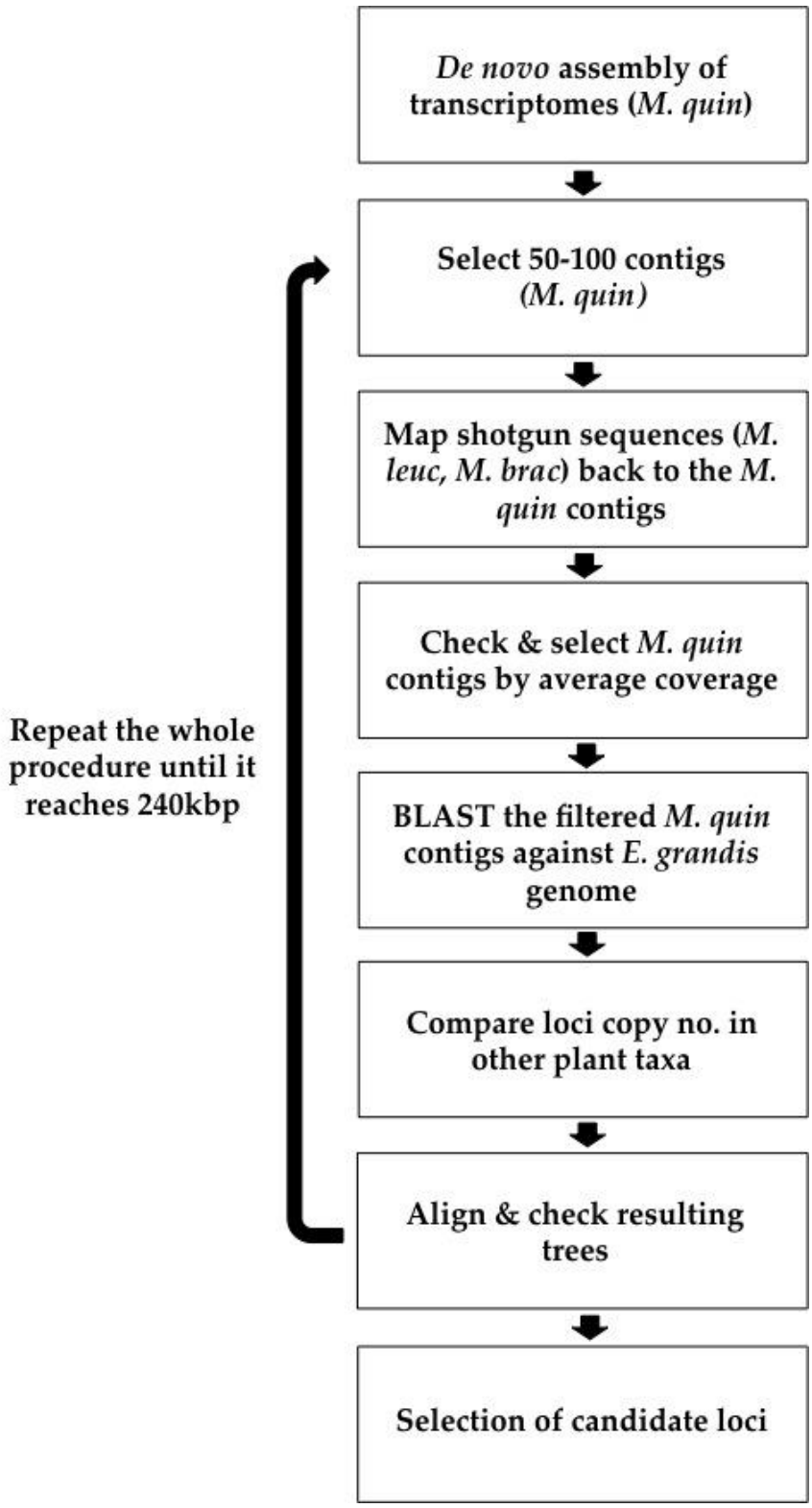

FIGURE 1. A workflow of the nuclear target loci identification. Species names are as follows: *M. brac* (*M. bracteata*), *M. leuc* (*M. leucadendra*), *M.quin* (*M. quinquenervia*).

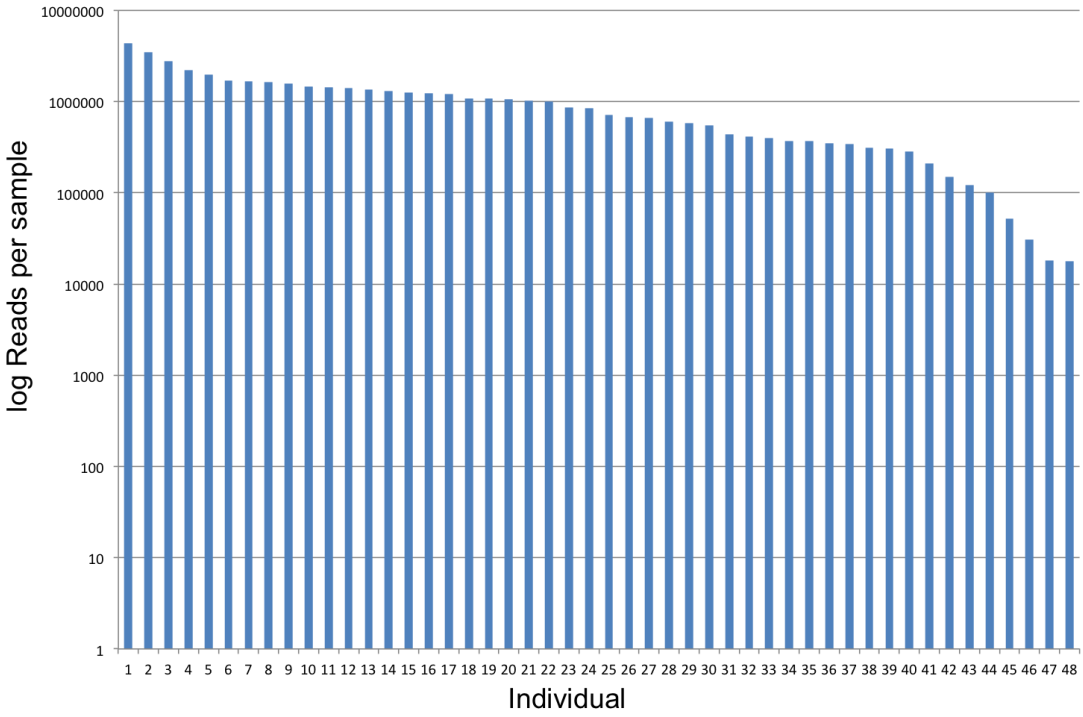

FIGURE 2. Number of reads mapped back to the *Eucalyptus grandis* targets, sorted by log reads.

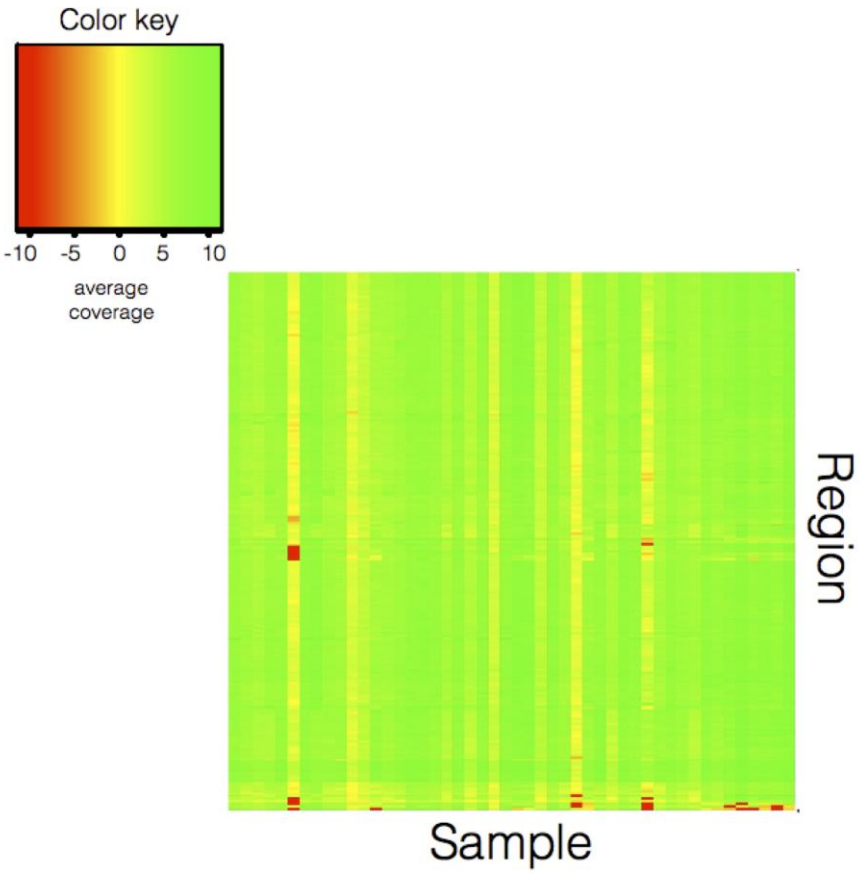

FIGURE 3. A heat map of average coverage of 209 loci of 47 samples. Samples are from left to right ordered serially 1-47 as in the 96-plate order. Y axis is the loci. Red represents the lowest coverage and green indicates the highest.

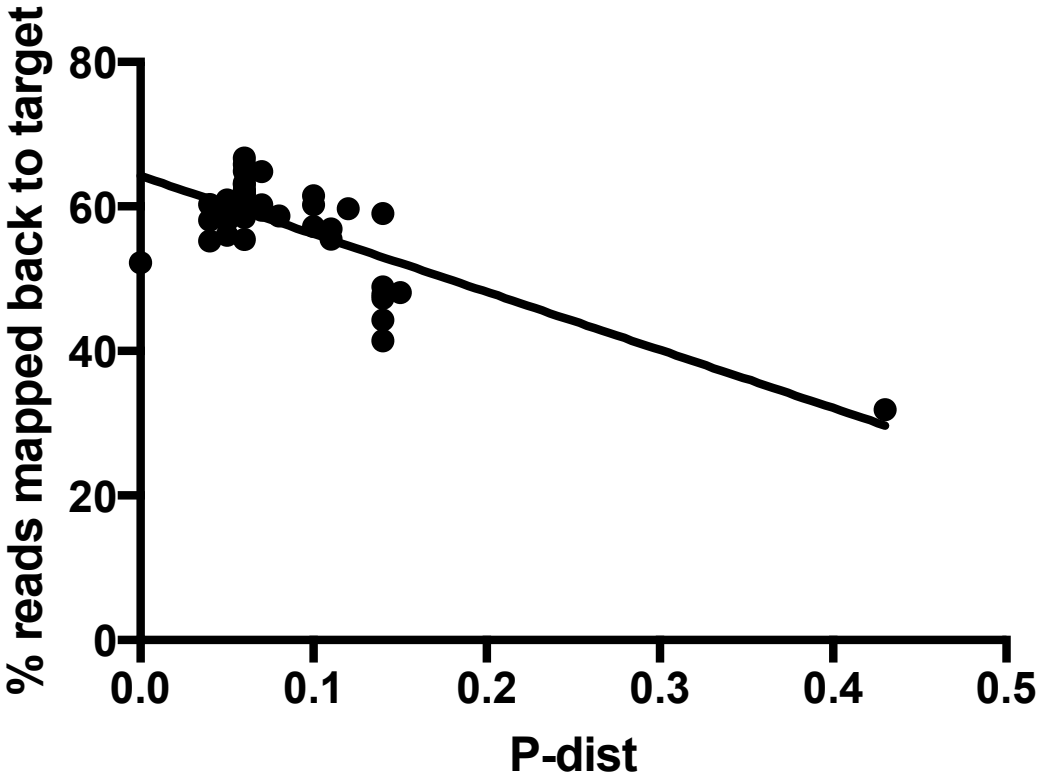

FIGURE 4. Linear regression of P-distance of each samples compared to *Eucalyptus globulus* versus percentage of reads mapped back to *Eucalyptus grandis* targets. The percenge of reads mapped and the P-distance is significantly negatively correlated (Rsquare=0.36, $P<0.0001$).

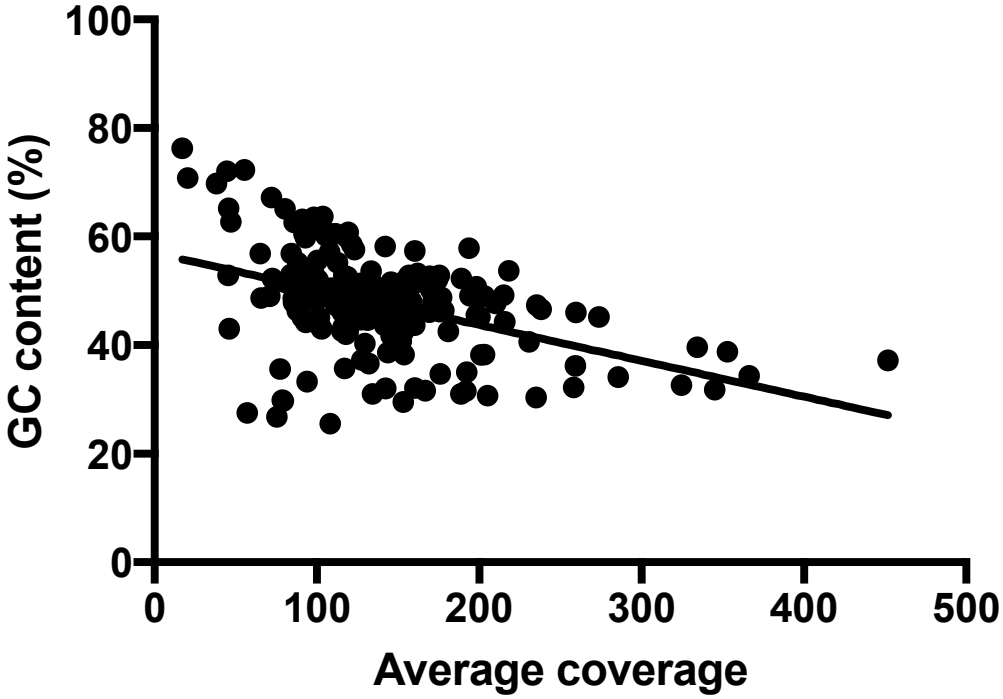

FIGURE 5. Linear regression of average coverage versus GC content. GC content and the average coverage are significantly negatively correlated (Rsqure = 0.2070, P<0.0001).

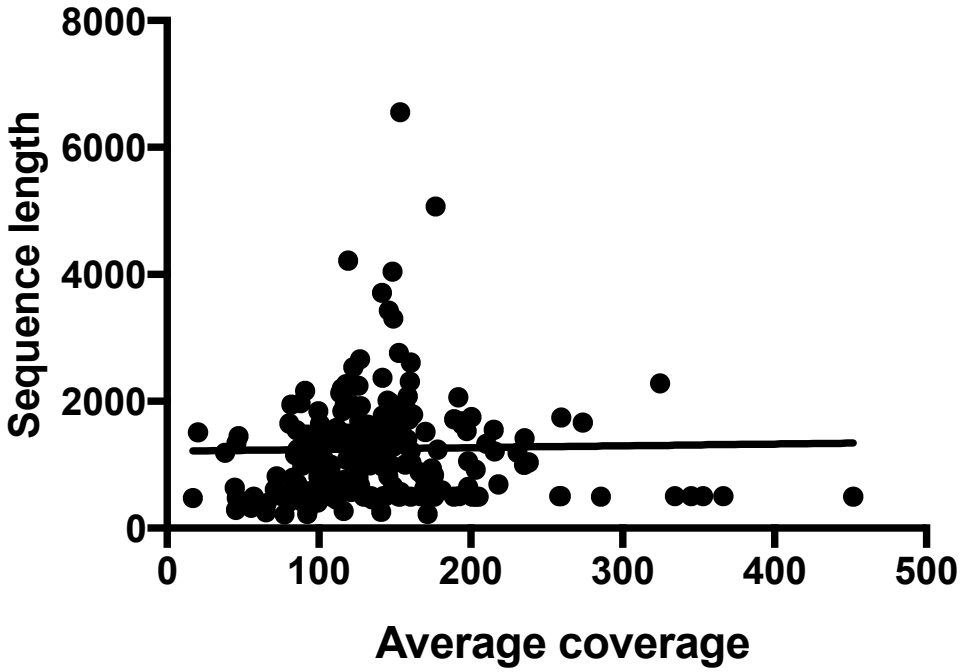

FIGURE 6. Linear regression of the target sequence length versus average coverage. The average and sequence length are not significantly correlated.

Supplementary Table S1. Average coverage of 47 Myrtaceous samples across 209 chloroplast and nuclear loci. Cells highlighted in light grey represent average coverage of 20 and dark grey cells indicate average coverage of 0. Note: In the brackets of the species colum indicate the collector and collection number for the investigated plant materials.

	Species	c_20012500	c_25013000	c_30013500	c_9220192700	c_35014000	c_1740117900	c_1980120300	c_2080121300	c_4510145600	c_4902049519	c_5020150700	c_5230152800	c_5550156000	c_6490165400	c_6760168100
1	***A. gummiferum* (McCoys.n.)**	243.66	231.28	174.7	297.1	314.75	145.74	261.61	483.45	182.13	204.91	261.34	512.9	176.25	196.13	367.38
2	***A. gummiferum* (McCoys.n.)**	93.82	96.97	77.45	184.74	136.19	81.76	132.33	221.6	94.55	82.09	109.74	235.23	75.57	71.33	177.36
3	***Corymbia bunites* (Cs229)**	382.33	448.48	325.62	511.45	265.62	305.98	457.5	751.56	377.49	381.62	463.31	753.99	380.15	353.6	563.96
4	***C. calophylla* (Cs377)**	3.66	5.83	6.25	6.2	8.06	5.5	6.73	21.69	9.29	9.28	7.16	15.9	8.41	10.17	10.22
5	***C. haematoxylon* (Cs380)**	174.54	143.49	160.18	291.95	132.09	155.37	249.21	380.67	168.07	137.6	277.63	382.55	148.36	106.46	307.18
6	***C. peltata* (Cs493)**	614.31	616.52	464.15	730.1	422.89	367.1	635.86	1036.76	461.8	554.93	589.89	1028.51	518.99	495.71	821.91
7	***E. brockwayi* (Cs474)**	287.36	260.67	331.47	536.01	607.7	233.97	340.29	628.25	278.35	262.35	370.06	533.6	315.15	208.74	425.79
8	***E. coccifera* (ANBG9404806)**	284.49	218.92	289.87	300.65	496.56	139.8	214.53	402.61	183.17	196.32	255.42	443.03	230.21	185.9	366.94
9	***E. extrica* (CANB632680)**	445.12	538.71	493.99	621.12	881.16	331.56	478.57	832.14	400.59	352.57	474.88	910.33	394.07	362.16	653.47
10	***E. globoidea* (Cs729)**	37.42	48.49	47.52	69.11	94.01	40.37	58.08	111.94	47.94	37.16	60.44	101.6	37.88	38.7	72.32
11	***E. globulus* (CANB688145)**	172.28	140.32	175.21	174.07	306.17	119.84	147.26	222.15	116.46	108.27	141.86	252.26	154.74	121.04	233.4
12	***E. globulus* (CANB494954)**	19.17	18.46	18.95	15.84	32.4	11.43	15.06	25.62	9.94	9.68	15.76	37.14	18.35	21.01	39.95
13	***E. gomphocephala* (Cs357)**	235.81	183.82	219.67	223.98	317.08	125.96	200.77	309.48	136.69	152.36	185.27	357.15	191.09	145.95	277.53
14	***E. goniantha* (Cs424)**	2.86	1.57	1.16	2.74	5.73	1.98	1.75	3	1.06	5.1	2.85	6.99	1.63	1.98	9.12
15	***E. howittiana* (Cs496)**	167.09	142.45	195.24	300.2	296.7	138.72	176.61	311.91	167.49	127.67	214.99	359.92	156.38	157.98	299.29
16	***E. insularis* (Cs275)**	97.54	105.19	105.63	138.04	191.75	64.41	105.44	182.51	83.13	115.76	86.04	186.8	112.27	60.63	129.44
17	***E. insularis* (CANB413807)**	28.89	30.31	25.76	40.29	47.72	19.82	22.96	36.61	19.79	18.02	25.49	51.81	21.76	18.6	47.64
18	***E. leucoxylon* (Cs716)**	181.86	154.77	181.99	191.67	314.32	93.7	153.3	268.55	99.16	147.35	150.05	339.75	194.82	165.08	199.1
19	***E. ligulata* (Cs721)**	21.2	14.44	20.97	25.32	33.03	26.24	19.77	39.29	28.86	17.92	19.02	37.18	20.09	16.46	20.55
20	***E. moorei* (Cs574)**	194.55	182.42	253.88	312.54	428.19	133.26	192.29	354.86	174.69	210.11	221.49	490.91	248.98	157.1	357.69
21	***E. nitida* (Cs744)**	112.18	109.84	124.01	142.7	233.69	70.99	98.81	199.74	70.32	106.35	122.24	221.67	121.69	62.28	152.18
22	***E. optima* (Cs708)**	43.11	45.58	50.93	75.04	96.95	39.82	58.28	93.23	47.96	50.08	58.2	100.13	58.72	43.55	66.29

23	***E. pachycalyx* (Cs305)**	17.2	17.54	15.97	22.25	32.97	15.89	15.35	37.91	13.63	15.87	16.42	26.91	10.37	19.55	22.59
24	***E. pachyloma* (Cs426)**	78.7	66.58	86.45	111.13	159.7	58.58	98.39	154.81	68.56	69.91	88.45	181.49	58.84	51.49	140.52
25	***E. perriniana* (Cs742)**	137.15	113.09	138.01	173.57	229.17	99.73	113.66	206.85	94.03	81.61	103.99	224.9	92.38	105.66	199.35
26	***E. pilularis* (Cs482)**	245.9	227.25	302.9	317.86	554.65	169.75	277.2	479.03	210.38	236.31	275.17	499.26	261.54	197.67	403.94
27	***E. platydisca* (CANB638520)**	195.28	175.06	253.19	321.05	420.58	151.31	224.87	411.68	183.02	210.56	276.71	485.04	183.58	178.64	340.13
28	***E. pleurocarpa* (Cs431)**	189.83	206.16	199.88	231.86	352.01	127.52	195.34	309.48	134.33	135.75	206.18	402.12	152.97	142.4	233.5
29	***E. preissiana* (Cs425)**	39.36	32.64	29.67	59.74	38.69	30.81	40.26	52.78	30.81	42.08	23.4	74.36	25.19	21.27	37.84
30	***E. pumila* (Cs227)**	282.77	217.76	249.68	265.43	473.45	154.37	207.85	336.73	198.72	191.55	214.64	446.74	183.55	183.46	306.05
31	***E. selachiana* (CANB632673)**	148.83	109.52	136.4	233.06	272.95	129.18	206.22	255.82	139.24	117.01	156.5	340.03	121.47	117.19	235.16
32	***E. selachiana* (CANB693174)**	178.72	159.72	197.48	231.11	346.76	124.1	178.46	319.63	135.11	153.99	187.26	384.32	151.82	132.76	225.2
33	***E. spathulata* (Cs368)**	167.79	148.79	176.96	163.9	211.88	72.74	115.98	190.64	85.12	123.16	124.65	249.8	143.61	100.53	202.73
34	***E. staeri* (Cs366)**	18.26	14.36	13.66	26.23	36.74	10.74	22.77	44.63	21.69	14.74	29.44	33.19	17.02	12.7	26.03
35	***E. stenostoma* (Cs733)**	4.01	6.18	3.82	10.22	10.33	2.65	2.25	7.55	5.42	3.19	4.93	15.49	3.08	2.33	5.52
36	***E. stoatei* (Cs191)**	315.67	260.94	306.84	316.61	436.82	188.13	240.54	473.54	211.65	206.01	294.19	481.6	244.4	177.51	353.25
37	***E. tereticornis* (Cs769)**	134.59	108.39	132.41	139.29	205.82	86.3	98.12	205.76	89.12	88.7	113.95	187.44	88.6	87.76	142.68
38	***E. verrucata* (Cs715)**	90.64	108.12	107.92	151.65	185.91	76.47	100.92	175.24	75.26	70.93	117.94	186.41	85.41	58.45	155.69
39	***E. virginea* (Cs727)**	28.86	33.37	31.87	45.47	52.38	36.3	34.34	66.97	29.12	37.15	43.01	59.54	24.64	20.97	42.31
40	***E. willisii* (Cs492)**	120.86	107.89	138.58	152.33	263.12	67.21	124.63	168.16	89.06	101.98	114.39	233.66	111.97	95.05	182.96
41	***Heteropyxis natalensis* (CANB576168)**	73.01	55.06	72.41	66.59	112.53	49.03	62.31	97.6	50.57	47.02	76.42	80.64	27.96	55.03	104.34
42	***Melaleuca capitata* (RDE176)**	878.91	1150.87	1990.01	1499.02	2306.12	687.39	1185.99	1673.75	735.39	1068.9	1089.79	2007.25	570.33	1203.24	1649.92
43	***M. cornucopiae* (Harwood1546)**	607.59	638.06	1380.84	915.77	1567.33	369.97	633.71	969.11	476.82	734.82	530.86	1234.87	522.81	644.15	976.22
44	***M. foliolosa* (AF4308)**	309.76	311.58	482.26	351.2	518.87	183.43	324.19	520.92	206.11	252.94	289.48	386.8	192.6	390.8	453.78
45	***M. squarrosa* (RDE75)**	159.62	288.52	637.22	389.84	590.96	180.88	312.93	499.12	230.16	254.65	314.06	503.53	216.88	273.57	457.83
46	***M. sylvana* (AF4286)**	195.3	249.67	355.2	254.68	422.45	106.47	235.73	421.61	173.48	158.09	219.23	353.94	115.95	146.92	257.05
47	***Calothamnus gracilis* (RDE154)**	408.27	497.15	945.58	543.89	752.44	254.59	477.22	713.51	299.66	566.95	422.41	772.78	351.94	426.64	671.03

Supplementary Table S1. (continued).

	Species	c_6810168600	c_6919169690	c_7290173400	c_8330183800	c_9002690525	c_150761151760	c_152701153200	c_154601155100	c_157101157600	c_158701159200	c_164501165000	c_165301165800	c_166301166800	c_167001167500	c_ndhF
1	***A. gummiferum* (McCoys.n.)**	155.51	174.4	482.43	96.32	403.26	266.01	86.73	102.38	152.81	173.17	179	258.98	82.02	95.35	205.69
2	***A. gummiferum* (McCoys.n.)**	79.32	80.15	290.46	40.62	194.07	116.1	25.35	41	96.58	65.19	70.8	102.72	22.65	28.07	86.3
3	***Corymbia bunites* (Cs229)**	298.25	251.07	1073.65	161.19	676.51	397.18	177.57	188.34	288.32	312.24	219.53	364.57	95.78	142.69	385.03
4	***C. calophylla* (Cs377)**	10.65	3.32	14.63	4.67	15.05	9.72	3.06	4.6	2.04	5.7	3.84	6.08	3.66	1.1	4.33
5	***C. haematoxylon* (Cs380)**	136.21	153.14	540.2	58.99	418.17	189.77	62.26	66.73	129.01	156.58	112.87	192.44	28.2	71.73	170.7
6	***C. peltata* (Cs493)**	390.55	379.25	1306.69	236.5	945.17	622.9	273.89	342.01	410.93	458.22	340.54	554.94	114.8	218.93	544.03
7	***E. brockwayi* (Cs474)**	235.46	256.93	930.82	123.97	680.84	226.45	119.68	153.13	264.79	259.73	267.72	245.68	74.21	126.53	339.77
8	***E. coccifera* (ANBG9404806)**	168.64	204.81	534.77	100.57	425.43	210.25	68.69	113.65	185.22	145.71	207.73	219.65	66.06	82.62	247.16
9	***E. extrica* (CANB632680)**	329.27	352.34	1124.99	117.38	819.42	304.82	147.99	155.06	398.15	415.03	276.17	440.16	97.13	161.09	443.65
10	***E. globoidea* (Cs729)**	45.94	35.36	160.94	14.36	145.47	35.82	9.12	20.61	38.14	38.04	38.76	27.69	8.11	12.84	50.91
11	***E. globulus* (CANB688145)**	105.95	127.76	305.85	69.19	251.63	127.89	50.32	55.62	117.06	107.52	131.76	144.67	40.49	73.22	142.79
12	***E. globulus* (CANB494954)**	12.19	20.48	26.41	7.86	43.02	18.02	4.31	5.27	10.13	17.01	11.93	22.76	6.86	8.44	23.15
13	***E. gomphocephala* (Cs357)**	124.23	170.65	432.19	70.68	379.97	171.15	62.78	94.46	144.03	112.91	139.77	217.65	50.87	79.75	215.45
14	***E. goniantha* (Cs424)**	1.71	0.11	8.91	0.29	9.46	3.22	2.78	4.02	4.55	6.33	2.61	1.82	0	0.84	4.75
15	***E. howittiana* (Cs496)**	129.76	145.01	487.55	59.24	403.79	141.17	56.15	76.31	178.2	131.94	145.2	234.22	41.57	67.06	193.89
16	***E. insularis* (Cs275)**	60.15	74.47	256.91	33.67	205.09	87.73	32.27	50.51	84.51	94.21	61.41	88.95	22.88	42.88	106.28
17	***E. insularis* (CANB413807)**	16.81	14.57	53.83	12.39	44.16	18.01	6.16	7.19	17.66	20.7	15.46	25.77	4.73	10.19	24.95
18	***E. leucoxylon* (Cs716)**	90.64	116.28	401.68	77.48	298.04	130.58	60.34	89.16	153.27	133.93	131.56	167.69	53.44	69.25	186.38
19	***E. ligulata* (Cs721)**	10.1	10.7	36.08	6.07	54.85	14.12	5.32	5.97	15.85	15.43	10.26	14.2	9.57	10.64	17.66
20	***E. moorei* (Cs574)**	136.92	147.42	540.98	89	431.98	209.45	85.03	109.57	194.91	147.04	154.01	133.52	62.63	82.43	254.72
21	***E. nitida* (Cs744)**	42.65	78.68	235.45	51.06	220.38	93.64	40.83	52.98	99.61	63.12	63.33	106.25	32.92	41.39	131.67
22	***E. optima* (Cs708)**	38.06	42.92	130.09	10.34	121.7	41.74	15.27	27.1	39.51	43.28	33.99	31.62	11.72	24.2	50.97
23	***E. pachycalyx* (Cs305)**	7.44	9.06	48.3	5.98	40.55	13.32	3.69	8.38	12.95	12.2	10.38	12.31	5.46	11.53	17.29
24	***E. pachyloma* (Cs426)**	75.39	70	220.81	28.05	167.48	64.75	19.4	34.83	79.09	59.8	45.76	82.13	13.86	29.01	90.98
25	***E. perriniana* (Cs742)**	83.37	102.32	313.64	37.65	281.6	103.97	33.71	50.85	95.68	82.87	105.53	110.46	24.24	54.56	110.23

26	***E. pilularis* (Cs482)**	181.73	193.67	672.51	120	508.24	216.28	83.49	144.37	217.46	177.06	158.86	179.06	64.15	89.63	288.32
27	***E. platydisca* (CANB638520)**	149.93	177.51	554.49	90.96	413.94	211.01	74.58	87.23	182.18	202.76	146.77	196.24	84.61	91.32	272.91
28	***E. pleurocarpa* (Cs431)**	131.62	122.9	377.13	69.62	273.51	142.59	50.52	74.71	129.49	132.29	111.85	179.79	39.24	75.77	194.64
29	***E. preissiana* (Cs425)**	27.5	35.01	82.88	8.41	107.75	22.97	8.42	11.22	33.39	20.98	27.96	26.63	3.32	11.27	30.39
30	***E. pumila* (Cs227)**	163.72	166.04	514.66	78.17	457.21	185.13	54.87	108.89	192.36	149.13	180.72	192.72	55.65	88.19	212.1
31	***E. selachiana* (CANB632673)**	101.31	123.63	301.42	34.25	311.09	92.01	23.26	43.47	107.07	102.71	64.97	145.1	23.63	54.36	134.16
32	***E. selachiana* (CANB693174)**	119.89	117.1	392.72	83.35	302.81	122.39	50.77	94.81	125.97	119.33	146.76	169.21	46.32	63.8	164.75
33	***E. spathulata* (Cs368)**	78.58	114.72	264.17	61.74	209.15	106.67	47.38	81.35	106.41	98.13	116.13	185.03	48.31	72.81	161.03
34	***E. staeri* (Cs366)**	9.36	13.7	40.81	8.39	46.14	17.92	4.1	10.94	9.8	17.11	10.67	15.37	3.2	6.05	19.71
35	***E. stenostoma* (Cs733)**	6.86	4.98	11.13	3.35	11.35	3.41	0.94	2.46	4.37	4.61	4.54	2.11	0.69	2.12	4.32
36	***E. stoatei* (Cs191)**	195.86	214.51	757.61	102.81	508.24	211.26	81.43	119.74	211.14	170.48	178.98	248.27	64.42	108.2	281.93
37	***E. tereticornis* (Cs769)**	62.61	72.95	223.2	35.44	200	94.65	27	44.91	83.93	80.21	81.44	101.47	23.13	33.36	101.41
38	***E. verrucata* (Cs715)**	86.64	60.15	269.64	30.4	185.52	74.22	32.6	34.58	83.42	73.39	51.4	76.17	23.31	34.32	106.51
39	***E. virginea* (Cs727)**	26.03	19.74	80.39	6.14	79.83	24.65	9.14	7.11	35.45	23.58	20.59	29.6	5.1	20.05	31.85
40	***E. willisii* (Cs492)**	82.23	96.29	317.8	53.53	200.56	120.3	46.92	56.19	89.22	87.55	91.43	132.61	37.91	62.36	137.81
41	***Heteropyxis natalensis* (CANB576168)**	38.44	34.85	118.44	20.26	115.9	65.55	10	35.13	56.12	66.39	52.44	76.5	11.58	32.91	57.72
42	***Melaleuca capitata* (RDE176)**	1075.15	766.97	2469.72	577.52	1904.5	2407	283.88	511.33	536.83	689.88	1068.35	1517.24	1575.92	487.31	1174.09
43	***M. cornucopiae* (Harwood1546)**	430.3	544.59	1113.83	274.74	938.89	1366.55	129.16	456.85	243.87	370.56	680.33	994.05	710.65	250.79	605.66
44	***M. foliolosa* (AF4308)**	264.86	160.22	565.25	117.99	472.22	395.13	31	132.74	118.02	150.02	230.81	287.91	315.16	120.28	229.54
45	***M. squarrosa* (RDE75)**	236.74	201.02	701.49	124.83	483.21	498.86	69.03	109.84	153.92	139.88	289.08	352.92	293.24	113.78	231.69
46	***M. sylvana* (AF4286)**	186.61	136.4	505.94	97.21	420.48	337.36	20.14	90.62	76.52	101.66	162.64	205.28	209.44	56.98	151.39
47	***Calothamnus gracilis* (RDE154)**	311.43	289.55	946.49	191.58	760.89	714.54	93.2	297.93	199.83	252.76	544.85	522.39	476.45	121.95	376.93

Supplementary Table S1. (continued).

	Species	c_psbA-trnH	c_trnL-trnF	c_matK-trnK	n_39171	n_44831	n_30177	n_30296	n_30355	n_30410	n_30431	n_30519	n_30906	n_31003	n_31220	n_31275	n_31348
1	***A. gummiferum* (McCoys.n.)**	60.88	145.58	414.96	262.29	190.95	234	117.61	274.34	201.52	278.78	201.25	232.87	189.92	292.25	338.13	156.94
2	***A. gummiferum* (McCoys.n.)**	25.42	58.72	173.45	163.7	90.83	141.54	58.44	144.84	100.95	152.11	101.08	123.74	100.13	154.03	191.65	88.52
3	***Corymbia bunites* (Cs229)**	99.44	275.38	717.94	399.23	381.91	431.81	167.22	647.73	306.96	257.34	388.1	398.87	347.38	479.04	500.36	300.36
4	***C. calophylla* (Cs377)**	2.77	1.98	15.68	2.14	2.96	3.68	2.74	3.3	1.34	2.1	2.57	1.42	0.99	2.04	1.26	1.24
5	***C. haematoxylon* (Cs380)**	24.14	131.16	313.54	53.87	52.22	43.84	30.14	84.41	34.74	31.74	52.99	53.58	50.76	54.48	65.26	43.65
6	***C. peltata* (Cs493)**	125.91	358.87	1011.14	287.51	261.73	287.37	155.92	442.88	228.05	212.72	302.14	288.02	262.75	361.87	328.94	258.92
7	***E. brockwayi* (Cs474)**	147.51	216.82	468.54	197.14	204.88	165.4	137.01	262.7	142.85	112.08	178.43	158.49	173.31	248.94	227.86	118.9
8	***E. coccifera* (ANBG9404806)**	102.41	138.09	337.07	211.4	153.05	163.92	155.22	218.9	152.31	110.54	261.85	150.16	160.5	155.03	253.73	137.24
9	***E. extrica* (CANB632680)**	152.53	296.09	691.34	157.71	129.29	132.9	74.93	210.62	163.64	95.88	126.55	137.6	175.87	251.81	156.79	98.12
10	***E. globoidea* (Cs729)**	20.72	32.52	78	131.22	106.06	96.12	95.22	179.83	89.41	67.2	86.25	95.44	84.02	169.19	153.91	81
11	***E. globulus* (CANB688145)**	65.38	80.21	210.47	170.64	137.96	99.62	89.61	185.26	136.79	101.8	139.6	129.68	164.75	165.18	175.24	111.67
12	***E. globulus* (CANB494954)**	10.56	10.54	26.33	43.45	38.24	42.74	19.4	41.89	49.08	30.81	48.38	31.3	34.63	53.51	50.83	31.87
13	***E. gomphocephala* (Cs357)**	85.1	135.24	284.06	198.19	121.09	133.48	90.32	257.54	144.28	110.72	139.6	115.58	159.15	208.06	205.99	133.64
14	***E. goniantha* (Cs424)**	3.88	2.89	4.65	2.65	2.02	1.07	1.03	4.01	0.42	2.32	1.09	4.22	1.36	8.73	2.89	1.63
15	***E. howittiana* (Cs496)**	52.13	113.92	255.06	129.16	93.77	90.34	80.09	155.33	84.51	63	91.61	99.16	95.93	118.55	136.37	76.44
16	***E. insularis* (Cs275)**	37.05	56.94	152.57	122.75	72.65	77.29	65.97	136.83	78.13	68.07	70.92	81.27	71.09	211.13	109.89	68.8
17	***E. insularis* (CANB413807)**	8.68	20.33	41.96	82.62	44.81	52.79	40.25	87.94	47.97	54.7	48.77	54.22	49.18	148.48	72.73	34.88
18	***E. leucoxylon* (Cs716)**	50.47	110.03	226.56	154.81	101.67	102.09	77.94	155.36	88.86	90.84	106.66	82.12	113.23	142.88	171.44	92.09
19	***E. ligulata* (Cs721)**	12.61	14.87	28.74	18.75	14.19	15.59	15.85	19.41	14.04	12.63	10.94	13.88	10.72	22.78	15.52	13.88
20	***E. moorei* (Cs574)**	97.4	158.46	333.57	231.25	186	168.85	118.72	220.23	151.35	161.8	173.92	190.85	184.88	325.88	258	151.32
21	***E. nitida* (Cs744)**	40.79	72.01	175.37	93.16	75.18	72.96	48.36	96.39	66.31	66.59	66.11	75.91	73.32	132.27	95.84	56.45
22	***E. optima* (Cs708)**	20.92	34.73	72.28	53.06	34.3	37.17	38.83	49.62	19.17	21.7	39.56	37.61	33.98	64.77	49.63	24.61
23	***E. pachycalyx* (Cs305)**	7.04	7.76	23.66	8.03	7.97	5.9	5.03	9.68	3.93	3.36	5	7.48	8.34	11.54	8.54	4.22
24	***E. pachyloma* (Cs426)**	35.42	62.72	129.81	75.5	51.72	51.68	40.4	89.52	47.96	37.92	47.38	48.94	59.44	210.05	74.2	49.93
25	***E. perriniana* (Cs742)**	39.38	70.31	160.76	65.14	62.36	38.87	34.76	63.89	30.69	43.06	55.06	48.55	57.69	50.94	63.5	34.43
26	***E. pilularis* (Cs482)**	112.91	169.6	420.53	246.95	210.28	157.08	150.31	266.9	197.38	149.33	188.73	173.66	186.38	344.54	299.18	162.67

27	***E. platydisca* (CANB638520)**	101.51	141.72	349.77	270.31	189.48	183.89	122.14	307.73	206.13	186.98	196.59	236.53	192.31	571.11	276.19	149
28	***E. pleurocarpa* (Cs431)**	76	123.49	296.42	120.06	73.56	103.15	57.14	124.96	118.61	72.91	73.53	77.3	110.96	162.54	134.82	60.23
29	***E. preissiana* (Cs425)**	12.07	30.23	46.07	19.34	21.71	18.91	13.44	17.55	11.36	11.8	13.69	11.36	22.26	20.45	16.84	12.58
30	***E. pumila* (Cs227)**	75.81	137.21	309.69	247.07	161.57	143.42	144.65	259.16	167.85	135.43	200.29	176.33	194.52	327.84	263.5	172.34
31	***E. selachiana* (CANB632673)**	56.66	107.35	211.99	54.6	45.44	28.66	21.76	73.17	57.69	28.21	53.07	60.82	41.41	93.6	63.3	21.79
32	***E. selachiana* (CANB693174)**	65.01	107.77	268.87	189.84	129.57	120.97	101.47	221.28	182.74	100.8	140.25	123.62	136.17	302.01	190.82	115.52
33	***E. spathulata* (Cs368)**	59.71	85.24	207.61	190.89	145.31	149.61	94.29	208.4	125.48	111.56	152.03	149.86	168.67	238.18	207.5	112.53
34	***E. staeri* (Cs366)**	8.88	12.75	24.08	41	33.95	28.28	27.44	34.01	32.43	20.87	23.67	30.31	45.01	75.68	31.74	19.63
35	***E. stenostoma* (Cs733)**	2.45	4.04	7.03	4.32	5.95	6.35	3.67	8.08	2.98	3.69	4.73	1.73	1.94	9.88	5.12	5.2
36	***E. stoatei* (Cs191)**	130.25	159.98	388.98	220.56	147.67	170.94	145.84	225.62	199.55	109.92	163.06	147.13	155.32	309.95	201.75	120.04
37	***E. tereticornis* (Cs769)**	54.36	70.14	150.65	335.53	249.82	248.11	211.89	372.94	239.01	198.73	300.99	207.83	269.2	279.57	354.78	209.09
38	***E. verrucata* (Cs715)**	51.35	59.47	154.84	77.91	59.19	59.43	40.2	96.47	39.71	45.65	48.93	47.2	55.97	114.43	89.38	54.33
39	***E. virginea* (Cs727)**	15.06	27.27	54.77	61.34	59.45	51.67	45.73	64.35	64.6	47.47	44.87	41.72	40.55	112.93	71.96	29.29
40	***E. willisii* (Cs492)**	55.78	82.88	188.83	57.55	46.96	47.29	31.72	68.88	47.16	41.79	49.17	52.6	55.19	92.13	45.15	29.96
41	***Heteropyxis natalensis* (CANB576168)**	12.28	33.72	112.19	2.86	1.42	3.57	1.85	6.13	13.7	5.83	2.96	7.4	6.7	18.03	0.77	4.76
42	***Melaleuca capitata* (RDE176)**	410.14	592.97	2178.62	376	362.98	433.11	212.85	633.09	320.95	208.71	264.65	344.97	342.92	445.88	156.19	304.88
43	***M. cornucopiae* (Harwood1546)**	711.98	359.95	1066.82	609.29	373.28	598.03	230.03	893.16	277.45	276.78	378.35	569.9	506.53	624.18	201.55	401.19
44	***M. foliolosa* (AF4308)**	62.2	127.23	545.94	188.04	92.11	228.96	82.83	324.49	106.1	135.44	136.45	167.75	160.54	210.23	54.32	129.08
45	***M. squarrosa* (RDE75)**	108.19	164.34	587.21	237.79	186.34	303.36	126.43	427.89	213.13	111.38	184.87	243.5	219.23	324.82	97.87	186.97
46	***M. sylvana* (AF4286)**	54.94	87.9	437.78	194.59	190.11	297.08	88.24	398.36	133.34	80.49	158.38	210.86	231.46	248.07	85.12	155.26
47	***Calothamnus gracilis* (RDE154)**	157.26	215.53	899.04	385.67	105.56	198.64	69.41	245.93	142.74	70.34	122.87	123.45	145.22	165.93	47.6	146.72

Supplementary Table S1. (continued).

	Species	n_31545	n_31598	n_31662	n_32080	n_32112	n_32478	n_32479	n_32563	n_32592	n_32660	n_32747	n_32861	n_32864	n_32883	n_33080	n_33128
1	***A. gummiferum* (McCoys.n.)**	379.19	256.42	269.57	154.41	211.7	250.26	136.23	331.48	235.17	65.63	230.02	158.73	189.36	26.55	239.26	242.88
2	***A. gummiferum* (McCoys.n.)**	250.89	148.05	174.68	85.16	124.88	138.2	68.49	198.18	141.91	36.28	140.62	89.2	112.41	18.35	139.87	135.78
3	***Corymbia bunites* (Cs229)**	826.03	483.93	535.94	263.91	486.99	428.16	402.49	713.43	360.95	56.39	467.66	303.77	293.35	51.45	516.97	465.18
4	***C. calophylla* (Cs377)**	5.17	2.61	4.59	0.91	2.86	5.88	1.03	4.27	3.88	1.18	2.73	1.6	0.4	0.66	3.32	1.98
5	***C. haematoxylon* (Cs380)**	90.59	73.59	62.89	33.04	45.57	70.06	48.84	87.36	56.89	6.28	50.05	33.14	27.98	6.16	53.38	59.42
6	***C. peltata* (Cs493)**	482.33	358.75	404.6	172.34	331.89	230.12	481.97	517.97	297.22	42.45	334.11	231.93	185.33	35.52	346.79	376.3
7	***E. brockwayi* (Cs474)**	330.8	205.08	189.1	108.12	236.64	493.23	176.24	246.52	191.96	87.69	249.03	260.67	177.71	26.66	170.73	188.93
8	***E. coccifera* (ANBG9404806)**	222.79	199.77	180.96	107.26	225.51	292.86	208.24	242.35	185.38	79.58	146.28	137.16	163.63	19.32	153.62	208.92
9	***E. extrica* (CANB632680)**	146.26	157.78	162.09	80.19	119.61	123.39	135.96	218.69	135.16	40.76	185.56	109.58	149.74	7.14	167.59	119.35
10	***E. globoidea* (Cs729)**	202.4	102.34	136.77	68.56	93.94	124.25	92.05	166.98	124.49	54.1	145.65	80.86	101.44	23.26	108.31	168.37
11	***E. globulus* (CANB688145)**	184.72	171.25	150.35	99	216.17	246.02	222.95	212.59	139.13	54.59	128.92	130.37	113.09	11.17	139.76	121.78
12	***E. globulus* (CANB494954)**	58.92	51.44	58.39	25.85	42.78	67.32	42.31	66.23	34.31	12.19	41.54	30.49	33.74	4.72	44.01	44.99
13	***E. gomphocephala* (Cs357)**	247.4	160.82	201.54	93.94	125.95	255.75	140.11	245.61	138.94	40.94	221.06	135.91	140.16	11.42	140.72	139.77
14	***E. goniantha* (Cs424)**	8.02	2.63	3.14	0.57	2.32	1.6	2.1	2.71	2.01	1.66	1.79	2.19	2.35	0.33	2.78	2.98
15	***E. howittiana* (Cs496)**	154.48	114.43	125.95	63.6	150.27	231.33	94.5	149.58	94.54	44.7	98.85	86.85	95.52	16.51	93.4	102.14
16	***E. insularis* (Cs275)**	95.92	96.87	109.97	41.13	60.16	84.17	69.09	147.85	73.1	27.71	85.05	82.6	83.95	7.62	70.72	263.72
17	***E. insularis* (CANB413807)**	97.44	63.68	74.62	35.5	53.23	60.44	43.68	80.86	61.67	19.73	60.73	61.04	45.49	4.03	54.34	176.87
18	***E. leucoxylon* (Cs716)**	159.35	132.38	163.94	79.95	133.95	245.07	116.57	215.05	122.24	40.7	109.66	108.09	108.16	13.79	235.52	108.87
19	***E. ligulata* (Cs721)**	23.51	16.19	28.04	7.24	13.7	21.02	18.06	15.94	18.88	6.37	18.89	11.72	12.33	1.96	11.56	16.47
20	***E. moorei* (Cs574)**	278.65	206.31	194.47	125.06	180.98	170.86	174.11	264.34	156.83	61.64	192.8	179.72	169.15	27.18	187.32	331.05
21	***E. nitida* (Cs744)**	109.44	82.04	95.65	59.1	85.87	68.74	77.53	93.8	88.46	22.83	74.01	63.74	56.45	7.24	89.32	136.09
22	***E. optima* (Cs708)**	66.7	45.72	66.16	24.63	35.55	110.39	36.31	65.61	39.15	19.69	55.64	29.01	37.15	2.84	40.32	39.5
23	***E. pachycalyx* (Cs305)**	11.01	10.04	9.58	4.95	6.55	9.32	6.74	6.57	7.7	1.24	6.09	4.05	5.9	2.41	8.68	7.04
24	***E. pachyloma* (Cs426)**	89.18	63.03	71.61	39.39	59.37	60.1	51.16	79.56	50.82	24.61	68.53	46.63	58.06	5.91	43.58	110.67
25	***E. perriniana* (Cs742)**	85.87	54.66	61	37.92	58.8	102.6	53.04	75.98	52.11	22.35	62.4	37.61	70.35	7.54	46.77	63.15
26	***E. pilularis* (Cs482)**	326.4	205.65	230.8	119.83	179.43	200.5	174.97	280.35	181.57	69.99	275.9	176.86	183.2	30.96	184.78	429.15

27	***E. platydisca* (CANB638520)**	282.37	246.88	234.32	148.5	204.97	211.06	198.58	325.76	185.09	58.55	223.61	197.5	215.04	19.11	227.96	289.5
28	***E. pleurocarpa* (Cs431)**	141.54	106.73	94.4	58.72	58.52	77.03	82.56	122.7	107.16	25.87	79.64	60.3	94.07	8.16	86.42	80.42
29	***E. preissiana* (Cs425)**	26.52	19.82	25.59	7.71	10.59	32.64	16.24	28.64	17.65	5.75	13.92	10.19	15.99	2.78	29.29	15.57
30	***E. pumila* (Cs227)**	255.71	219.06	232.85	122.22	165.49	625.76	207.39	292.23	190.52	63.67	195.12	166.39	188.81	14.78	180.46	246.54
31	***E. selachiana* (CANB632673)**	46.94	54.82	57.82	24.7	31.64	33.31	71.15	77.74	38.18	6.89	37.12	29	45.2	3.28	52.75	42.77
32	***E. selachiana* (CANB693174)**	212.16	174.03	155.36	115.2	115.16	149.15	270.35	259.92	158.06	64.41	233.73	121.29	156.53	19.8	168.46	139.04
33	***E. spathulata* (Cs368)**	209.72	155.66	145.34	107.58	179.87	189.44	140.49	247.42	153.75	54.45	149.23	131.21	147.54	19.44	298.18	138.31
34	***E. staeri* (Cs366)**	44.38	33.07	30.25	21.04	18.79	25.47	23.22	49.45	22.87	12.68	32	30.31	28.06	2.9	28.98	57.68
35	***E. stenostoma* (Cs733)**	10.22	5.52	5.71	5	3.5	3.55	4.27	7.24	4.92	1.3	5.04	6.96	6.64	0.75	4.46	12.45
36	***E. stoatei* (Cs191)**	272.55	192.62	216.65	109.16	155.79	350.97	160.26	246.9	163.56	84.83	263.92	145.79	130.87	21.44	310.13	150.34
37	***E. tereticornis* (Cs769)**	471.82	290.59	306.21	176.63	353.68	275.64	239.55	392.73	246.39	107.37	237.27	220.92	296.19	32.2	267.65	311.09
38	***E. verrucata* (Cs715)**	91.15	69.35	64.74	38.12	44.45	63.81	59.76	89.74	53.13	16.46	71.08	43.83	52.26	4.28	52.58	166.1
39	***E. virginea* (Cs727)**	92.87	54.97	84.22	30.69	41.43	50.23	52.27	69.06	55.76	19.05	54.68	46.73	51.51	7.54	48.63	160.96
40	***E. willisii* (Cs492)**	75.55	59.5	42.8	40.9	51.03	50.32	62.07	62.61	33.98	19.78	57.75	43.54	61.66	5	58.4	71.21
41	***Heteropyxis natalensis* (CANB576168)**	15.67	11.35	2.53	2.09	2.65	2.77	4.68	9.05	2.36	1.82	9.73	3.61	2.59	2.4	9.08	10.06
42	***Melaleuca capitata* (RDE176)**	853.28	392.92	489.51	278.03	292.23	282.59	573.43	556.16	468.69	179.33	442.54	205.98	407.38	114.8	396.9	319.34
43	***M. cornucopiae* (Harwood1546)**	677.87	572.99	626.51	303.63	390.13	369.51	380.32	687.99	721.43	174.18	510.84	300.62	523.79	87.76	514.91	520.11
44	***M. foliolosa* (AF4308)**	322.66	195.55	275.34	130.92	138.03	98.46	129.71	221.87	264.68	55.06	195.33	109.82	124.26	52.05	180.83	164.02
45	***M. squarrosa* (RDE75)**	464.71	250.76	288.48	148.88	182.36	168.25	177.23	317.08	321.9	84.09	230.38	127.5	225.74	72.11	274.03	223.25
46	***M. sylvana* (AF4286)**	472.39	219.28	252.46	149.4	162.9	142.8	152.67	333.91	310.98	73.7	266.74	127.02	140.65	60.05	221.01	179.38
47	***Calothamnus gracilis* (RDE154)**	276.03	184.39	170.05	72.54	85.88	73.89	250.19	186.31	162.65	98.13	242.36	93.16	161.01	36.88	134.11	133.17

Supplementary Table S1. (continued).

	Species	n_33353	n_33384	n_33409	n_33459	n_33482	n_33521	n_33548	n_33877	n_33912	n_34024	n_34029	n_34206	n_34262	n_34341	n_34473	n_34730
1	***A. gummiferum* (McCoys.n.)**	372.73	352.19	175.85	52.03	115.23	296.02	143.72	162.81	245.2	241.75	316.87	119.28	297.44	173.14	193.17	186.82
2	***A. gummiferum* (McCoys.n.)**	206.26	224.1	86.87	23.35	70.47	145.16	82.42	100.16	142.14	159.5	210.75	64.21	157.49	94.75	115.05	117.58
3	***Corymbia bunites* (Cs229)**	588.33	779.04	235.42	138.36	160.61	447.98	202.11	267.74	486.47	583.59	400.12	314.74	576.47	207.66	398.51	289.46
4	***C. calophylla* (Cs377)**	5.92	3.34	0.1	0.89	2.52	3.07	1.56	0.84	2.4	3.38	2.4	3.58	1.01	2.03	1.39	1.94
5	***C. haematoxylon* (Cs380)**	73.74	95.36	41.11	14.79	25.41	76.36	31.94	37.48	61.22	143.87	42.28	41.12	88.14	18.58	52.21	35.37
6	***C. peltata* (Cs493)**	406.76	557.9	138.45	68.19	130.01	326.43	164.71	185.14	345.68	352.29	328.27	175.45	489.02	134.8	303.88	224.6
7	***E. brockwayi* (Cs474)**	192.1	266.11	135.37	148	118.89	164.1	111.65	130.43	187.78	184.2	206.34	136.04	177.9	149.71	151.9	101.79
8	***E. coccifera* (ANBG9404806)**	180.11	287.39	125.92	185.62	103.41	178.24	91.74	123.99	132.52	193.42	163.89	124.09	183.51	112.34	175.26	123.16
9	***E. extrica* (CANB632680)**	187.1	249.44	104	15.46	64.44	167.39	95.61	89	156.62	314.51	174.45	87.89	208.86	72.45	121.8	108.99
10	***E. globoidea* (Cs729)**	131.34	200.87	77.24	38.48	71.18	105.92	246.72	65.86	104.68	148.47	134.28	94.39	108.86	76.63	103.01	76.43
11	***E. globulus* (CANB688145)**	151.78	232.25	105.99	15.53	71.51	171.83	79.84	88.99	134.33	138.36	157.44	76.38	188.99	86.44	135.54	98.35
12	***E. globulus* (CANB494954)**	52.47	83.23	23.61	25.7	21.05	48.9	25.59	26.16	53.86	90.73	40.36	22	63.01	25.53	37.76	23.81
13	***E. gomphocephala* (Cs357)**	194.94	317.81	108.44	70	79.02	153.77	93.07	95.64	150.91	195.29	169.31	98.07	214.73	92.46	140.63	110.17
14	***E. goniantha* (Cs424)**	2.63	2.65	2.88	0	2.58	2.69	2.57	1.41	3.39	3.51	3.09	1.7	1.83	2.42	2.06	2.55
15	***E. howittiana* (Cs496)**	127.71	199.3	57.78	62.19	72.62	107.59	57.32	58.14	93.59	222.85	132.93	81.89	133.16	94.09	82.38	71.58
16	***E. insularis* (Cs275)**	102.76	123.51	51.11	14.86	49.66	109.96	110.25	55.53	75.97	82.23	91.71	85.87	102.27	64.56	83.73	59.61
17	***E. insularis* (CANB413807)**	75.35	85.87	46.42	9.61	30.19	65.44	72.27	37.99	64.94	67.6	74.99	61.11	85.62	25.82	55.29	46
18	***E. leucoxylon* (Cs716)**	141.34	214.47	71.63	54.49	61.45	141.51	75.41	96.69	106.84	208	136.5	106.4	152.28	91.74	134.67	82
19	***E. ligulata* (Cs721)**	17.48	22.71	7.32	3.52	11.29	22.05	33.26	12.42	15.74	22.72	17.59	18.33	8.84	12.38	13.4	16.04
20	***E. moorei* (Cs574)**	243.66	344.46	144.55	172.5	106.85	243.29	213.14	123.59	208.38	377.03	250.83	141.14	219.9	109.24	179.8	135.83
21	***E. nitida* (Cs744)**	132.18	106.43	42.88	43.19	36.59	82.14	159.84	48.17	92.66	153.79	91.83	56.74	81.12	48.29	69.68	63.72
22	***E. optima* (Cs708)**	49.3	57.9	26.44	24.52	28.55	46.08	31	28.53	33.56	87.73	48.61	29.2	47.93	34.89	38.89	30.68
23	***E. pachycalyx* (Cs305)**	7.23	13.63	4.83	1.7	6.75	8.39	3.53	5.82	6.25	11.14	7.18	5.28	6.13	8.19	5.58	5.44
24	***E. pachyloma* (Cs426)**	82	92.34	56.14	10.95	26.86	84.89	151.98	34.41	56.11	67.39	77.48	40.14	69.22	29.19	47.34	50.13
25	***E. perriniana* (Cs742)**	59.66	89.05	51.17	44.16	29.79	56.55	27.5	42.98	55.12	63.05	55.15	46	62.37	38.02	48.23	37.74
26	***E. pilularis* (Cs482)**	243.7	404.57	122.09	51.9	117.89	218.58	252.22	128.49	203.71	204.44	259.06	139.41	212.16	136.53	170.54	154.64

27	***E. platydisca* (CANB638520)**	296.41	356.98	140.64	44.5	85	307.75	464.18	141.52	227.35	227.96	284.14	142.8	246.78	117.19	205.35	171.73
28	***E. pleurocarpa* (Cs431)**	129.66	137.26	58.16	10.87	40.25	106.08	68.48	58.78	99.13	235.64	94.84	64.7	128.72	56.22	92.37	78.36
29	***E. preissiana* (Cs425)**	20.19	28.18	14.36	7.29	12.26	14.69	11.3	11.62	19.23	30.77	22.69	13.09	17.61	10.74	13.05	10.52
30	***E. pumila* (Cs227)**	274.54	261.05	133.97	162.19	87.55	236.3	115.81	133.1	172.98	275.9	217.22	131.71	252.38	123.62	192.62	155.3
31	***E. selachiana* (CANB632673)**	70.83	90.46	33.27	10.04	24.26	57.7	36.4	37.34	47.43	113.91	63.14	20.65	59.6	20.73	50.58	31.28
32	***E. selachiana* (CANB693174)**	200.13	223.55	86.44	19.51	78.85	164.57	104.41	109.16	150.85	370.02	165	89.34	169.68	85.84	138.81	100.25
33	***E. spathulata* (Cs368)**	212.39	221.09	84.88	79.7	79.03	181.95	85.35	120.72	152.92	140.54	159.1	107.56	213.69	115.72	127.31	127.12
34	***E. staeri* (Cs366)**	37.12	57.04	20.8	17.35	14.86	43.79	63.11	14.82	30.93	40.99	43.21	28.82	36.9	17.71	32.52	25.71
35	***E. stenostoma* (Cs733)**	5.17	10.55	4.55	1.84	3	6.1	7.67	3.2	4.89	8.27	6.05	5.2	4.82	1.73	4	3.59
36	***E. stoatei* (Cs191)**	190.29	267.77	119.38	163.85	135.33	166.61	100.85	132.94	155.64	236.3	197.09	147.28	197.8	142	148.75	112.23
37	***E. tereticornis* (Cs769)**	309.42	470.93	182.63	189.41	140.8	278.45	161.39	178.4	246.46	281.68	304.38	302.48	349.41	206.47	217.54	229.27
38	***E. verrucata* (Cs715)**	84.88	77.27	36.12	39.07	31.07	80.31	164.34	36.86	55.83	119.41	67.09	37.54	69.6	28.35	61.21	51.03
39	***E. virginea* (Cs727)**	64.37	57.84	38.21	39.38	35.17	60.32	89.67	31.61	52.04	58.54	64.55	53.04	67.71	36.29	44.96	32.7
40	***E. willisii* (Cs492)**	81.83	87.48	38.2	27.93	31.34	61.87	108.29	32.13	50.69	98.68	68.29	52.09	79.54	41.15	47.99	40.58
41	***Heteropyxis natalensis* (CANB576168)**	5.01	8.77	0.7	0	4.45	10.92	3.86	1.92	23.09	4.16	6.8	2.75	9.92	1.86	4.39	3.79
42	***Melaleuca capitata* (RDE176)**	564.69	501.72	168	0.71	450.02	416.95	238.67	218.93	392.63	525.22	428.07	402.1	519.32	868.85	334.61	341.19
43	***M. cornucopiae* (Harwood1546)**	680.64	775.59	254.89	0.41	610.91	682.52	359.23	210.95	447.3	606.64	592.83	305.77	764.08	443.09	460.32	459.2
44	***M. foliolosa* (AF4308)**	232.59	247.69	58.2	60.43	178.79	154.83	109.39	65.95	166.37	214.03	198.77	151.14	249.6	113.55	167.05	182.33
45	***M. squarrosa* (RDE75)**	307.54	320.52	119.37	0.15	246.58	263.62	154.52	144.68	241.92	497.72	277.83	267.31	660	846.31	211.92	239.36
46	***M. sylvana* (AF4286)**	269.72	284.68	75.93	0	240.5	132.68	134.11	71.78	201.65	301.61	245.54	260.58	244.36	201.97	195.07	219.21
47	***Calothamnus gracilis* (RDE154)**	186.03	244.24	46.22	0	137.17	183.59	111.25	34.06	129.6	171.56	153.65	122	227.14	141.26	132.22	135.95

Supplementary Table S1. (continued).

	Species	n_34792	n_34840	n_34877	n_40086	n_40184	n_40267	n_40333	n_40488	n_40593	n_40617	n_40707	n_40711	n_40715	n_40746	n_40755	n_40796
1	***A. gummiferum* (McCoys.n.)**	244.16	246.8	313.68	104.39	202.67	184.52	143.97	25.77	291.83	327.21	172.15	196.09	194.83	699.43	346.02	164.66
2	***A. gummiferum* (McCoys.n.)**	160.23	137.03	196.24	53.18	90.47	106.06	78.17	12.22	149.55	188.43	115.08	104.66	118.5	420.02	209.74	108.1
3	***Corymbia bunites* (Cs229)**	533.87	356.68	510.74	193.3	204.64	342.33	288.23	69.19	387.38	720.18	308.19	356.67	340.23	723.52	1151.12	287.21
4	***C. calophylla* (Cs377)**	2.18	1.28	1.58	0.36	0	1.08	3.34	0	3.5	5.73	0	3.86	2.01	4.58	6.39	0.07
5	***C. haematoxylon* (Cs380)**	60.13	42.82	66.29	18.79	30.12	39.76	37.35	3.34	59.89	89.72	51.01	41.93	41.64	76.26	103.18	29.7
6	***C. peltata* (Cs493)**	331.84	270.36	356.23	115.48	212.99	278.66	205.91	44.89	321.52	542.43	191.26	221.07	265.07	538.45	671.29	209.69
7	***E. brockwayi* (Cs474)**	228.47	154.76	209.32	115.48	132	164.8	122.57	111.27	173.24	264.07	156.09	174.39	164.19	224.5	259.8	145.82
8	***E. coccifera* (ANBG9404806)**	197.38	149.07	183.85	145.73	146.72	147.76	121.82	275.49	187.54	265.16	123.24	131.55	169.72	225.32	229.26	120.33
9	***E. extrica* (CANB632680)**	153.46	130.1	168.94	307.19	110.45	143.54	106.25	90.68	167.25	209.99	116.45	79.61	142.32	230.82	183.82	101.7
10	***E. globoidea* (Cs729)**	142.11	77.57	145.7	55.1	83.45	89.37	67.29	79.71	95.12	157.84	117.27	100.77	85.73	135.26	158.14	89.77
11	***E. globulus* (CANB688145)**	200.47	151.59	161.97	130.43	94.04	158.55	108.3	148.51	158.06	199.85	104.28	96.47	144.29	201.98	204.69	118.6
12	***E. globulus* (CANB494954)**	48.46	42.87	49.76	32.71	13.92	38.18	27.58	58.41	39.33	58.05	31.01	23.09	44.75	66.03	71.18	33.98
13	***E. gomphocephala* (Cs357)**	201.55	137.76	156.16	99.53	123.6	150.79	91.87	75.74	135.13	260.57	115.32	113.3	148.63	275.43	232.83	111.83
14	***E. goniantha* (Cs424)**	3.93	3.06	3.43	1.63	3.33	2.12	2.41	4.81	1.92	3.04	2.31	3.46	4.55	4.04	3.82	1.37
15	***E. howittiana* (Cs496)**	143.57	86.48	124.4	68.03	76.64	85.22	61.56	53.34	112.47	134.23	79.43	93.48	84.85	150.83	143.2	83.23
16	***E. insularis* (Cs275)**	116.71	70.72	106.47	61.24	59.83	93.41	61.67	112.02	98.73	116.37	62.03	64.91	73.77	141.95	135.2	58.15
17	***E. insularis* (CANB413807)**	84.38	57.31	78.59	37.37	36.61	54.63	40.86	42.03	58.12	92.08	45.43	43.22	54.27	97.39	87.5	44.57
18	***E. leucoxylon* (Cs716)**	138.25	92.23	129.29	77.65	105.94	112.43	110.45	103.78	123.34	178.25	90.05	92.63	118.77	187.32	216.43	90.71
19	***E. ligulata* (Cs721)**	19.63	7.29	18.08	10.73	8.61	14.67	10.41	6.9	13.89	19.2	10.71	15.77	10.49	22.42	13.5	11.22
20	***E. moorei* (Cs574)**	251.25	197.66	228.62	149.35	118.39	165.86	143.02	228.28	198.86	289.73	148.86	156.2	205.33	379.22	272.99	160.18
21	***E. nitida* (Cs744)**	105.71	75.5	95.02	55.13	73.34	87.26	52.33	76.39	64.68	112.86	49.41	62.34	78.4	114.24	102.5	60.45
22	***E. optima* (Cs708)**	74.96	36.19	41.91	27.23	18.28	30.39	33.1	26.58	45.5	63.91	42.22	36.96	34.64	62.51	67.32	22.91
23	***E. pachycalyx* (Cs305)**	7.87	10.25	5.12	2.36	1.91	9.31	4.6	7.89	7.43	8.39	5.96	7.06	6.73	7.56	11	4.34
24	***E. pachyloma* (Cs426)**	80.79	53.02	73.52	45.84	47.07	54.62	47.52	41.55	57.45	79.17	51.65	43.25	50.9	98.03	73.14	52.56
25	***E. perriniana* (Cs742)**	75.2	44.99	50.07	50.55	25.11	54.59	42.05	36.48	50.48	77.38	35.21	35.24	54.31	84.57	61.07	48.42
26	***E. pilularis* (Cs482)**	273.15	186.67	282.38	133.04	147.86	201.39	153.18	313.19	165.37	315.28	148.61	165.07	198.48	344.03	256.52	164.62

27	***E. platydisca* (CANB638520)**	266.51	344.69	293.56	163.82	117.46	236.55	163.96	218.4	219.13	347.9	136.34	139.04	223.03	378.23	327.24	185.01
28	***E. pleurocarpa* (Cs431)**	115.64	95.52	95.39	212.28	66.48	84.98	74.37	88.63	101.46	145.87	71.55	60.38	95.2	166.49	137.19	69.83
29	***E. preissiana* (Cs425)**	26.15	14.4	19.21	8.66	10.6	11.72	9.94	7.27	21.68	26.53	11.45	16.31	13.56	19.23	28.3	12.98
30	***E. pumila* (Cs227)**	247.45	175.54	214.86	130.34	126.66	184.37	134.02	205.27	191.63	272.87	121.94	137.61	190.7	311.17	323.07	173.59
31	***E. selachiana* (CANB632673)**	56.37	55.98	57	61.04	43.51	40.48	41.44	29.28	34.14	72.54	38.31	21.49	49.23	76.97	70.55	29.12
32	***E. selachiana* (CANB693174)**	179.71	164.53	203.59	204.9	124.4	129.76	125.05	133.85	181.75	241	128.19	133.78	148.4	236.31	192.09	109.8
33	***E. spathulata* (Cs368)**	187.03	133.73	167.16	102.75	77.13	136.89	117.14	88.87	142.69	217.67	117.18	131.3	165.7	283.4	238.41	116.76
34	***E. staeri* (Cs366)**	34.61	26.6	43.69	18.31	17.3	29.58	17.4	34.19	24.73	43.07	25.29	28.76	25.28	45.62	48.92	22.87
35	***E. stenostoma* (Cs733)**	7.36	2.42	5.63	3.63	0.51	3.21	2.91	0.65	7.55	8.44	5.59	3.83	4.39	4.65	5.6	4.86
36	***E. stoatei* (Cs191)**	184.37	142.52	200.74	134.81	147.78	141.44	135.42	115.68	173.95	250.83	124.05	160.9	135.91	229.34	231.9	136.93
37	***E. tereticornis* (Cs769)**	333.18	248.26	313.42	212.39	163.39	220.68	178.53	197.47	257.56	408.53	175.38	211.03	263.69	409.37	324.56	230.75
38	***E. verrucata* (Cs715)**	85.97	62.81	79.53	42.82	59.56	59.38	47.27	53.27	67.78	107.31	44.21	51.05	56.53	115.11	87.4	45.11
39	***E. virginea* (Cs727)**	82.47	44.69	63.71	40.64	22.05	42.53	34.13	43.78	52.59	86.59	45.45	44.75	46.78	85.14	80.57	42.34
40	***E. willisii* (Cs492)**	49.06	53.23	64.86	39.27	35.03	47.48	38.45	48.21	43.46	84.81	32.69	43.59	45.61	91.73	89.02	47.19
41	***Heteropyxis natalensis* (CANB576168)**	2.18	3.66	15.93	1.81	4.49	10.87	2.78	0.73	3.74	7.51	3.85	0.54	8.5	16.91	13.53	4.17
42	***Melaleuca capitata* (RDE176)**	527.64	417.69	497.9	144.08	211.41	297.39	212.82	146.4	507.99	648.77	357.19	486.48	266.67	701.93	1086.64	281.65
43	***M. cornucopiae* (Harwood1546)**	557.31	576.43	781.75	96.1	403.03	414.96	317.08	63.82	772.52	942.71	481.38	543.44	415.29	871.36	986.33	309.24
44	***M. foliolosa* (AF4308)**	289.13	187.38	246.24	40.66	130.6	145.47	108.33	0.28	261.14	312.81	180.33	211.71	136.2	280.13	284.63	139.67
45	***M. squarrosa* (RDE75)**	313.94	256.49	301.89	60.61	111.72	200.93	134.43	78.34	271.11	417.7	232.98	279.29	161.59	362.64	702.39	145.88
46	***M. sylvana* (AF4286)**	337.51	208.75	303.37	63.44	130.51	160.27	112.19	2.76	292.38	430.55	222.98	266.17	144.77	361.35	389.03	167.49
47	***Calothamnus gracilis* (RDE154)**	201.96	135.64	462.23	26.05	172.04	229.15	99.67	27.06	192.81	260.38	139.95	170.88	103.5	284.55	277.54	83.47

Supplementary Table S1. (continued).

	Species	n_40821	n_40832	n_40849	n_40909	n_40919	n_41002	n_41012	n_41115	n_41143	n_41174	n_41186	n_41200	n_41215	n_41348	n_41364	n_41372
1	***A. gummiferum* (McCoys.n.)**	203.97	225.56	202.49	29.95	282.19	252.8	161.06	183.95	152.88	275.87	233.65	145.79	171.03	2.81	80.89	215.77
2	***A. gummiferum* (McCoys.n.)**	122.47	117.91	141.03	18.22	168.94	150.01	86.57	115.76	75.09	153.64	124.54	84.97	103.45	4.86	34.99	119.37
3	***Corymbia bunites* (Cs229)**	305.07	437.58	366.63	48.8	492.78	556.71	334.39	439.42	252.72	474.01	358.5	302.91	321.2	35.49	98.19	366.22
4	***C. calophylla* (Cs377)**	0.55	2.07	0.48	0	2.79	4.67	2.55	2.57	1.45	3.3	0.99	0.98	0.86	0	0.97	3.06
5	***C. haematoxylon* (Cs380)**	35.15	53.69	55.52	7.75	59.16	84.48	35.09	49.26	31.41	62.32	46.47	33.69	42.28	5.67	14.04	49.23
6	***C. peltata* (Cs493)**	238.81	304.65	222.06	43.14	364.53	416.97	222.89	341.36	188.01	356.84	266.76	207.21	270.62	24.5	75.55	257.78
7	***E. brockwayi* (Cs474)**	159.86	159.94	162.12	19.23	183.5	226.24	115.42	177.15	151.48	189.39	208.21	117.69	131.03	101.8	77.74	158.83
8	***E. coccifera* (ANBG9404806)**	179.32	156.26	156.88	12.43	187.63	189.71	184.03	164.11	161.23	187.22	177.14	107	146.19	85.29	65.87	147.18
9	***E. extrica* (CANB632680)**	132.71	139.8	181.67	11.35	143.39	170.48	98.89	128.9	124.49	167.59	131.45	129.6	127.56	68.67	52.03	123.11
10	***E. globoidea* (Cs729)**	101.72	110.94	94.94	6.93	116.06	134.17	68.19	94.45	97.8	114.62	112.95	71.04	81.1	51.19	57.07	100.64
11	***E. globulus* (CANB688145)**	158.66	132.2	110.61	9.1	158.63	154.81	304.89	113.96	135.19	142.59	138.96	106.44	125.57	81.12	46.18	117.58
12	***E. globulus* (CANB494954)**	45.93	48.58	38.41	3.7	37.83	45.49	81.61	43.6	35.38	41.97	41.25	26.94	37.32	26.27	15.57	35.52
13	***E. gomphocephala* (Cs357)**	169.59	157.31	123.12	7.39	157.16	183.75	122.25	145.48	139.15	157.85	161.78	114.08	128.45	94.14	38.83	129.89
14	***E. goniantha* (Cs424)**	2.85	1.31	2.43	0.3	4.12	3.72	2.05	1.7	3.66	2.18	3.01	2.8	2.32	0	1.41	1.86
15	***E. howittiana* (Cs496)**	102.32	93.33	92.22	3.89	102.13	133.2	64.45	94.51	86.2	102.05	88.56	62.05	68.56	84.36	49.72	99.7
16	***E. insularis* (Cs275)**	97.47	89.48	73.43	7.88	87.21	103.69	56.92	80.09	83.38	91.37	83.53	66.16	70.59	28.08	24.59	80.23
17	***E. insularis* (CANB413807)**	66.2	51.23	50.99	2.29	71.37	66.18	37.51	47.83	54.56	66.14	67.47	46.64	50.4	37.63	23.85	50.37
18	***E. leucoxylon* (Cs716)**	147.95	122.52	91.5	6.11	105.8	130.21	92.24	98.22	121.72	126.18	121.99	93.55	91.94	68.84	37.17	102.69
19	***E. ligulata* (Cs721)**	13.35	8.07	11.28	0.61	18.48	14.07	8.68	14.62	15.58	17.14	15.46	9.38	10.42	7.37	5.08	12.04
20	***E. moorei* (Cs574)**	207.06	334.08	201.61	11.25	197.14	217.65	134.14	174.82	194.42	220.66	212.19	146.03	176.66	111.8	54.43	167.7
21	***E. nitida* (Cs744)**	86.34	95.21	65.37	7.43	103.7	91.3	61.62	67.22	61.39	91.54	73.06	64.36	56.42	40.68	25.84	62
22	***E. optima* (Cs708)**	307.86	41.31	24.28	3.99	41.08	47.92	27.56	41.02	33.92	41.46	44.69	33.44	30.07	18.05	11.94	32.27
23	***E. pachycalyx* (Cs305)**	6.37	6.74	6.25	1.21	6.77	10.12	6.3	6.77	7.4	8.27	7.81	3.84	5.66	3.33	4.49	7.89
24	***E. pachyloma* (Cs426)**	65.13	54.98	57.75	4.34	68.23	62.22	45.06	54.6	51.94	66.97	61.9	46.44	43.38	32.15	19.72	42.48
25	***E. perriniana* (Cs742)**	52.5	53.68	47.94	7.09	57.35	54.38	56.26	56.03	50.18	60.34	53.2	29.28	49.25	30.38	17.97	56.26
26	***E. pilularis* (Cs482)**	196.01	264.44	180.92	18.47	233.4	245.11	169.95	192.72	202.62	238.92	211.71	158.28	152.83	106.08	71.09	170.86

27	***E. platydisca* (CANB638520)**	277.39	226.85	212.49	24.3	236.83	245.87	163.55	188.31	195.06	285.29	244.77	177.01	186.99	89.05	56.52	205.78
28	***E. pleurocarpa* (Cs431)**	88.73	97.16	92.42	2.28	102.53	106.55	77.32	82.4	88.78	113.93	104.74	84.07	77.74	47.97	31	90.13
29	***E. preissiana* (Cs425)**	13.63	14.19	10.44	0.75	19.59	19.11	10.79	16.7	13.05	15.87	19.96	10.63	10.7	14.37	10.03	14.56
30	***E. pumila* (Cs227)**	192.95	209.72	140.96	8.84	220.72	201.68	154.74	188.8	189.88	193.7	205.02	163.68	164.37	119.66	64	180.52
31	***E. selachiana* (CANB632673)**	51.56	49.82	35.72	2.09	51.51	52.97	33.98	46.05	37.33	60.09	50.01	38.22	34.46	21.11	18.53	46.87
32	***E. selachiana* (CANB693174)**	156.72	136.18	133.5	12.01	164.74	190.37	112.32	145.98	127.3	185.69	152	124.2	124.32	108.7	49.6	137.55
33	***E. spathulata* (Cs368)**	167.06	168.01	161.75	13.5	176.5	150.61	128.26	150.01	120.21	185.91	169.16	101.59	104.03	94.64	58.9	139.09
34	***E. staeri* (Cs366)**	36.53	27.57	25.73	1.79	37.33	38.48	23.34	28.09	28.55	32.72	31.58	27.96	24.66	19.34	8.82	33.62
35	***E. stenostoma* (Cs733)**	3.43	4.21	4.01	0.71	3.41	6.75	3.27	3.12	4.09	5.57	3.85	1.84	2.29	10.11	0.82	3.02
36	***E. stoatei* (Cs191)**	149.83	175.23	157.34	17.47	172.74	192.87	122.1	173.42	149.08	176.35	160.84	108.42	137.76	89.6	74.5	168.58
37	***E. tereticornis* (Cs769)**	303.18	252.59	203.4	20.68	286.08	294.72	303.05	258.45	261.15	300.28	277.51	210.42	208	145.87	110.32	225.72
38	***E. verrucata* (Cs715)**	56.44	62.38	67.21	9.75	66.32	75.3	42.13	54.07	58.38	71.88	60.33	46.64	43.68	44.54	31.04	40.86
39	***E. virginea* (Cs727)**	51.66	49.08	55.26	6.17	57.21	63.07	31.87	51.32	52.93	59.11	53.42	39.96	36.02	38.64	19.61	42.4
40	***E. willisii* (Cs492)**	52.27	53.31	41.51	2.88	67.65	62.85	42.22	48.2	44.14	51.38	39.56	40.17	45.85	31.13	12.32	46.01
41	***Heteropyxis natalensis* (CANB576168)**	0.75	4.95	5.01	0.4	10.64	12.09	10.7	1.05	2.46	5.2	7.5	6.01	5.82	0	1.09	6.3
42	***Melaleuca capitata* (RDE176)**	139.46	490.2	299.98	113.38	381.65	345.16	193.24	353.22	191.27	320.95	318.26	196.12	240.68	0	227.73	290.17
43	***M. cornucopiae* (Harwood1546)**	191.12	614.44	378.9	114.89	401.8	418.35	262.83	478.79	321.36	384.37	282.21	210.92	337.4	56.16	120.97	499.18
44	***M. foliolosa* (AF4308)**	61.21	215.35	85.12	34.35	162.53	163.45	90.85	187.01	68.7	133.3	116.93	100.38	117.64	11.3	59.97	144.83
45	***M. squarrosa* (RDE75)**	82.48	262.7	147.92	55.23	238.65	206.32	116.94	196.31	116.4	195.36	178.26	106.84	151.79	4.34	113.05	174.22
46	***M. sylvana* (AF4286)**	87.6	283.39	165.81	47.79	238.87	202.21	101.24	210.91	90.38	193.61	136.91	99.42	115.21	25.88	83.18	172.68
47	***Calothamnus gracilis* (RDE154)**	50.42	182.89	56.53	26.07	126.22	121.21	81.9	134.34	65.03	128.78	99.07	71.5	104.74	12.4	51.68	146.07

Supplementary Table S1. (continued).

	Species	n_41395	n_41455	n_41455	n_41625	n_41627	n_41712	n_41818	n_41838	n_41845	n_41892	n_41898	n_41911	n_41930	n_41999	n_42036	n_42060
1	***A. gummiferum* (McCoys.n.)**	227.24	203.91	3.2	57.68	205.2	180.46	267.97	73.6	372.37	160.16	263.94	109.32	135.01	182.13	240.09	63.95
2	***A. gummiferum* (McCoys.n.)**	130.83	121.18	0.3	44.12	119.36	99.39	184.32	41.82	195.99	70.22	154.54	63.42	66.47	112.13	143.97	37.44
3	***Corymbia bunites* (Cs229)**	448.91	427.67	0.26	40.46	390.74	268.8	310.85	163.21	444.16	209.14	415.87	175.63	248.62	322.37	450.03	163.5
4	***C. calophylla* (Cs377)**	3.79	2.98	0	0	2.21	2.28	1.62	1.33	1.31	0	2.1	1.73	0.74	1.6	2.34	1.99
5	***C. haematoxylon* (Cs380)**	50.38	53.43	0	1.77	55.98	39.22	53.59	20.34	53.33	29.17	55.78	33.17	36.26	33.32	64.17	17.57
6	***C. peltata* (Cs493)**	391.71	289.81	2.39	14.38	282.37	190.27	220	115.05	302.5	142.54	324.58	140.46	201.71	215.53	367.28	94.95
7	***E. brockwayi* (Cs474)**	181.33	200.4	160.82	170.15	156.42	90.38	424.62	139.54	144.39	169.62	178.03	121.42	91.82	175.83	167.73	149.15
8	***E. coccifera* (ANBG9404806)**	168.71	144.4	153.04	228.55	144.02	99.35	189.88	96.81	132.46	216.49	216.4	154.45	107.57	155.2	163.82	147.52
9	***E. extrica* (CANB632680)**	142.22	137.4	146.46	195.59	109.52	85.07	140.89	201.05	142.04	129.02	165.38	119.08	86.95	130.23	162.53	90.99
10	***E. globoidea* (Cs729)**	112.65	113.2	70.19	109.34	72.93	203.43	118.79	46.07	77.83	103.91	106.84	189.49	58.15	118.8	108.89	95.74
11	***E. globulus* (CANB688145)**	142.94	151.73	151.73	193.62	124.35	82.24	104.19	96.24	94.72	143.55	203.47	94.61	96.56	125.98	157.41	103.53
12	***E. globulus* (CANB494954)**	33.7	34.31	33.51	34.31	40.22	26.45	28.57	57.8	22.99	32.72	60.72	20.62	29.79	42.4	43.87	26.53
13	***E. gomphocephala* (Cs357)**	141.35	148.56	110.91	84.41	127.01	104.17	250.79	157.9	154.21	140.4	194.16	133.8	160.59	131.22	167.39	106.74
14	***E. goniantha* (Cs424)**	2.27	3.86	5.97	4.05	2.77	2.39	1.86	1.65	3.99	1.3	3	1.52	2.61	1.46	3.58	2.4
15	***E. howittiana* (Cs496)**	107.39	100.35	74.2	48.15	83.4	61.27	134.09	63.6	75	101.93	124.68	57.88	58.74	81.64	98.98	91.58
16	***E. insularis* (Cs275)**	83.34	75.15	74.45	84.65	69.91	88.71	79.89	36.62	70.84	107.89	93.21	125.67	58.49	74.05	96.72	61.42
17	***E. insularis* (CANB413807)**	29.97	54.19	32.31	56.98	49.79	56.38	70.4	23.54	51.94	79.41	71.7	67.1	56.48	62.67	74.76	36.84
18	***E. leucoxylon* (Cs716)**	126.1	117.75	85.95	138.87	113.38	78.44	252.69	131.69	98.64	111.73	144.9	64.74	72.42	106.24	119.96	99.74
19	***E. ligulata* (Cs721)**	12.31	15.01	9.34	17.69	5.7	26.86	13.69	11.7	12.06	15.32	14.61	16.3	10.14	14.8	14.41	11.75
20	***E. moorei* (Cs574)**	184.09	178.01	176.23	180.28	150.68	180.94	191.76	84.27	134.13	195.87	219.39	173.93	126.29	166.95	220.65	136.88
21	***E. nitida* (Cs744)**	97.35	74.64	72.69	41.16	60.28	129.69	72.74	36.04	70.09	70.4	82.27	161.84	50.75	78.5	94.75	66.11
22	***E. optima* (Cs708)**	39.04	36	29.56	38.28	34.03	24	117.02	34.44	29.99	36.75	48.16	31.6	21.13	35.82	38.53	34.58
23	***E. pachycalyx* (Cs305)**	5.01	5.4	4.92	9.45	5.41	3.46	7.8	3.66	1.42	16.98	8.62	8.88	4.42	7.43	8.33	8.31
24	***E. pachyloma* (Cs426)**	57.87	58.53	71.21	57.93	52	119.56	53.58	30.24	45.51	48.26	56.32	63.84	35.05	67.49	67.71	57.56
25	***E. perriniana* (Cs742)**	46.34	45.56	62.47	52.45	45.1	29.83	61.66	35.07	40.43	59.82	65.74	52.16	33.12	56.2	48.39	43.54
26	***E. pilularis* (Cs482)**	184.16	194.13	162.49	237.27	179.8	225.96	184.98	95.3	167.26	163.38	221.72	202.86	157.71	211.09	223.11	172.19

27	***E. platydisca* (CANB638520)**	196.56	188.68	202.48	256.29	186.18	392.08	222.86	128.55	145.57	241.56	253.06	186.14	151.29	195.65	263.56	164.98
28	***E. pleurocarpa* (Cs431)**	124.59	111.83	121.9	75.7	79.21	53.86	139.53	130.39	82.79	89.61	111.07	70.8	69.02	87.92	105.92	80.99
29	***E. preissiana* (Cs425)**	13.98	19.24	14.07	10.95	12.38	11.63	20.32	10.72	13.18	18.69	15.59	17.5	13.32	15.7	15.11	12.82
30	***E. pumila* (Cs227)**	170.31	172.4	179.63	195.23	162.61	91.3	223.29	130.42	169.24	185.01	255.12	145.31	136.1	172.71	209.09	127.8
31	***E. selachiana* (CANB632673)**	61.98	40.45	29.21	33.69	35.66	29.7	61.04	40.19	50.97	44.14	53.5	29.65	28.5	44.62	52.33	30.21
32	***E. selachiana* (CANB693174)**	227.84	146.86	124.65	90.95	136.92	79.5	120.43	159.36	151.2	122.07	189.11	122.83	94.18	133.14	156.48	113.91
33	***E. spathulata* (Cs368)**	137.29	146.39	153.6	181.29	147.98	93.65	165.93	123.76	167.66	206.06	172.38	124.68	96.29	140.78	157.89	135.17
34	***E. staeri* (Cs366)**	34.13	35.82	9.63	31.28	19.48	51.65	53.17	36.56	21.14	32.46	32.72	31.93	28.97	27.32	32.99	21.8
35	***E. stenostoma* (Cs733)**	5.29	4.4	2.9	1.43	2.96	7.08	5.59	3.84	0.26	5.15	4.53	5.17	2.7	4.79	5.33	6.31
36	***E. stoatei* (Cs191)**	148.88	151.52	135.7	194.72	152.63	116.45	168.81	99.92	139.16	175.2	204.51	190.23	204.62	172.72	162.86	143.02
37	***E. tereticornis* (Cs769)**	388.26	292.66	233.06	288.88	238.85	135.49	246.98	180.64	225.5	221.03	320.8	256.46	174.98	245.06	266.17	208.54
38	***E. verrucata* (Cs715)**	63.5	54.48	54.42	50.99	57.3	132.63	69.58	24.25	49.22	74.92	75.51	82.55	37.1	63.24	62.05	43.46
39	***E. virginea* (Cs727)**	51.64	60.72	48.32	59.01	41.92	70.08	30.36	26.73	36.42	67.74	56.85	81.31	48.27	43.89	54.3	45.94
40	***E. willisii* (Cs492)**	37.32	39.94	61.79	56.73	39.37	85.7	58.3	15.24	46.97	73.94	60.74	40.91	43.78	38.77	58.3	31.25
41	***Heteropyxis natalensis* (CANB576168)**	4.04	9.15	0	0.79	18.18	2.95	0	0.27	7.43	0.23	10.55	1.37	7.96	0.27	9.67	0.25
42	***Melaleuca capitata* (RDE176)**	420.05	327.44	0	172.41	311.95	300	361.02	70.11	344.75	63.94	385.17	162.82	155.95	400.63	383.37	221.32
43	***M. cornucopiae* (Harwood1546)**	388.14	444.43	0.3	121.98	426.87	447.73	417.58	22.93	296.25	85.45	624.04	274.65	291.26	368.67	519.03	259.79
44	***M. foliolosa* (AF4308)**	189.34	160.87	0	39.03	163.55	137.72	155.5	9.67	114.76	21.28	198.05	80.91	90	152.04	187.52	73.99
45	***M. squarrosa* (RDE75)**	273.92	236.56	0.3	112.41	201.19	189.44	229.59	34.81	167.7	43.53	268.38	91.48	269.33	234.2	234.4	142.18
46	***M. sylvana* (AF4286)**	258.74	238.04	0.17	65.26	197.05	162.5	209.94	25.65	179.5	31.88	229.43	97.35	105.15	206.19	220.96	111.95
47	***Calothamnus gracilis* (RDE154)**	146.32	149.19	0	47.31	118.4	135.33	134.4	21.15	115.6	19.33	158.92	133.86	81.29	123.3	150.29	56.99

Supplementary Table S1. (continued).

	Species	n_42126	n_42166	n_42230	n_42393	n_42490	n_42561	n_42584	n_42721	n_42754	n_dxs2	n_idi2	n_2	n_13	n_208	n_1873	n_1931
1	***A. gummiferum* (McCoys.n.)**	171.28	276.25	161.73	48.81	126.9	142.44	67.68	91.16	239.68	184.23	313.43	202.8	173.76	246.69	228.23	215.04
2	***A. gummiferum* (McCoys.n.)**	98.47	152.71	79.83	39.63	67.7	74.54	38.8	68.92	139.43	105.8	184.73	116.23	109.5	144.5	130.22	131.12
3	***Corymbia bunites* (Cs229)**	343.03	33.53	292.13	159.5	181.72	269.86	47.03	137.54	486.03	317.69	530.51	391.69	122.6	322.16	438.95	410.51
4	***C. calophylla* (Cs377)**	0.36	2.87	3.98	0.55	1.62	1.93	0.85	0.77	4.03	2.23	2.5	1.46	0	0.81	2.94	3.22
5	***C. haematoxylon* (Cs380)**	36.26	74.71	46.45	17.96	16.37	34.84	5.71	28.3	85.22	46.27	68.37	47.34	13.55	47.3	49.22	52.29
6	***C. peltata* (Cs493)**	228.01	22.46	179.19	97.95	114.3	179.95	34.14	125.78	363.62	262.29	372.56	313.9	101.22	279.96	296.21	339.85
7	***E. brockwayi* (Cs474)**	126.86	373.41	119.74	58.38	89.79	89.5	73.38	73	377.9	150.21	194.99	156.86	228.01	168.55	139.93	187.11
8	***E. coccifera* (ANBG9404806)**	131.3	214.17	90.92	49.72	87.33	90.31	63.56	74.5	278.36	168.88	196.15	139.8	361.31	193.41	122.12	172.94
9	***E. extrica* (CANB632680)**	116.32	281.78	69.22	27.23	77.41	81.85	22.8	40.71	191.38	117.92	170.93	140.34	136.57	177.26	147.17	149.44
10	***E. globoidea* (Cs729)**	76.48	17.35	77.17	49.3	47.1	55.5	20.66	58.01	238.8	92.39	128.63	91.81	128	116.94	95.41	105.49
11	***E. globulus* (CANB688145)**	126.96	92.1	63.09	19.8	80.85	87.6	32.45	47.41	211.51	131.56	179.69	130.28	321.74	188.81	110.92	135.33
12	***E. globulus* (CANB494954)**	24.89	3.78	22.24	5.6	18.86	24.55	9.05	16.63	58.5	33.18	47.29	42.98	99.98	51.22	32.64	37.94
13	***E. gomphocephala* (Cs357)**	143.44	188.33	69.3	31.01	78.2	92.19	34.8	58.71	304.08	132.87	183.25	133.42	166.09	169.83	127.34	148.13
14	***E. goniantha* (Cs424)**	0.86	2.31	2.3	0.91	1.72	3.24	1.46	4.1	5.59	1.41	3.3	3.95	2.82	3.21	1.72	1.85
15	***E. howittiana* (Cs496)**	66.4	130.91	68.18	36.23	40.29	60.52	31.06	59.4	220.63	91.28	109.57	92.39	248.11	104.99	75.41	105.23
16	***E. insularis* (Cs275)**	60.99	163.58	52.34	40.33	39.28	41.67	14.69	53.74	135.18	60.7	110.75	70.04	138.85	118.8	86.02	82.07
17	***E. insularis* (CANB413807)**	35.25	152.15	42.39	28.36	24.99	30.08	7.28	30.34	86.6	45.43	87.56	60.59	87.69	79.91	52.65	55.99
18	***E. leucoxylon* (Cs716)**	107.41	151.11	60.25	39.16	47.78	79.69	37.61	43.4	217.6	118.5	141.58	89.19	255.96	130.95	98.33	114.26
19	***E. ligulata* (Cs721)**	9.33	22.71	10.01	4.2	5.68	12.59	4.71	14.92	33.1	11.35	17.03	14.79	18.31	20.32	11.81	12.8
20	***E. moorei* (Cs574)**	145.4	106.99	102.88	43.66	96.75	94.78	29.8	81.3	274.45	141.99	226.89	172.57	242.59	243.73	160.88	164.97
21	***E. nitida* (Cs744)**	57.53	77.96	42.85	12.12	44.23	37.78	8.49	38.85	125.43	47.74	108.77	72.56	119.48	90.06	67.7	73.89
22	***E. optima* (Cs708)**	34.97	81.05	22.95	15.25	16.5	25.42	23.01	19.44	100.91	44.37	41.7	25.78	62.09	40.2	29.36	36.72
23	***E. pachycalyx* (Cs305)**	4.8	6.41	9.07	3.51	3.62	3.85	2.82	4.25	24.52	5.68	7.12	5.4	3.13	7.56	3.86	5.09
24	***E. pachyloma* (Cs426)**	46.63	71.4	42.65	16.33	31.84	36.19	10.47	28.93	100.22	40.89	71.58	54	105.18	53.84	57.45	56.34
25	***E. perriniana* (Cs742)**	28.8	4.64	30.35	4.39	38.48	37.06	13.92	24.33	96.69	46.29	70.02	42.45	131.92	65.9	52.08	45.34
26	***E. pilularis* (Cs482)**	164.73	244.06	109.35	60.77	120.96	92.4	36.87	89.14	405.19	142.38	262.13	180.85	334.97	213.11	179.92	185.78

27	***E. platydisca* (CANB638520)**	197.03	113.16	107.29	37.32	119.52	112.8	20.51	80.93	421.9	161.21	312.02	211.11	362.02	256.58	219.34	222.37
28	***E. pleurocarpa* (Cs431)**	86.12	210.44	47.27	23.42	56.63	63.05	13.19	33.97	112.83	82.54	104.01	86.16	78.68	126.86	106.56	97.02
29	***E. preissiana* (Cs425)**	13.57	13.77	12.41	5.33	11.5	10.45	4.87	11.15	16.14	20.55	19.06	12.71	34.64	11.68	11.87	13.42
30	***E. pumila* (Cs227)**	157.36	375.81	82.25	48.82	111	122.34	58.31	51.32	236.59	169.82	230.76	189.52	351.08	250.54	172.82	204.75
31	***E. selachiana* (CANB632673)**	43.89	81.38	38.18	4.78	28.25	27.88	7.69	22.9	54.53	31.26	67.43	40.11	43.1	64.16	60.34	56.49
32	***E. selachiana* (CANB693174)**	135.94	304.83	101.23	40.91	77.99	81.01	45.4	82.12	298.79	134.59	174.42	138.6	114.84	200.51	149.32	142.53
33	***E. spathulata* (Cs368)**	124.36	170.93	89.15	40.36	58.89	85.73	39.19	68.3	213.81	126.81	182.43	146.52	196.19	151.45	142.65	137.19
34	***E. staeri* (Cs366)**	31.54	104.87	16.73	3.77	10.04	14.85	4.1	20.85	52.68	25.31	35.67	30.08	66.4	35.62	28.71	27.54
35	***E. stenostoma* (Cs733)**	9.09	6.75	3.08	1.09	2.31	3.73	1.1	3.62	7.99	5.09	5.47	2.84	5.74	5.09	2.06	3.96
36	***E. stoatei* (Cs191)**	135.38	456.45	115.51	63.95	76.23	107.35	60.45	98.23	362.3	159.99	194.24	128.95	266.6	203.77	129.93	157
37	***E. tereticornis* (Cs769)**	211.51	279.22	127.52	87.95	123.77	151.95	76.19	96.91	558.03	255.82	327.96	252.6	751.52	297.01	250.5	259.83
38	***E. verrucata* (Cs715)**	60.54	46.8	26.19	16.66	29.6	27.76	6.63	19.22	154.23	45.69	59.19	54.18	124.27	74.47	46.54	59.66
39	***E. virginea* (Cs727)**	34.81	67.76	41.67	19.81	26.92	52.45	11.93	27.61	91.28	34.81	71.28	50.86	68.88	59.29	41.4	54.68
40	***E. willisii* (Cs492)**	44.59	64.38	23.01	8.53	25.54	29.64	4.6	21.22	97.13	43.02	63.06	55.24	60.59	58.18	48.27	46.59
41	***Heteropyxis natalensis* (CANB576168)**	0.83	11.16	8.13	4.66	5.35	4.34	0.24	12.42	4.58	1.82	7.31	10.8	1.85	7.17	8.51	16.35
42	***Melaleuca capitata* (RDE176)**	142.59	399.34	418.07	214.83	288.71	334.49	172.68	316.56	379.76	312.92	333.74	404.28	594.15	389.1	385.57	393.8
43	***M. cornucopiae* (Harwood1546)**	314.45	457.72	380.57	225.37	301.7	400.51	203.04	467.63	494.84	502.92	334.39	346.33	471.81	465.95	439.34	595.37
44	***M. foliolosa* (AF4308)**	158.25	168.6	159.51	73.85	131.62	160.82	70.11	168.13	151.76	131.31	131.9	138.7	145.31	120.43	168.22	151.11
45	***M. squarrosa* (RDE75)**	106.48	225.48	215.79	113.04	178.76	249.34	91.11	171.31	409.65	206.59	268.44	243.66	282.16	459.35	264.75	257.97
46	***M. sylvana* (AF4286)**	141.78	254.8	302.41	89.71	139.89	199.82	90.12	182.42	197.4	152.53	182.29	226.08	225.44	162.81	235.56	199.04
47	***Calothamnus gracilis* (RDE154)**	123.52	162.17	136.69	67.09	101.84	128.49	141.85	120.45	159.11	325.33	95.77	121.7	103.76	140.53	137.07	132.01

Supplementary Table S1. (continued).

	Species	n_3214	n_3712	n_4206	n_4706	n_6015	n_6199	n_22854	n_23033	n_25721	n_30327	n_30471	n_30562	n_30839	n_31043	n_31075	n_31209
1	***A. gummiferum* (McCoys.n.)**	452.6	212.24	268.74	135.96	489.12	236.15	250.51	252.08	149.47	68.02	234.52	272.54	333.33	327.59	206.5	276.51
2	***A. gummiferum* (McCoys.n.)**	251.76	105.06	134.53	84.84	275.27	129.43	142.63	134.49	100.26	35.59	164.45	143.72	191.51	160.02	129.46	165.69
3	***Corymbia bunites* (Cs229)**	385.85	336.5	536.52	239.67	778.74	444.04	420.77	478.23	461.68	116.57	307.42	389.73	536.97	513.39	417.39	590.35
4	***C. calophylla* (Cs377)**	1.65	2.71	0.55	0.7	3.1	3.58	2.78	2.3	0.62	0	1.6	1.55	3.45	2.73	1.95	2.66
5	***C. haematoxylon* (Cs380)**	51.12	44.7	62.23	47.85	74.07	46.05	56.32	48.26	43.81	11.82	35.12	58.95	74.85	60.37	45.27	67.24
6	***C. peltata* (Cs493)**	316.68	248.84	394.19	153.94	541.52	297.75	330.3	333.46	340.34	92.63	241.62	283.99	398.3	431.5	325.52	432.97
7	***E. brockwayi* (Cs474)**	330.73	223.75	142.61	160.91	148.98	146.33	176.36	161.88	244.32	58.92	100.11	224.2	229.8	172.13	158.46	218.45
8	***E. coccifera* (ANBG9404806)**	305.56	144.62	157.9	140.93	136.28	142.01	177.5	141.68	483.01	66.09	202.93	171.46	210.08	176.83	137.02	157.82
9	***E. extrica* (CANB632680)**	130.85	122.03	186.57	79.43	159.66	156.69	149.13	135.99	103.83	19.56	80.85	125.14	184.78	149.11	252.15	171.07
10	***E. globoidea* (Cs729)**	91.33	84.52	101.06	69.38	92.34	136.69	121.53	79.69	48.91	25.9	59.9	114.11	146.53	105.14	85.17	128.7
11	***E. globulus* (CANB688145)**	330.13	138.08	134.11	92.06	144.98	145.47	142.46	142.6	387.01	11.74	86.99	189.61	192.73	144.1	113.98	152.53
12	***E. globulus* (CANB494954)**	70.86	39.43	35.54	32.91	40.94	39.58	51.88	38.38	121.33	1.34	29.42	37.21	57.94	35.14	22.74	50.13
13	***E. gomphocephala* (Cs357)**	229.98	137.24	127.13	106.92	148.6	154.79	157.99	153.37	235.02	32.39	120.75	177.17	199.21	164.65	127.45	160.66
14	***E. goniantha* (Cs424)**	2.4	2.3	1.47	2.15	2.78	3.52	4.69	4.43	1.49	4.71	0.92	3.5	2.59	1.96	1.97	1.61
15	***E. howittiana* (Cs496)**	90.54	80.75	89.37	65.81	92.87	91.58	99.05	99.35	142.63	24.59	68.63	113.68	124.74	90	74.24	114
16	***E. insularis* (Cs275)**	90.12	67.47	89.77	52.06	85.72	114.06	90.31	81.32	60.2	12.83	65.35	85.31	117.28	94.81	70.59	95.26
17	***E. insularis* (CANB413807)**	47.96	53.18	59.69	39.62	58.86	81.28	71.04	67.11	42.67	2.14	48.2	56.52	86.91	63.87	53.03	62.39
18	***E. leucoxylon* (Cs716)**	180.66	158.56	115.85	65.27	103.51	121.82	93.52	114.72	322.76	25.82	73.04	130.77	145	121.26	81.6	131.5
19	***E. ligulata* (Cs721)**	13.24	12.71	12.18	12.91	14.75	19.89	17.88	13.84	7.79	5.21	9.01	13.41	22.03	15.53	9.22	17.18
20	***E. moorei* (Cs574)**	194.46	186.31	192.8	127.36	195.7	321.15	210.75	153	125.67	28.45	105.31	185.55	259.54	209.84	165.85	213.48
21	***E. nitida* (Cs744)**	66.03	83.12	76.16	46.43	81.63	133.89	80.12	71.49	45.73	8.82	36.44	79.39	98.8	86.91	74.32	97.7
22	***E. optima* (Cs708)**	62.16	26.58	35.5	18.84	35.42	33.4	43.99	34.32	52.92	14.74	24.87	42.97	54.68	35.49	26.02	40.14
23	***E. pachycalyx* (Cs305)**	10.36	6.63	3.09	6.49	4.92	6.29	6.87	6.89	3.81	7.05	4.54	12.72	7.53	5.76	7.07	7.72
24	***E. pachyloma* (Cs426)**	44.25	63.92	53.73	36.37	61.68	93.88	56.2	55.35	38.84	3.54	42.32	68.62	80.57	58.48	49.05	70.22
25	***E. perriniana* (Cs742)**	84.54	43.65	46.48	46.05	54.2	48.76	54.48	46.19	124.83	8.97	50.33	62.85	61.34	52.67	46.35	65.33
26	***E. pilularis* (Cs482)**	178.88	171.57	189.95	103.36	202.41	285.25	214.21	183.94	138.09	72.63	126.9	203.27	274.59	205.94	171.59	239.95

27	***E. platydisca* (CANB638520)**	223.13	204.93	248.08	133.98	231.12	261.13	278.03	242.56	166.09	28.06	114.96	225.89	309.94	241.92	191.86	235.54
28	***E. pleurocarpa* (Cs431)**	93.4	93.8	77.6	58.93	102.45	81.27	111.36	101.56	79.41	25.37	46.05	84.29	128	122	161.7	103.77
29	***E. preissiana* (Cs425)**	21.16	12.69	12.27	9.24	12.76	13.79	56.89	10.59	22.11	1.81	15.21	16.28	19.68	15.23	12.13	21.31
30	***E. pumila* (Cs227)**	205.99	179.28	211.53	90.19	178.45	173.52	195.21	171.06	397.08	34.49	126.27	211.77	252.59	192.82	168.91	214.97
31	***E. selachiana* (CANB632673)**	48.13	40.17	38.99	32.26	47.73	54.95	47.45	40.73	39.91	9.92	19.51	40.22	71.5	52.16	78.39	62.04
32	***E. selachiana* (CANB693174)**	155.73	134.87	147.16	53.02	153.92	164.13	155.56	142.07	81.53	27.94	97.58	148.58	204.8	166.16	278.87	177.37
33	***E. spathulata* (Cs368)**	220.23	126.05	108.54	85.14	161.38	147.5	149.71	145.6	258.3	38.96	98.06	187.03	191.35	176.21	129.3	181.51
34	***E. staeri* (Cs366)**	29.02	24.2	26.54	16.7	31.16	37.22	41.76	29.38	15.27	12.87	27.71	50.16	37.6	35.89	24.38	29.19
35	***E. stenostoma* (Cs733)**	2.62	2.68	3.88	2.24	3.91	5.95	7.41	3.62	3.89	1.29	3.12	4.35	4.63	4.03	3.9	6.85
36	***E. stoatei* (Cs191)**	212.54	147.93	146.12	95.36	157.01	128.74	303.22	155.95	242.9	37.73	191.17	187.44	205.79	175.72	143.17	193.39
37	***E. tereticornis* (Cs769)**	508.21	198.57	232.82	181.67	240.1	241.63	276.3	266.2	589.19	53.21	143.15	320.13	336.42	247.53	204.94	305.47
38	***E. verrucata* (Cs715)**	52.12	61.32	52.81	37.5	66.64	100.96	60.52	50.32	36.17	11.76	33.99	67.88	80.86	55.96	55.91	71.14
39	***E. virginea* (Cs727)**	42.75	52.03	42.71	42.25	53.19	62.8	66.17	48.92	28.74	5.19	36.35	63.1	69.63	56.91	43.44	52.7
40	***E. willisii* (Cs492)**	56.25	53.99	42.35	32.35	50.89	75.69	67.8	46.78	30.64	10.56	52.31	55.24	65.93	50.85	50.51	51.75
41	***Heteropyxis natalensis* (CANB576168)**	17.35	1.82	9.57	4.26	22.44	6.37	29.33	14.46	3.53	0	1.59	7.24	10.32	15.31	10.73	13.41
42	***Melaleuca capitata* (RDE176)**	289.67	344.12	448.68	211.81	421.98	425.87	445	367.06	328.27	525.08	318.54	423.05	499.48	438.69	421.15	487.12
43	***M. cornucopiae* (Harwood1546)**	589.99	524.59	452.76	440.6	559.72	648.43	561.94	558.16	486.15	243.91	434.57	519.03	643.45	652.21	598.33	593.02
44	***M. foliolosa* (AF4308)**	243.23	185.67	191.76	96.31	201.02	204.58	202.26	202.68	132.3	79.11	142.11	191.17	258.64	232.09	212.16	253
45	***M. squarrosa* (RDE75)**	195.81	228.79	201.62	113	244.36	259.96	251.64	232.62	173.33	395.63	463.77	262.68	305.3	287.12	259.35	320.37
46	***M. sylvana* (AF4286)**	174.63	244.66	245.16	139.97	224.99	245.13	235.13	256.93	140.56	132.66	156.07	358.57	323.99	230.06	213.98	316.13
47	***Calothamnus gracilis* (RDE154)**	146.91	143.75	154.49	105.67	323.51	161.4	184.54	162.03	142.32	171.73	122.21	181.84	195.12	154.84	162.89	200.12

Supplementary Table S1. (continued).

	Species	n_31501	n_31750	n_31806	n_32269	n_32608	n_32708	n_33152	n_33440	n_33442	n_33580	n_33688	n_33888	n_34100	n_34416	n_34636	n_34751
1	***A. gummiferum* (McCoys.n.)**	242.81	210.04	206.88	242.23	340.29	203.24	274.31	224.93	209.83	205.11	246.25	169.51	265.49	267.65	270.46	197.32
2	***A. gummiferum* (McCoys.n.)**	162.96	137.98	114.6	132.77	207.5	128.07	147.76	113.29	111.97	104.71	121.98	78.69	121.83	134.23	186.93	112.82
3	***Corymbia bunites* (Cs229)**	445.15	374.67	299.92	573.02	468.97	369.17	470.73	122.32	411.09	311.51	392.26	342.33	386.19	503.6	287.08	332.48
4	***C. calophylla* (Cs377)**	1.97	2.48	0.43	4	3.54	1.33	1.76	0.74	1.71	1.56	1.41	2.07	2.5	1.26	1.83	1.79
5	***C. haematoxylon* (Cs380)**	66.13	52.65	41.93	83.12	59.25	43.18	53.32	17.56	53.92	51.82	46.79	43.54	45.28	60.94	35.52	48.07
6	***C. peltata* (Cs493)**	306.89	281.73	215.2	290.48	323.08	269.54	383.61	87.91	289.97	232.94	292.49	240.61	262.19	368.27	194.87	236.1
7	***E. brockwayi* (Cs474)**	198.53	170.63	125.24	152.97	212.32	136.34	165.79	120.01	196.55	163.01	154.1	119.14	163.55	171.45	232.12	147.86
8	***E. coccifera* (ANBG9404806)**	201.43	186.32	142.3	168.69	215.28	176.11	178.67	107.63	166.52	208.75	170.29	87.82	166.49	165.92	233.46	131.76
9	***E. extrica* (CANB632680)**	142.99	142.58	130.85	127.18	188.59	121.27	154.01	81.52	139.69	119.09	138.61	93.09	135.03	147.19	159.37	118.3
10	***E. globoidea* (Cs729)**	120.4	111.82	69.11	87.01	111.83	90.72	103.28	72.55	94.4	269.75	78.32	75.8	93.34	96.8	142.87	92.5
11	***E. globulus* (CANB688145)**	178.13	133.03	132.1	144.23	179.98	137.74	156.79	94.91	142.27	172.19	161.33	87.71	164.29	179.12	158.44	129.33
12	***E. globulus* (CANB494954)**	42.9	39.8	40.97	42.1	47.93	45.51	45.91	27.18	42.25	46.84	45.82	27.4	49.07	41.21	44.87	30.69
13	***E. gomphocephala* (Cs357)**	158.26	158.97	132.08	143.4	220.86	136.2	157.52	136.99	151.84	165.64	158.62	93.4	173.44	170.69	188.31	130.25
14	***E. goniantha* (Cs424)**	4.05	3.27	2.28	0.81	5.79	1.26	2.84	1.4	2.05	3.63	2.07	1.66	1.89	1.7	5.11	1.58
15	***E. howittiana* (Cs496)**	120.39	85.89	87.52	89.97	135.63	102.83	106.07	71.79	89.73	116.31	111.63	72.68	116.36	108.47	128.7	106.7
16	***E. insularis* (Cs275)**	100.52	85.05	70.78	89.77	93.8	75.17	92.96	51.32	76.28	111.67	71.57	50.29	67.57	100.51	178.26	86.03
17	***E. insularis* (CANB413807)**	76.74	63.83	51.26	54.83	80.38	56.39	66.96	41.12	55.32	89.63	61.15	42.78	47.05	69.97	105.72	53.33
18	***E. leucoxylon* (Cs716)**	139.6	125.84	96.74	123.73	195.58	100.15	140.59	106.08	119.71	140.37	136	64.28	141.82	151.69	149.63	118.55
19	***E. ligulata* (Cs721)**	15.5	21.98	13.8	11.96	13.45	12.16	15.22	14.05	11.32	18.32	13.24	9.07	9.47	15.28	37.54	12.33
20	***E. moorei* (Cs574)**	216.55	198.66	183.4	171.29	215.15	177.64	223.85	142.54	184.23	336.39	179.71	122.05	187.71	194.79	363.91	180.56
21	***E. nitida* (Cs744)**	86.14	88.94	72.23	74.36	85.05	67.43	87.28	58.54	68.63	141.77	66.06	58.9	63.95	64.43	116.04	62.38
22	***E. optima* (Cs708)**	52.08	32.34	25.43	38.68	39.19	30.94	42.06	36.29	41.7	36.29	35.15	26.71	42.73	45.9	90.69	39.92
23	***E. pachycalyx* (Cs305)**	8.56	10.69	3.78	4.38	6.68	9.21	7.32	5.02	6.26	8.68	7.58	5.53	7.46	9.5	8.06	8.09
24	***E. pachyloma* (Cs426)**	61.18	61.81	43.05	57.54	91.34	56.29	56.11	43.17	54.15	70.18	51.34	35.7	60.4	53.11	115.2	46.31
25	***E. perriniana* (Cs742)**	66.08	49.69	52.94	52.6	75.32	49.54	56.94	35.42	50.99	72.21	63.02	32.28	50.77	52.72	68.78	56.86
26	***E. pilularis* (Cs482)**	217.97	223.15	170.22	179.31	251.34	194.68	230.89	142.32	177.39	317.6	167.13	146.05	179.39	174.55	326.99	172.21

27	***E. platydisca* (CANB638520)**	255.42	226.1	213.64	218.87	258.22	212.13	248.05	151.49	224.18	313.18	213.89	114.46	230.11	263.36	332.9	185.08
28	***E. pleurocarpa* (Cs431)**	104.92	94.81	91.85	95.64	118.44	86.46	112.15	62.24	89.36	88.04	86.94	56.07	82	112.32	111.24	84.05
29	***E. preissiana* (Cs425)**	19.27	14.75	11.27	15.26	18.37	17.96	16.75	12.54	17.15	18.1	20.23	12.21	14.46	13.49	35.58	15.6
30	***E. pumila* (Cs227)**	217.58	188.01	173.98	197.93	248.96	167.24	213.31	130.19	208.75	267	185.47	112.27	214.75	230.03	239.74	186.04
31	***E. selachiana* (CANB632673)**	48.08	46.91	34.46	45.47	47.73	41.12	59.58	31.59	44.77	41.26	35.49	34.85	38.97	49.06	53.71	45.21
32	***E. selachiana* (CANB693174)**	169.62	160.48	113.79	125.78	188.29	134.99	172.5	113.55	141.67	136.52	131.29	94.07	149.25	191.54	195.77	139.67
33	***E. spathulata* (Cs368)**	168.89	171.19	138.02	145.32	188.43	124.29	166.4	140.36	156.47	173.65	179.91	83.95	156.22	174.41	188.21	126.7
34	***E. staeri* (Cs366)**	45.14	21.69	21.87	25.7	31.8	25.26	34.44	18.12	24	44.66	28.15	22.01	32.96	26.15	59.11	28.16
35	***E. stenostoma* (Cs733)**	4.23	5.35	3.07	3.56	8.07	4.28	3.97	2.77	3.42	5.91	5	5.21	2.99	8.07	11.79	4.88
36	***E. stoatei* (Cs191)**	188.07	161.21	114.46	154.5	183.16	150.73	168.56	142.27	149.91	214.77	181.85	114.54	172.96	181.74	369.67	150.17
37	***E. tereticornis* (Cs769)**	292.2	262.27	244.24	262.92	325.98	240.83	270.49	199.74	290.02	406.88	277.93	196.61	272.86	279.64	327.69	233.97
38	***E. verrucata* (Cs715)**	82.31	69.04	51.41	51.57	97.74	53.68	71.05	41.04	58.9	154.5	54.34	35.76	61.33	75.1	79.96	55.12
39	***E. virginea* (Cs727)**	59.36	59.2	39.78	47.88	54.39	49.08	51.45	35.17	49.16	76.6	40.6	38.35	45.59	67.84	71.28	51.68
40	***E. willisii* (Cs492)**	67.95	55.44	47.66	44.81	68.91	50.03	52.91	37.49	46.22	108.67	45.3	24.5	44.81	79.18	70.58	44.26
41	***Heteropyxis natalensis* (CANB576168)**	3.8	27.5	4.85	7.12	2.49	3.57	11.12	3.47	8.79	1.38	12.77	4.55	8.59	157.9	2.94	3.67
42	***Melaleuca capitata* (RDE176)**	386.3	499.31	237.54	315.21	498.88	262.01	381.24	235.83	308.38	407.78	437.26	340.53	311.92	390.92	328.94	328.65
43	***M. cornucopiae* (Harwood1546)**	532.96	465.36	311.76	421.89	795.4	321.36	568.03	268.09	382.4	509.28	480.79	420.19	415.89	633.11	492.81	432.93
44	***M. foliolosa* (AF4308)**	190.22	163.81	113.06	153.65	243.55	113.8	189.07	93.63	149.45	175.88	206.4	165.43	151.6	231.7	161.89	150.1
45	***M. squarrosa* (RDE75)**	264.13	281.41	133.93	195.98	298.44	241.42	248.26	161.3	192.27	244.3	257.35	213.62	275.3	245.73	222.95	180.2
46	***M. sylvana* (AF4286)**	222.23	174.45	119.25	178.33	234.64	154.67	227.94	107.54	172.11	216.17	218.92	205.45	195.8	265.32	233.26	180.96
47	***Calothamnus gracilis* (RDE154)**	157.54	113.12	88.05	129.16	179.91	107.12	176.65	71.86	98.54	159.55	146.83	131.86	112.25	167.62	117.98	121.69

Supplementary Table S1. (continued).

	Species	n_35801	n_40263	n_49139	n_BI096866	n_BI096871	n_BI096904	n_BI097026	n_3075	n_3659	n_4167	n_4178	n_4460	n_4589	n_5190	n_5930	n_27210
1	***A. gummiferum* (McCoys.n.)**	159.79	247.94	270.58	198.51	192.68	114.16	161.77	221.72	369.51	271.14	242.26	239.88	127.78	237.44	283.74	322.44
2	***A. gummiferum* (McCoys.n.)**	82.61	141.19	143.2	117.09	101.97	75.57	102.96	127.14	204.75	146.78	142.18	141.66	68.96	150.85	145.91	190.7
3	***Corymbia bunites* (Cs229)**	256	373.75	482.62	474.78	334.5	256.33	396.97	488.42	716.18	466.73	403.71	456.63	269.28	440.15	484.05	611.58
4	***C. calophylla* (Cs377)**	1.17	3.17	1.7	3.73	1.39	0.57	1.3	2.19	3.82	3.15	1.55	3.38	1	2.85	1.15	2.96
5	***C. haematoxylon* (Cs380)**	35.23	49.67	72.05	53.8	46.95	36.88	34.38	62.35	86.07	61.65	50.94	60.45	34.14	61.31	67.97	75.99
6	***C. peltata* (Cs493)**	193.78	274.12	389.36	273.84	265.55	168.23	206.33	365.71	503.56	348.36	307.72	339.69	174.43	314.6	373.85	454.6
7	***E. brockwayi* (Cs474)**	97.09	144.23	179.49	130.55	174.39	150.63	203.93	218.92	288.64	399.25	132.09	175.25	108.58	322.21	177.02	216.84
8	***E. coccifera* (ANBG9404806)**	96.37	151.82	145.01	125.53	152.73	106.62	243.08	154.76	226.5	310.57	131.88	155.69	80.98	175.79	193.83	194.72
9	***E. extrica* (CANB632680)**	101.03	172.99	148.24	72.33	123.19	67.23	100.4	169.16	199.3	268.29	477.4	148.01	66.14	188.11	170.67	154.93
10	***E. globoidea* (Cs729)**	52.71	90.34	113.07	89.3	109.27	90.25	63.97	111.06	167.14	723.5	80.7	107.93	74.07	106.68	91.41	130.83
11	***E. globulus* (CANB688145)**	103.74	137.6	162.04	64.62	117.77	89.03	207.09	160.24	212.22	337.41	138.92	148.74	77.53	234.97	164.55	174.07
12	***E. globulus* (CANB494954)**	22.04	40.14	38.43	27.81	35.68	31.11	54.62	37.26	53.64	79.58	34.71	41.57	14.27	54.72	53.95	49.8
13	***E. gomphocephala* (Cs357)**	77.09	142.13	152.22	104.7	137.13	85.78	227.68	175.97	236.11	282.51	129.47	150.46	90.83	359.51	155.47	178.57
14	***E. goniantha* (Cs424)**	2.78	3.3	2.31	0.8	2.88	1.36	2.2	3.16	3.79	6.98	1.56	2.91	0.79	2.79	2.79	3.03
15	***E. howittiana* (Cs496)**	64.53	78.56	95.06	90.35	105.74	69.99	139.07	98.47	158.48	126.3	77.41	102.17	72.67	105.98	100.01	122.7
16	***E. insularis* (Cs275)**	37.91	74.25	93.79	65.57	72.74	50.26	66.32	84.64	107.56	278.08	69.03	88.01	54.29	89.3	97.24	101.39
17	***E. insularis* (CANB413807)**	26.54	47.34	68.7	35.78	42.06	32.44	40.73	60.27	91.96	157.49	48.29	71.01	29.38	58.48	69.13	68.54
18	***E. leucoxylon* (Cs716)**	52.15	104.92	116.22	83.84	115.14	75.1	156.76	120.51	201.55	226.6	101.42	126.28	74.54	168.73	126.69	137.8
19	***E. ligulata* (Cs721)**	7.07	12.14	16.17	18.74	16.66	11.73	10.18	16.47	20.99	87.25	10.01	17.29	10.38	18.19	17.67	19.24
20	***E. moorei* (Cs574)**	121.57	204.59	199.62	140.42	167.16	108.9	134.21	180.77	277.58	905.67	158.41	189.02	98.5	197.62	223.9	223.55
21	***E. nitida* (Cs744)**	45.68	77.12	81.97	65.52	82.83	54.62	50.2	93.19	117.18	328.25	68.92	90.17	46.28	82.61	86.78	93.16
22	***E. optima* (Cs708)**	22.4	32.2	36.29	42.13	37.62	30.94	51.61	37.98	52.18	50.59	30.79	41.06	20.58	49.59	43.9	50.45
23	***E. pachycalyx* (Cs305)**	3.2	7.34	9.48	7.9	6.63	5.58	10.8	7	8.27	13.66	4.97	5.57	6.51	13.91	5.94	7.34
24	***E. pachyloma* (Cs426)**	29.99	53.23	56.68	40.47	53.85	48.09	38.27	66.16	87.4	331.04	50.78	68.26	30.7	57.8	63.68	85.47
25	***E. perriniana* (Cs742)**	27.75	53.79	50.56	37.48	42.21	36.72	87.48	50.45	76.14	118.87	47.29	58.78	23.43	50.8	67.3	62.85
26	***E. pilularis* (Cs482)**	102.31	235.13	212.57	124.96	185.29	116.84	129.63	234.26	321.19	872.57	164.11	217.87	113.38	208.14	229.98	485.29

27	***E. platydisca* (CANB638520)**	137.13	217.19	220.06	163.05	170.38	107.67	162.1	239.34	335.16	877.99	194.48	226.41	96.61	223.6	240.89	285.65
28	***E. pleurocarpa* (Cs431)**	58.62	147.56	93.43	48.47	69.82	51.93	78.56	126.06	136.16	174.36	246.11	93.01	48.4	142.86	104.75	111.02
29	***E. preissiana* (Cs425)**	7.78	17.32	13.51	13.31	16.51	9.93	24.96	16.02	24.52	47.63	15.15	19.19	11.99	17.94	18.51	17.28
30	***E. pumila* (Cs227)**	119.76	202.71	186.18	130.69	156.62	118.1	295.42	194.77	282.65	490.48	158.64	188.63	115.04	187.21	206.2	258.32
31	***E. selachiana* (CANB632673)**	30.47	65.79	57.28	23.79	34.85	17.98	26.96	53.72	63.09	105.96	149.17	56.87	23.85	65.77	47.01	71.02
32	***E. selachiana* (CANB693174)**	91.82	231.73	152.18	129.73	160.37	103.59	122.73	142	202.28	350.28	428.38	156.15	82.55	213.62	164.8	179.4
33	***E. spathulata* (Cs368)**	93.28	167.45	144.43	96.75	108.94	89.98	215.85	177.71	223.96	295.53	139.14	158.75	86.33	256.39	153.02	169.67
34	***E. staeri* (Cs366)**	19.45	26.47	32.96	35.44	33.74	19.72	20.46	34.16	54.85	176.17	21.05	23.54	20.36	32.3	32.45	31.16
35	***E. stenostoma* (Cs733)**	3.96	4.28	4.67	2.45	7.33	4.36	2.34	4.53	8.17	13.8	4.12	4.11	0.61	4.76	3.68	4.67
36	***E. stoatei* (Cs191)**	85.41	164.71	163.51	128.96	177.37	127.38	223.37	168.37	218.39	159.02	153.77	172.06	118.54	412.85	185.87	173
37	***E. tereticornis* (Cs769)**	169.4	273.62	272.57	172.9	231.17	174.88	301.22	285.61	377.28	463.24	216.56	263.24	157.76	411.83	285.85	317.76
38	***E. verrucata* (Cs715)**	33.13	53.17	65.42	49.37	54.89	31.62	33.2	67.95	80.63	238.66	53.74	71.71	31.28	68.6	70.72	76
39	***E. virginea* (Cs727)**	20	53.08	61.73	44.15	42.82	42.68	35.45	63.85	82.22	183.55	46.35	58.35	30.24	70.49	55.48	57.85
40	***E. willisii* (Cs492)**	34.38	65.55	51.19	31.89	36.54	33.04	37.73	49.66	73.24	256.8	51.96	57	26.69	50.85	54.72	71.74
41	***Heteropyxis natalensis* (CANB576168)**	2.63	10.52	7.24	8.73	0.68	4.12	6.75	4.54	11.63	10.02	11.87	11.98	4.05	6.42	6.36	3.73
42	***Melaleuca capitata* (RDE176)**	168.11	356.82	483.12	318.39	420.21	368.52	306.62	513.85	807.47	398.41	353.16	444.82	267.52	462.1	505.98	517.4
43	***M. cornucopiae* (Harwood1546)**	141.67	496.07	552.57	597.57	456.9	409.04	457.38	466.35	698.63	558.83	582.98	672.75	359.9	586.72	790.07	626.95
44	***M. foliolosa* (AF4308)**	72.35	183.34	240.07	191.58	191.35	154.71	133.32	197.01	326.73	186.85	187.2	202.95	318.64	257.04	234.05	212.93
45	***M. squarrosa* (RDE75)**	96.15	230.33	286.64	219.19	264.42	240.57	195.08	325.75	448.45	251.21	232.53	253.68	135.64	307.71	308.61	340.37
46	***M. sylvana* (AF4286)**	80.45	202.26	187.99	191.97	211.49	230.2	186.45	302.59	405.68	271.29	172.07	249.05	239.97	281.88	250.8	311.66
47	***Calothamnus gracilis* (RDE154)**	43.77	141.11	191.91	148.3	142.52	116.79	138.4	139.12	224.93	154.88	144.2	156.88	89.58	180.71	198.11	314.11

Supplementary Table S1. (continued).

	Species	n_40152	n_40685	n_40951	n_41116	n_41171	n_41451	n_41541	n_41647	n_41657	n_42256	n_42353	n_42511	n_42576	n_42717	n_42723	n_hdr
1	***A. gummiferum* (McCoys.n.)**	200.89	218.97	407.82	335.92	182.75	293.03	158.32	188.26	330.18	212.76	173.95	194.29	134.85	270.2	119.5	277.5
2	***A. gummiferum* (McCoys.n.)**	121.15	120.14	207.99	150.89	111.68	169.91	92.86	87.16	195.46	121.54	90.37	143.46	65.79	156.3	70.9	162.99
3	***Corymbia bunites* (Cs229)**	379.54	371.08	589.09	946.18	408	601.34	286.58	235.69	434.76	364.51	313.19	334.26	250.63	527.06	227.62	814.65
4	***C. calophylla* (Cs377)**	1.77	1.33	3.1	5.12	0.66	4.08	1.49	1.53	3	3.22	1.42	1.65	2.05	2.46	2.61	3.67
5	***C. haematoxylon* (Cs380)**	49.63	52.99	90.2	104.84	34.13	84.27	34.44	41.02	97.7	46.06	43.73	36.65	33.2	67.76	26.07	58.72
6	***C. peltata* (Cs493)**	252.89	238.91	448.56	618.8	282.84	463.76	194.46	203.08	332.94	261.49	227	233.31	176.3	399.56	165.27	599.27
7	***E. brockwayi* (Cs474)**	173.46	186.89	241.81	216.93	170.72	172.92	120.53	111.89	424.5	242.73	143.15	357.43	124.93	170.49	704.41	187.26
8	***E. coccifera* (ANBG9404806)**	145.87	161.71	232.51	204.15	129.54	219.7	109.01	124.11	301.07	141.93	327.37	308.15	137.49	190.61	347.64	179.27
9	***E. extrica* (CANB632680)**	141.67	162.29	203.4	234.95	106.21	156.33	292.25	86.13	232.92	142.29	147.98	135.98	148.83	160.89	239.39	179.95
10	***E. globoidea* (Cs729)**	112.76	123.75	150.07	139.82	107.15	124.32	77.94	77.34	216.44	103.31	84.37	237.55	99.8	95.88	269.33	147.98
11	***E. globulus* (CANB688145)**	134.7	156.87	199.95	180.85	97.17	178.39	88.93	86.59	269.73	133.25	192.13	420.58	106.93	174.56	232.72	146.12
12	***E. globulus* (CANB494954)**	44.92	49.43	53.26	63.69	32.06	51.58	29.44	29.49	72.62	36.56	64.17	76.56	28.51	47.31	70.47	50.89
13	***E. gomphocephala* (Cs357)**	137.96	174.91	231.86	212.54	97.49	198.38	103.9	106.04	280.99	169.8	151.56	620.8	113.83	173.06	295.44	176.5
14	***E. goniantha* (Cs424)**	1.97	3.23	2.49	4.84	1.37	3.09	2.11	1.43	6.28	1.53	0.89	3.63	3.85	2.27	6.03	3.96
15	***E. howittiana* (Cs496)**	83.26	100.2	137.07	144.66	69.21	118.41	79.16	79.74	158.21	88.35	247.03	185.67	81.82	98.58	224.12	110.24
16	***E. insularis* (Cs275)**	86.38	89.39	126.93	127.06	80.4	106.63	64.1	66.94	169.26	100.79	65.4	86.91	95.75	96.61	199.93	132.66
17	***E. insularis* (CANB413807)**	42.84	53.52	73.38	92.06	50.9	74.2	45.54	39.91	137.54	63.19	41.45	58.02	51.64	65.89	128.22	72.52
18	***E. leucoxylon* (Cs716)**	119.71	118.23	177.59	169.55	86.03	130.01	75.25	79.91	169.6	105.05	136.49	662.69	98.37	144.78	237.34	135.85
19	***E. ligulata* (Cs721)**	15.81	14.33	19.31	14.96	9.93	15.26	10.94	14.82	37.58	15.98	10.73	13	25.73	14.16	34.72	20.15
20	***E. moorei* (Cs574)**	184.13	258.16	285.74	271.26	160.9	244.56	135.71	99.66	324.05	199	148.5	278.54	229.02	233.92	265.34	276.87
21	***E. nitida* (Cs744)**	76.97	108.15	121.38	99.63	69.42	94.02	47	44.62	125.33	80.68	65.38	339.78	68.21	88.82	106.97	121.2
22	***E. optima* (Cs708)**	43.2	35.82	50.14	40.66	30.92	46.21	37.05	33.84	89.7	45.14	29.64	73.03	27	39.87	194.85	51.6
23	***E. pachycalyx* (Cs305)**	6.71	7.22	9.47	5.37	5.8	8.11	4.98	4.9	9.47	5.59	7.06	6.4	6.27	6.08	11.18	9.16
24	***E. pachyloma* (Cs426)**	59.23	66.96	90	97.63	42.26	69.1	38.63	38.59	124.99	49.82	49.39	70.49	71.3	59.45	126.13	76.76
25	***E. perriniana* (Cs742)**	54.81	55.3	78.12	60.93	37.33	65.6	35.01	39.47	112.3	45.38	55.19	163.25	34.56	58.22	111.54	85.88
26	***E. pilularis* (Cs482)**	208.35	212.9	315.28	285.56	167.97	237.04	150.21	128.8	430.02	189.26	170.33	365.52	217.09	234.16	480.5	284.46

27	***E. platydisca* (CANB638520)**	236.98	245.41	342.07	358.19	203.1	296.2	137.53	137.97	402.4	207.89	192.89	223.57	257.81	265.94	389.13	344.37
28	***E. pleurocarpa* (Cs431)**	93.06	101.92	148.62	132.82	78.23	119.64	202.16	61.76	184.45	109.34	134.9	109.6	100.32	110.96	218.95	106.67
29	***E. preissiana* (Cs425)**	11.01	13.15	23.28	20.83	10.51	15.94	12.55	14.66	60.85	17.35	31.37	46.05	11.24	16.26	39.72	19.95
30	***E. pumila* (Cs227)**	164.03	191.52	251.26	282.52	148.56	228.43	122.53	87.6	412.9	213.84	255.01	1047.82	148.65	204.55	346.17	218.74
31	***E. selachiana* (CANB632673)**	45.63	62.46	69.57	66.6	35.83	62.56	86.92	36.69	88.17	41.54	50.78	37.62	46.85	54.58	92.6	59.78
32	***E. selachiana* (CANB693174)**	153.19	144.66	233.66	202.24	112.76	192.36	358.43	91.69	288.95	141.63	109.79	150.58	155.57	175.66	438.77	172.63
33	***E. spathulata* (Cs368)**	141.78	167.82	249.75	238.75	102.92	163.29	111.98	98.86	272.98	150.41	114.43	364.35	139.12	188.74	330.34	177.86
34	***E. staeri* (Cs366)**	27.4	32.99	36.09	54.94	22.47	36.35	24.3	26.44	56.35	41.45	26.79	58.27	34.83	30.23	81.74	37.73
35	***E. stenostoma* (Cs733)**	6.1	4.43	4.73	7.81	4.9	5.25	3.56	3.64	8.22	4.33	3.21	5.19	3.87	4.77	6.56	5.96
36	***E. stoatei* (Cs191)**	158.47	177.06	249.65	207.88	107.63	190.33	108.99	140.67	470.17	152.18	264.78	953.15	153.9	164.2	391.93	174.51
37	***E. tereticornis* (Cs769)**	263.46	248.64	344.59	330.6	226.29	289.15	180.2	193.36	473.65	263.32	392.16	717.35	205.29	308.27	756.97	291.28
38	***E. verrucata* (Cs715)**	49.3	69.11	75	104.89	54.11	74.19	40.63	36.42	85.15	71.87	40.27	143.09	38.59	68.9	122.22	74.53
39	***E. virginea* (Cs727)**	47.99	56.36	75.01	89.06	42.78	59.38	46.61	37.27	120.74	59.04	53.06	59.49	41.4	50.15	137.8	71.26
40	***E. willisii* (Cs492)**	46.48	60.93	66.92	82.51	29.83	66.26	26.34	31.06	76.25	56.68	41.67	78.73	52.99	59.86	70.22	67.29
41	***Heteropyxis natalensis* (CANB576168)**	2.16	3.63	10.42	3.87	2.72	6.06	6.09	2.27	0.12	9.6	3.53	24.88	3.45	11.8	1.38	9.69
42	***Melaleuca capitata* (RDE176)**	349.83	203.38	631.85	599.75	325.57	255.32	208.9	506.23	25.14	250.9	226.53	340.85	160.73	373.89	199.48	302.75
43	***M. cornucopiae* (Harwood1546)**	457.63	205.05	828.45	774.97	387.99	441.34	316.85	638.31	56.61	325.09	340.24	571.77	285.78	682.59	233.5	378.96
44	***M. foliolosa* (AF4308)**	186.37	101.77	271.61	290.02	157.76	147.94	88.98	190.37	22.14	122.9	123.96	175.69	73.38	205.43	68.27	117.07
45	***M. squarrosa* (RDE75)**	213.93	133.12	391.92	368.98	178.81	162.41	120.41	305.9	15.25	179.22	156.31	189.18	99.83	244.5	113.74	166.46
46	***M. sylvana* (AF4286)**	207.27	115.46	339.54	290.72	198.53	129.47	138.69	229.83	25.96	149.07	140.3	203.4	97.39	224.04	104.55	174.05
47	***Calothamnus gracilis* (RDE154)**	147.65	131.3	224.73	208.78	123.72	126.84	65.45	156.69	17.39	119.84	96.59	150.54	72.66	183.57	60.57	95.16

Supplementary Table S1. (continued).

	Species	n_idi1	n_BI096903	n_4744
1	***A. gummiferum* (McCoys.n.)**	653.59	125.8	466.14
2	***A. gummiferum* (McCoys.n.)**	445.01	65.99	294.71
3	***Corymbia bunites* (Cs229)**	466.81	232.96	996.96
4	***C. calophylla* (Cs377)**	3.22	1.51	7.39
5	***C. haematoxylon* (Cs380)**	66.76	19.47	132.7
6	***C. peltata* (Cs493)**	302.22	127.88	723.9
7	***E. brockwayi* (Cs474)**	169.31	165.94	292
8	***E. coccifera* (ANBG9404806)**	183.68	154.69	261.12
9	***E. extrica* (CANB632680)**	146.08	117.9	276.33
10	***E. globoidea* (Cs729)**	109.94	118.39	186.25
11	***E. globulus* (CANB688145)**	159.55	139.82	243.95
12	***E. globulus* (CANB494954)**	45.78	45.55	77.85
13	***E. gomphocephala* (Cs357)**	141.79	112.66	266.13
14	***E. goniantha* (Cs424)**	3.9	0.99	5.01
15	***E. howittiana* (Cs496)**	105.85	86.75	175.67
16	***E. insularis* (Cs275)**	102.62	77.7	153.89
17	***E. insularis* (CANB413807)**	62.67	56.85	106.79
18	***E. leucoxylon* (Cs716)**	155.3	105.59	218.54
19	***E. ligulata* (Cs721)**	13.78	13.52	20.6
20	***E. moorei* (Cs574)**	210.99	165.02	369.64
21	***E. nitida* (Cs744)**	75.51	68.49	114.89
22	***E. optima* (Cs708)**	42.27	35.73	65.65
23	***E. pachycalyx* (Cs305)**	10.11	9.72	10.54
24	***E. pachyloma* (Cs426)**	52.28	59.69	106.38
25	***E. perriniana* (Cs742)**	52.88	45.05	81.19
26	***E. pilularis* (Cs482)**	188.98	178.54	352.29

27	***E. platydisca* (CANB638520)**	219.48	197.8	385.89
28	***E. pleurocarpa* (Cs431)**	91.95	56.58	182.09
29	***E. preissiana* (Cs425)**	24.7	10.73	28.83
30	***E. pumila* (Cs227)**	209.13	152.71	366.01
31	***E. selachiana* (CANB632673)**	43.98	30.8	85.16
32	***E. selachiana* (CANB693174)**	157.47	113.45	270.56
33	***E. spathulata* (Cs368)**	207.76	121.37	260.73
34	***E. staeri* (Cs366)**	23.74	35.87	52.72
35	***E. stenostoma* (Cs733)**	6.37	3.74	6.56
36	***E. stoatei* (Cs191)**	178.42	137.11	291.31
37	***E. tereticornis* (Cs769)**	415.98	241.4	438.46
38	***E. verrucata* (Cs715)**	57.51	57.16	114.97
39	***E. virginea* (Cs727)**	52.25	45.6	94.98
40	***E. willisii* (Cs492)**	50.64	48.77	91.97
41	***Heteropyxis natalensis* (CANB576168)**	14.02	1.33	35
42	***Melaleuca capitata* (RDE176)**	559.43	253.54	790.95
43	***M. cornucopiae* (Harwood1546)**	559.08	213.08	982.1
44	***M. foliolosa* (AF4308)**	211.69	79.59	372.6
45	***M. squarrosa* (RDE75)**	310.27	111.55	513.6
46	***M. sylvana* (AF4286)**	229.84	116.99	484.61
47	***Calothamnus gracilis* (RDE154)**	222.34	59.13	342.2

Supplementary laboratory protocol for exon capture

The overall work flow is based on Rohland & Reich (2012) and Maricic *et al.* (2010). Modifications are **highlighted in Bold**.

Summary of workflow

1. Shearing
2. 1.6 X Sera-Mag Bead clean up
3. Blunt end repair
4. 2 X Sera-Mag Bead clean up
5. Adapter ligation
6. 1.6 X Sera-Mag Bead clean up
7. Nick fill in
8. 1.6 X Sera-Mag Bead clean up
9. Enrichment PCR
10. Sample quantification and pooling
11. 1.2 X Sera-Mag Bead clean up
12. Hybridisation
13. Recovery and wash
14. Semi-qPCR
15. Indexing PCR
16. 1.2 X Sera-Mag Bead clean up
17. Gel run and Qubit quantification
18. Final pooling of indexed library

Reagents

Agarose and agarose gel reagents

Bst large fragment polymerase

DNA ladder (100 bp, e.g. Axygen cat.no. M-DNA-100bp)

Dynabeads M-270 Streptavidin (Invitrogen cat.no. 65305)

Ethanol 100%

Orange loading dye

HotStarTaq Plus DNA polymerase (Qiagen cat.no. 203603)

Human COT-1 DNA (Invitrogen cat.no. 15279-011)

Milli Q water

Phusion Hot Start II DNA polymerase (Thermo Scientific cat.no. F-549S)

Qubit dsDNA HS Assay Kit (Invitrogen cat.no. Q32854)

Quick Blunting Kit (New England BioLabs cat.no. E1201L)

Quick Ligatioin Kit (New England BioLabs cat.no. M2200L)

SeqCap EZ Developer Library (NimbleGen 06471714001)

SeqCap Developer Reagent (NimbleGen 06684335001)

SeqCap Hybridisation and Wash Kit (NimbleGen 05634261001)

Sera-Mag SpeedBeads carboxylate-modified particles (GE Healthcare Life cat.no. 65152105050250)

Tris (10 mM, pH 8.5)

Equipment

Agarose gel electrophoresis unit

DNA shearing device (e.g. Bioruptor NGS, Diagenode)

PCR plates (96-Well, 200 µL) and seals

Filter tips

Ice

LabChip GX II (Caliper Life Sciences)

LabChip DS system (Caliper Life Sciences)

Ring magnet stand (e.g. SPRIplate 96-Ring Agencourt)

PCR machines

PCR 8-tube strips and caps (e.g. Axygen cat. no. PCR-02-FCP-C, PCR-0208-C)

Pipettor (single-channel, multi-channel)

Qubit 2.0 Fluorometer (Invitrogen 1011007409)

Speed vacuum concentrator

Tube (Bioruptor 0.65 mL mitrotubes for DNA shearing cat.no. C30010011)

Tube (microcentrifuge, 1.5 mL e.g. Eppendorf)

Tube (Qubit assay Invitrogen Q32856 500 µL)

Vortex mixer

Reagent preparations

Preparation of Sera-Mag Bead (See Faircloth & Glenn 2014)

10 X Blunting buffer (See Rohland & Reich 2012, 100 mM $MgCl_2$, 500 mM Tris-HCl pH 7.5, 75 mM DTT, dNTPs)

Preparation of barcoded-P5-adapter

Preparation PE-P7-adapter

Note

For the optimal results, it is suggested that the whole procedures are done consecutively by minimizing the time samples being stored in the freezer.
The library preparation was done on a PCR plate and the TIMING is per plate (96 samples) based.

Workflow

1. Shearing of Diluted DNA TIMING 8h

Shear 100 µL of 6 ng/µL of diluted genomic DNA using Bioruptor ultrasonicator using 30 cycles of 30 sec on and off with intensity set to H(igh).

2. 1.6X Sera-Mag Bead Clean Up TIMING 2 h (from Faircloth & Glenn 2014; Rohland & Reich 2012)

a) Pull the Sera-Mag Beads out 20 min in advance so that the beads are at room temperature. Vortex the beads before use.
b) Add 1.6 X volumes (160 µL) of Sera-Mag Beads to each well of a PCR plate using a multi-channel pipettor. Pipette mix 10 X.
c) Incubate the sample at room temperature for 10 min in dark.
d) Put the PCR plate on the ring magnet stand and wait until the liquid goes clear.
e) Remove and discard the supernatant.
f) Remove the magnetic plate.
e) Add 200 µL of fresh 80% ethanol and leave in for 1 min on a magnetic plate. Remove and discard the supernatant.
f) Repeat the ethanol wash in total of two washes.
g) Incubate the samples in a fume hood for 10-15 min uncovered until the pellet is dry.
h) Add 24.63 µL of 10 mM Tris (pH 8.5) and resuspend by pipette mixing 10 X.
i) Incubate at room temperature for 5 min. Spin down.
j) Put the PCR plate on the ring magnet stand and incubate at room temperature for 5 min.
k) Transfer 23.63 µL of the cleaned up DNA to a new PCR plate.

3. Blunt End Repair TIMING 1.5 h (from Rohland & Reich 2012)

a) Prepare the mix as below. Perform the following reaction in a 96-well PCR plate.

Reagent	Volume (µL) per sample	Final concentration
Sheared DNA (0.6 µg)	23.63	
dNTPs (25 mM)	0.12	0.1 mM
ATPs (10 mM)	**3**	**1 mM**

Blunting buffer (10 X)	3	1 X
Blunting enzyme mix	0.25	
Total volume	30 µL	

b) Incubate for 20 min at 12 °C, followed by 15 min at 37 °C.

4. 2X Sera-Mag Bead Clean Up TIMING 2 h

Perform the same as the first Sera-Mag Bead clean up except initially adding 2 X (60 µL) Sera-Mag Beads and using 70% EtOH for a total of two washes. Elute in 11.66 µL MiliQ water, and transfer 10.66 µL at the final step.

5. Adapter Ligation TIMING 1.5 h (from Rohland & Reich 2012)

a) Create the following mix. Perform the following reaction in a 96-well PCR plate.

Reagent	Volume (µL) per sample	Final concentration
Blunted DNA	20.34	
Quick Ligase Buffer (2 X)	15	0.75 X
Quick T4 DNA Ligase	**1**	
Barcoded-P5-adapter	1.67	
PE-P7-adapter	1.67	
Total volume	30 µL	

b) Incubate for 30 min at 22-24 °C.

6. 2X Sera-Mag Bead Clean Up TIMING 2h

Perform the same as the first Sera-Mag Bead clean up except initially adding 2 X (60 µL) Sera-Mag Beads and use 70% EtOH for the two times washes. Elute in 27.63 µL MiliQ water, and transfer 26.63 µL at the final step.

7. Nick Fill-in TIMING 1.5h (from Rohland & Reich 2012)

a) Prepare the master mix as below. Perform the following reaction in a 96-well plate.

Reagent	Volume (µL) per sample	Final concentration
Ligated DNA	26.63	
dNTPs (25 mM)	0.3	4 mM
ThermoPol Reaction Buffer (10 X)	3	1 X
Bst Polymerase large fragment (8000 U/mL)	0.067	240 U
Total volume	30 µL	

b) Incubate for 20 min at 37 °C.

8. 1.6X Sera-Mag Bead Clean Up TIMING 2h

Perform the same as the first Sera-Mag Bead clean up except initially adding 1.6 X (48 µL) Sera-Mag Beads and using 70% EtOH for a total of two washes. Elute in 17.51 µL MiliQ water, and transfer 16.51 µL at the final step.

9. Enrichment PCR TIMING 1.5h

a) Prepare the reaction mix as indicated below and perform the reaction in a 96-well PCR plate.

Reagent	Volume (µL) per sample	Final concentration
Nick filled-in DNA	16.51	
PCR buffer (10 X)	2	1X
$MgCl_2$ (25 mM)	0.8	1 mM
PreHyb-PE_F&R (24 µM)	0.2	0.25 µM
dNTP (25 mM)	0.16	0.2 mM
HotStarTaq Plus DNA Polymerase (5 U/µL)	0.33	0.08 U/µL
Total volume	20µL	

b) Run the program as below in a thermocylcler.

6 min 95 °C

(20 sec 95 °C, 45 sec 55 °C, 30 sec 72 °C) **X 13***

5 min 72 °C

∞ 4°C

*The number of cycles can be adjusted to 10-15 depending on the DNA concentration.

10. Sample quantification and Pooling TIMING 4h

Quantify average DNA concentration using Qubit per sample to achieve a minimum of 1 µg of total library DNA. In this study, 9.2 µL per sample were pooled.

11. Sera-Mag Bead Clean up (1.2X) TIMING 2h

Perform the same as the first Sera-Mag Bead clean up except initially adding 1.2 X Sera-Mag Beads of enrichment PCR products and using 70% EtOH for a total of two washes. Elute in 51 µL MiliQ water to transfer 50 µL at the final step.

12. Hybridisation TIMING 76.5h (Roche NimbleGen, Chapter 5. SeqCap EZ Library SR User's Guide)

a) Prepare the hybridisation mix as below.

Reagent	Volume (µL)/sample	Final concentration

Pooled and cleaned up DNA library	50	
BO1.P5.F (200 µM)	**0.83**	**2.4 µM**
BO2.P5.R (200 µM)	**0.83**	**2.4 µM**
BO3.P7.part1.F (200 µM)	**0.83**	**2.4 µM**
BO4.P7.part1.R (200 µM)	**0.83**	**2.4 µM**
BO5.P7.part2.F (200 µM)	**0.83**	**2.4 µM**
BO6.P7.part2.R (200 µM)	**0.83**	**2.4 µM**
Univ_Block_P5 (200 µM)	**2.5**	**7.2 µM**
Univ_Block_P7 (200 µM)	**2.5**	**7.2 µM**
Human COT-1 DNA (1 mg/mL)	**5**	**0.07 mg/mL**
Nimblegen Seq Capture EZ developer reagent	**5**	
Total volume	70 µL	

b) Add 20 µL of hybridisation mix to 49 µL of pooled and cleaned up DNA library in a PCR tube.

c) Make holes on the tube using a needle.

d) Speed vac until dry.

e) Change the lid so that the tube does not have a hole.

f) Add the following reagents to the pooled and cleaned up DNA and hybridisation mix.

Reagent	Volume (µL) per sample
Hybridisation buffer (2 X; vial 5)	7.5
Hybridisation component A (vial 6)	3
Ez developer library	4.5
Total volume	15 µL

g) Mix well to dissolve.

h) Dissolve the sample at 95 °C on a thermocycler up to 45 min. Mix by pipetting every 15 min.

i) After the sample is dissolved, incubate at 95 °C for 10 min.

j) Start gradual temperature drop down from 95 °C to 52 °C (-0.1 °C/sec).

k) Once the temperature reaches 52 °C quickly set the temperature to 47 °C and leave it for incubation for 72 hours.

13. Wash and Recovery of Captured Sample TIMING 2.5h (NimbleGen, Chapter 6. SeqCap EZ Library SR User's Guide)

STEP 1. Prepare the wash buffers

a) Pull out the streptavidin beads 20 min before use.

b) Prepare 1X working buffers of each concentrated wash buffer as directed below. Prepare extra 5% volume of the 1X working buffers as suggested below.

Concentrated buffer	Buffer (µL)	MiliQ (µL)	Total volume
10 X wash buffer I (vial 1)	31.5	283.5	315 µL
10 X wash buffer II (vial 2)	21	189	210 µL
10 X wash buffer III (vial 3)	21	189	210 µL
10 X stringent wash buffer (vial 4)	42	378	420 µL
2.5 X bead wash buffer (vial 7)	210	315	525 µL

c) Aliquot 400 µL of 1X stringent wash buffer and 100 µL of 1 X wash buffer I into each well of a PCR strip tube. **Heat** the two buffers and **magnetic bars** on a pcr machine at 47 °C.

STEP 2. Prepare the Capture Beads

d) Mix the streptavidin beads well by vortexing.

e) Aliquot 100 µL of the streptavidin beads into a pcr tube.

f) Place the tube on the ring magnet stand.

g) Remove the supernatant when it goes clear.

h) Add 2 X volume (200 µL) of bead wash buffer (vial 7) to beads on magnet.

i) Remove magnet and mix by pipetting.

j) Place the tube back in the magnet.

k) Discard supernatant when clear.

l) Repeat Step h)-k) for a total of two washes.

m) After removing the supernatant add 100 µL of bead wash buffer (vial 7).

n) Incubate at 47 °C on a PCR machine.

o) Place the heated bar magnet.

p) Once clear remove the supernatant. Beads are ready.

STEP 3. Bind DNA to the Capture Beads

a) Straight away transfer 15 µL hybridisation sample to the beads on PCR machine at 47 °C. Do not let the beads dry.

b) Mix thoroughly ten times by pipetting up and down.

c) Incubate for 45 min at 47 °C.

d) Mix by pipetting for 3 sec at 15 min interval.

STEP 4. Wash the Captured Beads Plus Bound DNA

a) After the 45 min incubation, add 100 µL wash buffer I (vial 1) heated to 47 °C to the bead plus captured DNA.

b) Mix by pipetting five times.

c) Place a heated bar magnet.

d) Remove supernatant once clear.

e) Remove the magnet.

f) Add 200 µL 1X stringent wash buffer (vial 4) heated to 47 °C.

g) Mix by pipetting five times.

h) Incubate at 47 °C for 5 min.

i) Place a heated bar magnet.

j) Remove supernatant once clear.

k) Repeat step f)-j) for a total of two washes.

l) Once the supernatant is removed, all the rest of the steps are done at room temperature.

m) Add 200 µL room temp wash buffer I (vial 1).

n) Mix by pipetting a few times then vortex for 2 min. Give PCR tube a quick hand spin.

o) Place the tube in a magnetic plate to bind the beads.

p) Remove and discard supernatant once clear.

q) Add 200 µL of room temperature wash buffer II (vial 2).

r) Mix by vortexing for 1 min and do a quick hand spin.

l) Place the tube on the ring magnetic plate.

m) Remove and discard supernatant once clear.

n) Add 200 µL of room temperature buffer III (vial 3)

o) Mix by vortexing for 30 sec.

p) Place the tube on the ring magnetic plate.

q) Remove and discard liquid once clear.

r) Remove the tube from the ring magnetic plate.

s) Add 60 µL 0.1 mM EDTA.

t) Store samples captured plus beads in a freezer.

14. Semi-qPCR TIMING 3h

a) Prepare the reaction mix.

Reagent	Volume (µL) per sample	Final concentration
Washed and recovered captured library	10	
MiliQ water	4.2	
Phusion HotStart HF buffer (5 X)	4	0.4 X
$MgCl_2$ (50 mM)	0.4	1 mM
Sol-PE-PCR_F (10µM)	0.4	0.2 µM
Sol-PE-PCR_R (10 µM)	0.4	0.2 µM

Phusion HotStart HF DNA polymerase (2 U/µL)	0.4	0.02 U/µL
Total volume	49.8 µL	

b) Run the following program (a total of 25 cycles). After 15, 20 and 25 cycles remove 4 µL of PCR product while the temperature is at 72 °C to check the concentration at each cycle.

1 min 98 °C

(15 sec 98 °C, 30 sec 60 °C, 30 sec 72 °C) X 15 cycles

2 min 72 °C

(15 sec 98 °C, 30 sec 60 °C, 30 sec 72 °C) X 5 cycles

2 min 72 °C

(15 sec 98 °C, 30 sec 60 °C, 30 sec 72 °C) X 5 cycles

7 min 72 °C

∞ 4 °C

c) Prepare 1.2% Agarose gel + EtBr. Load 4 µL of the PCR product with orange loading dye and 2.5 µL of 100 bp DNA ladder to run the gel.

* We estimated the number of indexing PCR at this step. The remaining semi-qPCR product from the 25 cycles can be used for Qubit quantification after Sera-Mag Bead clean up.

15. Indexing PCR TIMING 1.5h

a) Prepare the mix as below.

Reagent	Volume (µL) per sample	Final concentration
Captured DNA libraries	10	
MiliQ water	3.8	
Phusion HotStart HF buffer (5 X)	4	1 X
dNTPs (25 mM)	0.2	0.26 mM
Sol-PE-PCR_F (10 µM)	0.4	0.2 µM
Index-XX-PE Primer (5 µM)	0.8	0.2 µM
Phusion HotStart HF DNA polymerase (2 U/µL)	0.4	0.04 U/µL
Total volume	19.6 µL	

b) Run the following program:

1 min 98 °C

(30 sec 98 °C, 15 sec 60 °C, 15 sec 72 °C) X # cycles decided from semi-qPCR

7min 72 °C

∞ 4 °C

16. Sera-Mag Bead Clean up (1.2X) TIMING 2h

Perform the same as the first Sera-Mag Bead Clean up except initially adding 1.2 X (24 µL) Sera-Mag beads and using 70% EtOH for a total of two washes. Elute in 25µL of MiliQ water, and transfer 24 µL at the final step.

17. Gel Run and Qubit Quantification TIMING 1.5h

a) Prepare 1.2% agarose gel + EtBr. Load 3 µL of PCR product with orange loading dye and 2.5 µL of 200 bp DNA ladder to run the gel.

b) Quantify the concentration of PCR product (2 µL) using Qubit fluorometric quantification following the manufacturer's protocol.

18. Final Pooling and Submission of Indexed Library for Sequencing

Prepare 2 nM of 60 µL of indexed library in MiliQ water.

Chapter 4

Evaluation and comparison of 196 exon capture nuclear and chloroplast loci for Myrtaceae

Abstract

To date, resolving phylogenetic relationships within *Melaleuca* and *Eucalyptus* using chloroplast or nuclear markers has been challenging most likely due to lack of information (informative sites) because limited number of markers were available. In chapter 3, we identified in total of 209 chloroplast and nuclear loci to estimate species to generic phylogenies for Myrtaceae with focus on *Melaleuca* and *Eucalyptus*. After removing 13 loci based on average coverage, here we inferred maximum likelihood (ML) trees of the 196 loci to test which loci do not fit the prior taxonomic knowledge using 47 samples of 43 taxa from five genera in Myrtaceae. The criteria we assessed were as follows: 1) the samples were clustered by genus, 2) whether samples in *Melaleuca* formed three monophyletic clades. We then concatenated sequences as follows to estimate ML species trees: 1) all the chloroplast loci, 2) all the chloroplast and nuclear loci, 3) all the nuclear loci, 4) selected loci based on average coverage. We found that 144 loci are potentially useful for *Melaleuca* and 174 loci for the *Eucalyptus* group. Loci with outlying results were generally due to low coverage in a few *Corymbia* and *Eucalyptus* samples. In all the concatenated datasets, the topologies of *Melaleuca* clades were consistently (A,(B,C)) into three monophyletic clades. In Eucalypteae, the relationship was ((*Arillastrum*,*Corymbia*),*Eucalyptus*) in all datasets except in selected chloroplast loci based on coverage ((*Arillastrum*,(*Corymbia*,*Eucalyptus*))). In general, the potentially useful loci were able to better resolve the relationships in *Melaleuca* and *Eucalyptus* compared to the markers available to date although some internal branches remained unresolved.

Introduction

Resolving molecular phylogenetic relationships of species has been hampered in the past by a limited number of phylogenetically informative loci. Previous attempts to resolve the species relationships with fewer than five loci yielded a phylogeny in which many taxa remained unresolved with polytomies in *Melaleuca* (e.g. Edwards *et al.* 2010). That might indicate that the loci did not contain enough information to infer phylogenetic relationships, i.e. the loci did not have enough informative sites to resolve the relationship. Alternatively, it may also be due to incomplete lineage sorting, introgression or hybridisation among these taxa (e.g. Cook *et al.* 2008). In eucalypts, although a broad range of marker sets (e.g. Diversity Arrys Technology markers, ITS, *psbA-trnH*) have been used to date, the relationships in the genera *Corymbia* and *Angophora* remain problematic. With next generation sequencing it is now possible to sequence multiple loci in a short amount of time at lower costs (per sample per loci) compared to Sanger sequencing. To date, phylogenetic studies have been mostly based on single or up to a few chloroplast or nuclear loci. Alternatively, chloroplast whole genomes were also used in eucalypts (Bayly *et al.* 2013).

Different methods (e.g. gene tree estimation and/or reciprocal blasting) have been used to locate target loci for phylogenies (Misof *et al.* 2014; Teasdale *et al.* 2016). For phylogenetic studies we look for orthologous and low copy number loci. In chapter 3, we used gene tree estimation and assessed average coverage to identify orthologous loci with low copy number. For the gene tree estimation, we used priors based on existing taxonomic knowledge from molecular and morphological studies: 1) *Melaleuca* is sister to *Eucalyptus* (Thornhill *et al*. 2015; Wilson *et al*. 2001), 2) Within *Melaleuca* whether species from the same group were clustered together (Edwards *et al*. 2010). In this chapter, we used gene tree estimations of individual loci to further assess potentially useful loci.

With the development of Next-generation sequencing it is now possible to collect large sequence datasets (hundreds~thousands of loci; e.g. Blom *et al.* 2016; Valderrama *et al.* 2018). With the availability of large data it is crucial to decide which level of locus inclusion will be useful for robust phylogenetic estimation.

Different approaches have been taken to analyse multiple loci, and examples include concatenation and coalescence methods. In a concatenation approach, multiple loci are

concatenated into a single dataset and it is assumed that individual gene trees have the same branch length and phylogenetic relationship (Degnan & Rosenberg 2009; Edwards *et al.* 2015). In addition, often the loci are concatenated without checking the individual topologies (conflicts). This may result in inclusion of paralogous loci, resulting in erroneous estimations.

Gene trees of different loci do not necessarily concord with species trees (or species relationships) because different genes have different histories. Incomplete lineage sorting, horizontal gene transfer, hybridisation, and introgression can contribute to discordance of gene and species trees (Maddison 1997; Degnan & Rosenberg 2009). The coalescence method is increasingly used because it takes the evolutionary relationship of each locus into account (Liu *et al*. 2009; Blom *et al.* 2016). Concatenation approach and multispecies coalescence model work fine in trees with long branches and with low possibilities of incomplete lineage sorting, however, the concatenation approach can result in high support but disagreeing results from the coalescent approach (Edwards *et al.* 2016). This method can be computationally expensive with big dataset. The main aim of this study is to evaluate the marker set we identified in Chapter 3. Therefore a robust estimation of species phylogenies is out of the scope of this study.

With the recent taxonomic changes in the tribe Melaleuceae (*Melaleuca sensu lato*), eight genera were transferred to *Melaleuca* (Craven 2006; Craven *et al*. 2014) based on Edwards *et al.* (2010). They showed that eight genera (*Beaufortia*, *Calothamnus*, *Conothamnus*, *Eremaea*, *Lamarchea*, *Petraeomyrtus*, *Phymatocarpus*, and *Regelia*) which used to be in Melaleuceae (*Melaleuca sensu lato*) are nested within *Melaleuca.* In Edwards *et al.* (2010) showed strong monophyly of the tribe and the samples formed three main clades (A, B, C). based on chloroplast *ndhF* sequences. *Melaleuca sensu stricto* includes 12 informal species-groups and 10 of them were included in the study. Melaleuceae genera and/or species-group taxa included in the study (Edwards *et al.* 2010) is as the following: Clade A (*M. leucadendra* and *M. acacioides* species-groups, samples from the eight Melaleuceae genera); Clade B (Australian *Callistemon*, *M. fulgens, M. laxiflora, M. ordinifolia, M. lanceolata, M cuticularis* species groups); Clade C (*M. foliolosa, M. huegelii group, M. lanceolata, M. cuticularis* species-groups.).

The Eucalypt genera *Eucalyptus* and *Corymbia* are mostly endemic to Australia, with the distribution of some *Eucalyptus* species extending to New Guinea and Southeast

Asia while *Arillastrum* is native to New Caledonia (Bayly *et al*. 2013). Eucalypts are The genus *Eucalyptus* is the largest Myrtaceae genus in Australia. The taxonomic positions of *Angophora*, *Corymbia,* and *Eucalyptus* have been controversial in Eucalypteae. Since Brooker (2000) included *Angophora* and *Corymbia* as subgenera of *Eucalyptus,* the taxonomic change was not reflected in subsequent studies. For example, in phylogenetic analyses of ITS sequences by Steane *et al.* (2002) the three genera form a clade with good support and the genera *Angophora* and *Corymbia* were differentiated from *Eucalyptus sensu stricto.* Some studies supported the monophyly of *Angophora* and *Corymbia* while a recent study based on a genome phylogeny showed that *Corymbia* is paraphyletic with respect to *Angophora* (Bayly *et al.* 2013).

Using the 209 chloroplast and nuclear loci identified for Myrtaceae, with *Melaleuca* and *Eucalyptus* being the major target groups, we inferred gene trees of each locus. We assessed the topological conflicts to filter out loci that have outlying topologies and to test which of the 209 loci will be potentially useful to resolve the species to generic phylogenetic relationships in the family Myrtaceae. Finally, we concatenated all the loci based on which genome they are originating from (chloroplast/nuclear) or the average sequence coverage (Chapter 3) to estimate maximum likelihood species phylogenies.

Materials and Methods

Taxon sampling

Myrtaceae representatives including *Melaleuca* (6 taxa), Eucalypteae (36 taxa), and *Heteropyxis* (1 taxon) as an outgroup (Thornhill *et al.* 2015) were selected and collected from the field or from herbarium specimens. In total, we included 48 samples of 43 taxa (Table 1). The six taxa in *Melaleuca* represent the three main clades A (*Calothamnus gracilis*, *Melaleuca cornucopiae*), B (*M. squarrosa, M. capitata*) and C (*M. foliolosa, M. sylvana*; Edwards *et al*. 2010). Craven *et al*. (2014) transferred all the genera in the tribe Melaleuceae to *Melaleuca* and *Calothamnus gracilis* in Clade A is now called *Melaleuca gracilis* (Craven *et al*. 2014). See Figure 1 of Chapter 1 for the recent taxonomic changes in the tribe Melaleuceae. In Eucalypteae, the 36 taxa included represented *Arillastrum* (1 taxon), *Corymbia* (4 taxa), *Eucalyptus* (33 taxa). Samples in *Eucalyptus* are from three different subgenera, *Eudesmia* (3 taxa), *Symphyomyrtus* (14 taxa), *Eucalyptus* /formely *Monocalyptus* (14 taxa; Brooker 2000).

Genetic locus identification

We previously identified 209 loci for estimating species to genus level phylogenies in Myrtaceae, as described in Chapter 3. In brief, to identify the nuclear probes, we assembled randomly selected *Melaleuca* transcriptomes against whole genome shotgun sequences of *Melaleuca* species to discard loci that are present multiple times in the genome. We then blasted the selected transcriptomic contigs against the annotated *Eucalyptus grandis* genome on Phytozome (www.phytozome.net) to obtain the matching *E. grandis* coding DNA sequences and compared the copy number of the matching loci across the plant kingdom to remove potentially high copy number loci. Finally, we estimated gene trees for each locus and assessed whether the tree topologies match known molecular phylogenetic relationships (Edwards *et al.* 2010; Thornhill *et al.* 2015). To identify chloroplast marker sets, we aligned three chloroplast genomes from *Melaleuca* and *Eucalyptus* and assessed presence of insertions and deletions, and selected loci with nine or more single nucleotide polymorphisms per 500 bp because the chloroplast genome in general is conserved compared to the nuclear genome in plants (Gielly & Taberlet 1994). Nimblegen (Madison, USA) designed probes for selected loci using SeqCap EZ system. See Chapter 3 for further details.

DNA extraction, library preparation, exon capture, sequencing, read filtration

DNA was extracted from leaf tissues of the 48 samples of 43 taxa across Myrtaceae using Qiagen 96 Dneasy Kit (Hilden, Germany). We followed Rohland & Reich (2012) for the library preparation, and used SeqCap Ez Developer Library (Nimblegen, Madison, USA) to hybridise barcoded library to the probes following manufacturer's guide with small modifications. After recovery and wash the captured library was sequenced on Illumina Miseq platform (100 bp paired-end) at Bio-molecular Research Facilities at the Australian National University. We discarded low quality and duplicate reads and barcodes using Flexbar (Dodt *et al*. 2012) and checked quality of reads using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were mapped against the *Eucalyptus grandis* target loci using CLC genomics workbench (CLC Bio, Aarhus, Denmark) and when a nucleotide was present in more than 60% of the reads at a site we used it for a final sequence. See Chapter 3 for further details.

Locus selection and alignment for gene tree estimation

In total, 21,623 bp of 35 chloroplast and 236,869 bp of 161 nuclear loci were included for gene trees estimation. We excluded the following loci that had one or more samples with zero coverage (see supplementary Table 1 from Chapter 3): a chloroplast locus (c_166301166800) and 12 nuclear loci (n_13, n_30327, n_33459, n_40184, n_40488, n_40707, n_40909, n_41348, n_41455nof, n_41625, n_41818, n_41892). We also excluded one of the three *Arillastrum gummiferum* samples from subsequent analyses because the sample (RBGMelb) appeared to be misidentified.

We aligned sequences of each locus with MAFFT L-INS-i (Katoh & Stanley 2013) followed by manual editing using Geneious 7.1.4 (Kearse *et al.* 2012). Unalignable regions were excluded from analyses. Alignment ends were trimmed to remove ragged ends.

Phylogenetic analyses

For each locus we inferred the maximum likelihood (ML) phylogeny using PhyML (GTR+G substitution model; Guindon *et al*. 2010) implemented in Geneious 7.1.4 (Kearse *et al*. 2012). Branch supports were based on bootstrapping with 100 replicates (Felsenstein 1985). Each gene tree was rooted between the *Melaleuca* clade (Melaleuceae) and Eucalypteae clades because *Heteropyxis* (outgroup) was often not suitable for rooting due to short sequence lengths (< 50% length of aligned sequences, see Results section).

We concatenated 196 loci into four datasets: 1) 32 chloroplast loci of 47 samples (Cpl_all) 2) 164 nuclear loci of 47 samples (Ncl_all) 3) 164 nuclear and 32 chloroplast of 47 samples (Cpl_ncl_all), 4) 25 selected chloroplast loci in which a minimum of 43 samples or more were represented with the average coverage of 10 or more (Cpl_best) 5) 47 selected nuclear loci in which a minimum of 43 samples were represented with the average coverage of 10 or more (Ncl_best).

To assess topological conflicts or differences in the gene trees, we compared the topologies with respect to the main clades of *Arillastrum*, *Corymbia* and *Eucalyptus* in Eucalypteae. In *Melaleuca*, we assessed topologies to see whether *Melaleuca* species were grouped into three clades (A, B, C) as in Edwards *et al.* (2010), regardless of bootstrap support achieved. Furthermore, we assessed whether *Calothamnus* was clustered outside of all the rest of *Melaleuca,* as expected from the traditional taxonomy,

or if it was within clade A as expected from the phylogeny based on the nuclear 5S and ITS-1 ribosomal DNA marker and chloroplast *ndhF* gene respectively (Brown *et al.* 2001; Edwards *et al.* 2010). We further assessed whether samples in the same genus clustered together, and if not what could have contributed to the relationship. We can potentially filter out problem loci, if present, among the selected loci by comparing gene tree topologies. However, in this study, we do not have priors for the relationship among A, B and C in *Melaleuca* and among Eucalypteae genera, so we do not aim to make a claim on the above mentioned relationships. We also compared tree topologies among different sets of concatenated loci and among individual loci. Although we have assessed and compared gene trees of individual chloroplast loci (partitions/segments; and that we tentatively call it as loci here), all chloroplast loci are linked and should technically be considered as only one locus.

Results

We estimated Maximum likelihood (ML) gene trees from 196 loci represented by 43 species across Myrtaceae. The aligned length of loci varied from 268 bp to 6,588 bp, with the average length of 1,320 bp per locus (Table 1). All samples from *Melaleuca, Eucalyptus, Corymbia* and *Arillastrum* were clustered as monophyletic clades except for loci that have one or more sequences represented by less than 50% of the alignment length (e.g. *Corymbia calophylla, Eucalyptus goniantha, E. pachyphylla, E. stenostoma*). Short sequences of these species had low average coverage (See Supplementary Table S1 in Chapter 3). In 61.7% of the chloroplast and nuclear loci, sequences of *Heteropyxis* were less than 50% of the alignment length.

Gene Trees in General

In general, higher bootstrap support was achieved in nuclear than in chloroplast gene trees by visually assessing 196 gene trees based on bootstrap values and tree topologies.

Gene trees in Melaleuceae

The six *Melaleuca* species were grouped into three main clades as in Edwards *et al*. (2010) in 131 maximum likelihood gene trees (68.2%) out of 164 nuclear loci we estimated: clade A (*Calothamnus gracilis*, *M. cornucopiae*), clade B (*M. squarrosa, M. capitata*), clade C (*M. sylvana, M. foliolosa*) (e.g. Figs. 1, 2). However, when *Melaleuca* was grouped into three clades the relationship among them could be any one of the three possible topologies: ((A,B),C) (e.g. Fig. 2); ((A,C),B) (e.g. Fig. 3); (A,(B,C)) (e.g. Fig. 1). In

47 (24%) of gene trees, while B and C formed separate clades, species in clade A were not sister to each other, and the relationships among clade B and C and the other two taxa (*Calothamnus gracilis*, *M. cornucopiae*) varied.

In the chloroplast loci, species in gene trees of 10 loci had the relationship of (A,(B,C)), while seven loci showed ((A,C),B) and five had the topology of ((A,B),C). The remaining 10 loci did not resolve the three clades (A,B,C) (Table 2), while in 10 gene trees A, B, and C clades were not reciprocally monophyletic. In the nuclear loci, the most frequent topology was (A,(B,C)), in 48 gene trees, while the relationship of ((A,B),C) and (A,C),B) were found in 41 and 23 gene trees respectively. In 52 gene trees from nuclear loci, *Melaleuca* did not cluster into the three main clades. In 92% of the gene trees of chloroplast and nuclear loci, *Calothamnus gracilis* was nested within *Melaleuca.* In the remaining gene trees (8%), *C. gracilis* was sister to the rest of *Melaleuca* (Tables 1-2, Fig. 5).

Gene trees in Eucalypteae

In Eucalypteae, we assessed the relationship of *Arillastrum, Corymbia,* and *Eucalyptus*. Results of gene trees of nuclear loci showed that *Arillastrum* was either sister to *Corymbia* in 43% of the gene trees, or sister to *Eucalyptus* in (40 gene trees, Table 1). *Arillastrum* formed a sister group to both *Corymbia* and *Eucalyptus* in 40 gene trees. The other topologies were found in 12 gene trees of nuclear loci. The species in the same genus were either paraphyletic in the gene trees of nuclear loci or formed polytomies in the chloroplast. In the gene tree of nuclear loci n_42230 and n_BI097026, *Arillastrum* was nested within Eucalypteae. However, in both loci either a *Eucalyptus* or a *Corymbia* taxon was sister to the rest of Eucalypteae species and the sequence lengths were less than 50% of the alignment length. In addition, in cases where coverage of *Heteropyxis* was low and it was not possible to use the taxa to root the trees. In general, the relationship among major clades in *Eucalyptus* was better resolved in gene trees of nuclear loci compared to the chloroplast loci (e.g. Figs. 1-5). In the gene trees of the chloroplast loci, often larged polytomies were observed (e.g. Figs.1-2). The majority of taxa were clustered by subgenera but *E. coccifera, E. preissiana, E. virginea* and *E. goniantha* often did not cluster by subgenus (*Eudesmia*: Eud*, Eucalyptus*: Euc*, Symphyomyrtus:* Sym; e.g. Figs. 3-5).

Species tree based on concatenated data

In general, maximum likelihood tree estimations of all concatenated datasets better resolved the relationships within Eucalypteae and *Melaleuca* with an increase in percentage bootstrap support compared to individual gene trees. *Heteropyxis* sequence length was longer than 50% of the total alignment in only 121 loci, while 75 of the loci were represented with *Heteropyxis* sequences shorter than 50% of the alignment length. Rooting with *Heteropyxis* was possible in all the concatenated datasets.

In the concatenated tree of Cpl_all, branch lengths were twice as long among the six *Melaleuca* species than among all the *Eucalyptus* species when compared by eye (Fig. 6). On the other hand, the ranges of genetic differences (branch lengths) of the concatenated tree of all nuclear loci (Ncl_all) were similar among the six *Melaleuca* species and the *Eucalyptus* taxa (Fig. 7). When both nuclear and chloroplast loci were concatenated (Cpl_ncl_all) the branch lengths were similar based on the trees (Fig. 8).

Relationships among the three main clades in *Melaleuca* were better resolved in all concatenated datasets compared to individual gene trees. The bootstrap values were over 70 on all nodes except on four bootstrap internal branches of the Cpl_all dataset ML phylogeny. In all of the five concatenated dataset trees, clades B and C are sister to clade A while the relationship among the three clades differed in individual gene trees. In all trees of the concatenated datasets *Calothamnus gracilis* was nested within *Melaleuca*.

In trees of all concatenated datasets, the generic relationship within Eucalypteae was consistently ((Ari,Cor),Euc), except for the phylogeny of the Cpl_best concatenated dataset. In the Cpl_best dataset, the relationships within Eucalypteae was (Ari,(Cor,Euc)) (Fig. 9). Furthermore, compared to single loci, bootstrap values for the three genera were often increased to over 95% or more in all of the concatenated datasets, but some of the internal branches of *Eucalyptus* remained unresolved with less than 70% bootstrap support values.

In the species tree of the Cpl_all dataset, *Eucalyptus* spcies were clustered largely by subgenus (subg. *Eucalyptus*, *Eudesmia, Symphyomyrtus*; Fig. 6) into three clades. The first two clades mostly included taxa from subgenera *Symphymyrtus* and *Eucalyptus* respectively, and they were sister to each other. However, two species from subgenus *Eucalyptus* species (*E. preissiana, E.coccifera*) were nested within subgenus *Symphymyrtus*

while two species (*E.virginea, E.goniantha*) from the subgenus *Symphymyrtus* were nested in the subgenus *Eucalyptus.* Relationships within the third clade was (*Eudesmia, Symphyomyrtus*) and this clade was sister to the former two clades. In Ncl_all dataset (Fig. 7), subgeneric relationship within the *Eucalyptus* was (Eud,(Sym,(Euc,Eud,Sym),(Euc,(*E.goniantha*,Euc)))). In Cpl_ncl_all dataset, a clade of *Eudesmia* was sister to a large polytomy clade that included three main clades. Two of the clades were mostly clustered by subgenus (subg. *Eucalyptus, Symphyomyrtus*) but the last clade included samples from all the subgenera (Fig. 8).

In the species tree of Cpl_best dataset, taxa from subgenera *Symphyomyrtus* and *Eucalyptus* mostly formed a clade separately, forming a sister clade. These two clades were then sister to all *Eudesmia* taxa and two *Symphyomyrtus* species (Fig. 9). In the Ncl_best dataset taxa in *Eudesmia* were clustered by genus except for few taxa. Subgenera *Symphyomyrtus* and *Eucalyptus* formed a clade and this clade was sister to the clade of *Eudesmia* (Fig.10).

Discussion

In Eucalypteae, all gene trees of the remaining 196 loci topologies mostly adhere to one of the following although in some loci species did not cluster by genus: (Ari,(Cor,Euc)), ((Ari,Cor),Euc), ((Aril,Euc),Cor). When the species in Eucalypteae were not clustered by genus and therefore had outlying topologies, it was because a few taxa were nested within or sister to different genera. We found outlying topologies in 12 nuclear loci with paraphyletic relationships, which was mainly due to short sequences in *Corymbia calophylla*, *Eucalyptus goniantha* and a few other species (Table 1). In these cases, the sequence length of those species was less than 50% of the total alignment length and their average coverage was often below 10 (see supplementary table in Chapter 3). We would expect paralogs to have outlying topologies (such as species not being clustered by genus). Alternatively, considering the short sequence length, the outlying relationships in those loci are more likely due to the lack of information rather than due to actual difference/similarities among the sequences.

Different genes/loci have different evolutionary histories as demonstrated in the gene and evolutionary histories of gene trees might not match the species evolutionary relationships (Maddison 1997). Several different reasons can contribute to the topological discordances among gene trees and a species tree. Gene duplication events,

extinction of genes, horizontal gene transfer, and (incomplete) lineage sorting can all result in different topologies in gene trees (Maddison 1997). Previous research has shown evidence or at least the possibility of hybridisation (Griffin *et al*. 1988), introgression and incomplete lineage sorting in *Melaleuca* and *Eucalyptus* (e.g. Cook *et al*. 2008) so we would expect gene trees to have different topologies and that the species tree can have different topologies from individual gene trees.

In general, the relationships among the three main clades A,B, and C in *Melaleuca* and topologies among *Arillastrum, Eucalyptus* and *Corymbia* were better resolved in trees of concatenated datasets compared to individual gene trees of both chloroplast and nuclear loci. When different loci are concatenated into a dataset and analysed the assumption is that all the loci have the same evolutionary history, and the evolutionary histories of individual loci/genes are not taken into account (Degnan and Rosenberg 2009; Liu *et al.* 2009). Of the six concatenated datasets, only chloroplast_best of Eucalypteae resulted in a different topology: (Ari,(Cor,Euc)). However, individual gene trees showed an array of different topologies (Tables 1-3, Figs. 1-10) that is likely to be hidden in the concatenated dataset. Further investigation is required to explore characteristics of each locus/gene that might explain their different evolutionary histories.

Our results support the inclusion of *Calothamnus* into *Melaleuca*. The recent taxonomic changes of transferring all the genera in the tribe Melaleuceae including *Calothamnus,* into *Melaleuca* has been controversial (Udovicic & Spencer 2012; Edwards *et al.* 2010). Craven *et al.* (2014) transferred species in *Calothamnus* to *Melaleuca* based on Edwards *et al.* (2010). In 92% of gene trees, *Calothamnus* was nested within *Melaleuca* and this result agreed with Edwards *et al.* (2010). Based on the *ndhF* gene phylogeny and morphology of genera within Meleleuceae, Edwards *et al.* (2010) have shown that the eight genera (*Beaufortia, Calothamnus, Conothamnus, Eremaea, Lamarechea, Petraeomyrtus, Regilia* and *Phymatocarpus*) are nested within *Melaleuca,* and grouped into clades A, B and C. In this study, 68% of individual nuclear loci and all concatenated datasets support the *Melaleuca* clades A, B and C. Furthermore, in Edwards *et al.* (2010), *Calothamnus* and most of the genera in Melaleuceae clustered into clade A. Our results provide an evidence to support the claim. Gene trees of 68% of loci in this study support clustering of *Calothamnus* with clade A, and thus also support the transfer of *Calothamnus* to

Melaleuca. Further analyses that include a larger number of samples (taxa) from each genus is required to further resolve this issue.

In Eucalypteae, the relationship of *Arillastrum* with *Corymbia* and *Eucalyptus* was uncertain. Brooker *et al*. (2000) suggested that *Corymbia* should be included in *Eucalyptus sensu lato* as a subgenus. However, in subsequent studies this generic circumscription was not applied (e.g. Balyly *et al.* 2013). Steane *et al.* (2002), using the nuclear ribosomal internal transcribed spacer (ITS) found that *Corymbia* and *Arillastrum* are not a part of *Eucalyptus*. In this study, 43% of gene trees and all the concatenated datasets except Cpl_best supports the relationship of ((Ari,Cor),Euc). Recent phylogenetic analysis based on Eucalypt chloroplast genomes agreed with this result (Bayly *et al.* 2013). Because all the chloroplast loci are linked and they need to be regarded as a single locus, a different subset chloroplast dataset can potentially result in different relationship as can be seen in the Cpl_best dataset.

The dataset Cpl_best had the following relationship: (Ari,(Cor,Euc)). Whether *Corymbia* is paraphyletic with respect to *Angophora,* which was not included in this study, is still debated (Parra-O *et al.* 2010; Bayly *et al.* 2013). Using more samples of different species from each genus, including *Angophora,* which is very closely related to, will be able to provide more convincing evidences to make a claim on the relationship of the genera.

Examinations of concatenated species trees based on concatenated datasets showed non-monophyly of the three subgenera *Eudesmia, Symphyomyrtus* and *Eudesmia.* An exception is found in species trees of dataset Cpl_best and Ncl_best datasets wherein subgenus *Eudesmia* formed a monophyletic clade. In comparison, concatenated datasets of nuclear loci when all the loci were concatenated (Ncl_all), resulted in polytomies with one of the major clades containing taxa of all three subgenera. In addition, bootstrap values for one of the main nodes was <70%. In contrast, in the tree of Ncl_best dataset, major nodes of genus *Eucalyptus* were better resolved with good support (>93%).

In gene trees, two taxa (*E. preissiana, E. coccifera*) from subg. *Eucalyptus* were often nested within subgenus *Symphyomyrtus* and two taxa from subgenus *Symphyomyrtus* (*E. goniantha, E. virginea*) were nested within subgenus *Eucalyptus*. The outlying topology of *E. goniantha* and *E. preissiana* is most likely due to low average coverage of loci in

this study. However, the average coverage of *E. virginea* and *E. coccifera* was not problematic and we could find clear reasons for the outlying topology of the taxa. Whether this is due to a misidentification of samples, or if this reflects the species evolutionary history, need further exploration. The relationship of the three subgenera, ((Sym,Euc),Eud), agreed with the result of Bayly *et al.* (2013). Another thing to note from the nuclear concatenated dataset is that different combinations of marker sets can potentially increase tree resolution and with higher bootstrap support (e.g. Ncl_best. See below for the discussion).

In this study, some of the branches within *Eucalyptus* remained unresolved with low bootstrap values (< 70%) in the trees of all concatenated loci. Some of those low support values could have been due to low genetic differentiation among *Eucalyptus* species in this study (33 species) compared to *Melaleuca*, for example. Alternatively, it could be due to conflicts among different genes. In the case of *Melaleuca* clades (six species), likewise, if more closely related species are added to each clade A, B and C, the bootstrap support values of some of the nodes could potentially decrease.

Out of 196 loci, we have found 144 loci potentially useful to better resolve phylogenies of *Melaleuca* and 174 loci for Eucalypteae based on the gene trees and concatenated trees of the 47 samples across Myrtaceae. This was based on the average coverage and topologies within each group, whether the topologies met the taxonomic/phylogenetic priors. To conclude, the loci that we selected after filtering out the loci that does not meet the taxonomic/phylogenetic priors will serve as useful markers to come up with more robust hypothesis of the *Melaleuca, Eucalyptus* or other relationships within Myrtaceae.

We estimated species trees using concatenated datasets, yet to draw a conclusion about the species or generic relationship further phylogenetic estimations using other methods are needed to work out the best hypothesis. Coalescence, for example, explicitly deals with incongruent gene trees. Both in *Melaleuca* and *Eucalyptus,* use of overhanging intronic regions at the boundary of the selected loci (exons) in the nuclear loci could improve the resolution of species relationships and internal branches in *Eucalyptus*. Ultimately, by adding more samples, these loci provide an excellent opportunity to estimate more robust species to generic phylogenies of *Melaleuca* and Eucalypteae and also other genera in Myrtaceae in future. Furthermore, the

phylogenies will serve as useful backbones to explore the hypotheses regarding the Australian flora.

References

Bayly, M.J., Rigault, P., Spokevicius, A., Ladiges, P.Y., Ades, P.K., Anderson, C., Bossinger, G., Merchant, A., Udovicic, F., Woodrow, I.E. & Tibbits, J. (2013) Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* 69: 704-716.

Blom, M.P., Bragg, J.G., Potter, S. & Moritz, C. (2016) Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Systematic biology* 66: 352-366.

Brooker M.I.H. (2000) A new classification of the genus *Eucalyptus* L'Hér. (Myrtaceae). *Australian Systematic Botany* 13: 79-148.

Brophy, J.J., Craven, L.A. & Doran, J.C. (2013) Melaleucas: their botany, essential oils and uses. Australian Centre for International Agricultural Research (ACIAR).

Cook, L.G., Morris, D.C., Edwards, R.D. & Crisp, M.D. (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* 47: 506-522.

Craven, L.A. (2006) New combinations in *Melaleuca* for Australian species of *Callistemon* (Myrtaceae). *Novon: A Journal for Botanical Nomenclature* 16: 468-475.

Craven, L.A., Edwards, R.D. & Cowley, K.J. (2014) New combinations and names in *Melaleuca* (Myrtaceae). *Taxon* 63: 663-670.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332-340.

Dodt, M., Roehr, J.T., Ahmed, R. & Dieterich, C. (2012) FLEXBAR—flexible barcode and adapter processing for *Melaleuca* next-generation sequencing platforms. *Biology* 1: 895–905.

Edwards, R.D., Craven, L.A., Crisp, M.D. & Cook, L.G. (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L. (Myrtaceae) is not monophyletic. *Taxon* 59: 744-754.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R. & Leaché, A.D. (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular phylogenetics and evolution* 94: 447-462.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1: 783-791.

Gielly, L. & Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular Biology and Evolution* 11: 769–777.

Griffin, A.R., Burgess, I.P. & Wolf, L. (1988) Patterns of natural and manipulated hybridisation in the Genus *Eucalyptus* L'hérit.-1 A Review. *Australian Journal of Botany*: 36, 41-66.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kearse, M., Moir, R. & Wilson, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organisation and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Lanfear, R., Ho, S.Y., Davies, T.J., Moles, A.T., Aarssen, L., Swenson, N.G., Aarssen, L., Swenson, N.G., Warman, L., Zanne, A.E., & Allen, A.P. (2013) Taller plants have lower rates of molecular evolution. *Nature Communications* 4: 1879.

Liu, L., Yu, L., Pearl, D.K., & Edwards, S.V. (2009) Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58: 468-477.

Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology* 46: 523-536.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G. & Niehuis, O. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346: 763-767.

Slee, A.V., Brooker, H., Duffy, M. & West, J.G. (2006) EUCLID eucalypts of Australia. 3 rd edition DVD edn. Centre for Plant Biodiversity Research-CSIRO publishing, Canberra.

Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E. & Potts, B.M. (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* 15: 49-62.

Teasdale, L.C., Koehler, F., Murray, K.D., O'Hara, T. & Moussalli A. (2016) Identification and qualification of 500 nuclear, single- copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources* 16: 1107-1123.

Thornhill, A.H., Ho, S.Y., Külheim, C. & Crisp, M.D. (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Molecular Phylogenetics and Evolution* 93: 29-43.

Udovicic, F. & Spencer, R.D. (2012) New combinations in *Callistemon* (Myrtaceae). *Muelleria* 30: 23-25.

Valderrama, E., Richardson, J.E., Kidner, C.A., Madriñán, S. & Stone, G.N. (2018) Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L.f., Zingiberaceae). *Molecular phylogenetics and evolution* 119: 13-24.

Wilson, P.G., O'Brien, M.M., Heslewood, M.M. & Quinn, C.J. (2005) Relationships within Myrtaceae *sensu lato* based on *matK* phylogeny. *Plant Systematics and Evolution* 251: 3-19.

Tables.

Table 1. A summary of maximum likelihood gene trees of 196 chloroplast and nuclear loci. 47 samples across Myrtaceae were used. A, B, and C in *Melaleuca* represent the three main clades identified in Edwards *et al*. (2010). *Calothamnus gracilis*, which was transferred to *Melaleuca*, is in clade A. The abbreviations for the generic names in eucalypts are as the following: Ari (*Arillastrum*), Cor (*Corymbia*), Euc (*Eucalyptus*). Y and n is abbreviation for yes and no respectively. C_ and n_ are abbreviations for chloroplast and nuclear. * refers to those loci in which *Heteropyxis* was less than 50% of the alignment length.

					Which genera is *Arillastrum* sister to				
	Loci name	Relationship in *Melaleuca*	Relationship within Eucalypteae	(Cal,Mel)	Cor	Euc	Euc+Cor	nested in Eucalypteae	Alignment length (bp)
1	c_20012500	((A,C),B)	(Ari,(Ari,(Cor,Euc)))	n			y		497
2	c_25013000	((A,C),B)	(((Cor,Ari),Euc),Euc)	n	y			y	486
3	c_30013500	((A,B),C)	((Ari,Euc),Cor)	n		y			515
4	c_9220192700	(A,(B,C))	(((Ari,Cor),Euc),Cor.cal)	n	y				504
5	c_35014000	((((B,M.syl),C.gra),M.cor),M.fol)	((Ari,Euc),Cor)	n		y			501
6	c_1740117900	(((B,M.cor),C),Cal.gra)	((Ari,(E.gon,E,sta)),(Euc,(Cor,Euc)),Euc)	y		y			505
7	c_1980120300	(A,(B,C))	((Ari,Cor),(Cor,Euc))	n	y				500
8	c_2080121300	(A,(B,C))	((((Cor,E.pre),Ari),Ari),the rest of Euc)	n			y		500
9	c_4510145600	(((A,M.cap),M.squ),C)	((Ari,Euc),(Ari(E.moo,(Cor.hae,(Cor.bun,Euc))))	n			y		501
10	c_4902049519	((A,B),C)	(Ari,(Cor,Euc))	n			y		519
11	c_5020150700	(A,(B,C))	((((Ari,Euc),Euc),Euc),Cor)	n		y		y	499
12	c_5230152800	((C,M.cap),M.squ),A)	((Ari,E.gon,Cor),Euc)	n			y		504
13	c_5550156000	(A,(B,C))	((Ari,Euc),(E.ste,Cor))	n			y		511
14	c_6490165400	(M.cor,Cal.gra,(B,C))	((Ari,Ari,Cor),Euc)	n	y				500
15	c_6760168100	(((B,M.syl),M.fol), A)	(((Ari,(Ari,Euc)),Cor),E.gon)	n		y			503
16	c_6810168600	(Cal.gra,(M.cor,(B,C)))	((Ari,(Cor,Euc)),E.gon)	y			y		508*
17	c_6919169690	(M.cor,((M.syl,M.fol),(M.fol,B)))	((Ari,Cor),Euc)	n	y				507
18	c_7290173400	(M.cor,((Cal.gra,C),B))	((Ari,Cor,Euc),(Cor,Euc))	n		y			505
19	c_8330183800	(A,(B,C))	((Ari,Cor),Cor.cal)	n	y				509

20	c_9002690525	((A,B),C)	(((Ari,Cor),Cor),Euc)	n	y			502
21	c_150761151760	((C,A),B)	((Ari,Cor),Euc)	n	y			1023
22	c_152701153200	(A,(B,C))	((Cor,Euc),(Ari,Euc))	n		y		506
23	c_154601155100	((A,B),C)	((Ari,Euc),Cor)	n		y		508
24	c_157101157600	(A,(B,C))	((Ari,Cor),Euc)	n	y			500
25	c_158701159200	((A,C),B)	(Ari,(Cor,(Cor,Cor,Euc)	n			y	505
26	c_164501165000	((A,C),B)	(Ari,(Cor,Euc))	n			y	503
27	c_165301165800	(A,(B,C))	((Ari,Euc),Cor)	n		y		506
28	c_167001167500	((A,C),B)	((Ari,Cor),Euc)	n	y			506
29	c_*ndhF*	(B,(Cal.gra,(M.cor,C)))	(Ari,(Cor,Euc))	n			y	1992
30	c_*psbA-trnH*	(A,(B,C))	((Ari,Cor),Euc)	n	y			596
31	c_*trnL-trnF*	((A,C),B)	(Cor,(Euc,(Euc,(Euc,Ari))))	n		y		796
32	c_*matK-trnK*	((A,B),C)	((Ari,Euc),Cor)	n		y		2311
33	n_39171	(Cal.gra,(M.cor,(B,C)))	((Ari,Euc),Cor)	n		y		2110*
34	n_44831	(M.cor,(C.gra,(B,C)))	((Ari,Cor),Euc)	n	y			1999*
35	n_30177	(A,(B,C))	((Ari,Euc),Cor)	n		y		994
36	n_30296	(C.gra(B,(M.cor,C)))	(Ari,(Cor,Euc))	y			y	1673*
37	n_30355	(A,(B,C))	(Ari,(Cor,Euc))	n			y	653
38	n_30410	(M.cor,(C.gra,(B,C)))	(((Ari,Euc),Cor),E.gon)	n		y		618
39	n_30431	(M.syl,(M.fol,(M.cor,B)))	((Ari,Cor),Euc)	n	y			1485*
40	n_30519	((C.gra,B),(M.cor,C))	(Ari,(Cor,Euc))	n			y	1921*
41	n_30906	(M.cor((C.gra,C),B))	((Ari,Cor),Euc)	n	y			824*
42	n_31003	(A,(B,C))	((Ari,Cor),Euc)	n	y			1719
43	n_31220	(A,(B,C))	((Ari,Euc),Cor)	n		y		1656
44	n_31275	((C.gra,C),(M.cor,B))	((Ari,Euc),Cor)	n		y		1064*
45	n_31348	(C,(B,A))	((Ari,Euc),Cor)	n		y		753*
46	n_31545	(A,(B,C))	((Ari,Cor),Euc)	n	y			688
47	n_31598	(C,(A,B))	((Ari,Cor),Euc)	n	y			4126
48	n_31662	(A,(B,C))	(Ari,(Cor,Euc))	n			y	1031
49	n_32080	(A,(B,C))	((Ari,Euc),Cor)	n		y		714

50	n_32112	(A,(B,C))	((Ari,Cor),(Euc,Cor.cal))	n	y			678*
51	n_32478	(Cal.gra,(M.cor,(B,C)))	(Ari,(Cor,Euc))	y		y		1014
52	n_32479	((A,C),B)	((Ari,Euc),Cor)	n		y		1471*
53	n_32563	(((B,Cal.gra),M.cor),C)	(Ari,(Cor,Euc))	n			y	1060
54	n_32592	((A,B),C)	(Ari,(Cor,Euc))	y			y	982*
55	n_32660	(A,(B,C))	((Ari,Cor),Euc)	n	y			1359*
56	n_32747	(A,(B,C))	(Ari,(Cor,Euc))	n			y	1284
57	n_32861	(M.cor,((B,Cal.gra),C))	((Ari,Cor),(Cor.cal,Euc))	n	y			1439*
58	n_32864	((A,B),C))	(Ari,(Cor,Euc))	n			y	578*
59	n_32883	((A,C),B)	((Ari,Euc),Cor)	n		y		1553*
60	n_33080	(A,(B,C))	((Ari,Cor),Euc)	n	y			1553
61	n_33128	((A,C),B)	((Ari,Euc),Cor)	n		y		1684
62	n_33353	(A,(B,C))	((Ari,Cor),Euc)	n	y			852*
63	n_33384	((A,C),B)	(Ari,(Cor,Euc))	n			y	1205
64	n_33409	(Cal.gra,(C,(M.cor,B)))	(Ari,(Cor,Euc))	y			y	418*
65	n_33482	(A,(B,C))	(Ari,(Cor,Euc))	n			y	2192*
66	n_33521	((A,B),C)	((Ari,Euc),Cor)	n		y		1671
67	n_33548	(((Cal.gra,C),M.cor),B)	((Ari,Cor),Euc)	n	y			1586
68	n_33877	((M.cor,(Cal.gra,B)),C)	(Ari,(Cor,Euc))	n			y	1971*
69	n_33912	((A,C),B)	((Ari,Cor),Euc)	n	y			1489
70	n_34024	(((A,B),M.fol),M.syl)	((Ari,Euc),Cor)	n		y		1744*
71	n_34029	((Cal.gra(M.cor,B),C)	((Ari,Cor),Euc)	n	y			1644*
72	n_34206	(A,(B,C))	((Ari,Cor),Euc)	n	y			613*
73	n_34262	((A,B),C)	((Ari,Cor),Euc)	n	y			601
74	n_34341	((A,B),C)	((Ari,Cor),Euc)	n	y			1109*
75	n_34473	(A,(B,C))	(Ari,(Cor,Euc))	n			y	1484
76	n_34730	((A,C),B)	((Ari,Cor),Euc)	n	y			916
77	n_34792	(A,(B,C))	((Ari,Cor),Euc)	n	y			796*
78	n_34840	(M.cor,(M.gra,(B,C)))	((Ari,Cor),Euc)	n	y			1202*
79	n_34877	(A,(B,C))	((Ari,Cor),Euc)	n	y			947

80	n_40086	((A,C),B)	((Ari,Cor),Euc)	n	y			798*
81	n_40267	((A,C),B)	((Ari,Cor), Euc)	n	y			1093
82	n_40333	(A,(B,C))	((Ari,Euc),Cor)	n		y		565*
83	n_40593	((Cal.gra,B),(M.cor,C))	((Ari,Cor),Euc)	n	y			671*
84	n_40617	(Cal.gra,(M.cor,(B,C)))	(Ari,(Cor,Euc))	y			y	1555
85	n_40711	((A,B),C)	(Ari,(Cor,Euc))	n			y	1493*
86	n_40715	(A,(B,C))	((Ari,Euc),Cor)	n		y		2278
87	n_40746	((A,B),C))	((Ari,Cor),Euc)	n	y			1199
88	n_40755	((A,C),B)	((Ari,Cor),Euc)	n	y			1039
89	n_40796	((A,B),C)	((Ari,Cor),Euc)	n	y			813*
90	n_40821	((A,B),C)	((Ari,Cor),Euc)	n	y			1861*
91	n_40832	((A,B),C)	((Ari,Euc),Cor)	n		y		822
92	n_40849	(((A,M.syl),M.fol),B)	((Ari,Euc),Cor)	n			y	451*
93	n_40919	((A,B),C)	((Ari,Cor),Euc)	n	y			1785
94	n_41002	(A,(B,C))	((Ari,Cor),Euc)	n	y			1759*
95	n_41012	((B,M.cor),Cal.gra),C)	((Ari,Euc),Cor)	n		y		1852
96	n_41115	((B,Cal.gra),(C,M.cor))	((Ari,Cor),Euc)	n	y			2250*
97	n_41143	(A,(B,C))	((Ari,Cor),Euc)	n	y			1667*
98	n_41174	(A,(B,C))	(Ari,(Cor,Euc))	n			y	1617
99	n_41186	((A,B),C)	((Ari,Cor),Euc)	n	y			1460*
100	n_41200	(M.cor,(C.gra,(B,C)))	((Ari,Cor),Euc)	n	y			988
101	n_41215	(C.gra,((M.cor,C),B))	((Ari,Euc),Cor)	y		y		1436
102	n_41364	((A,B),C)	((Ari,Cor),Euc)	n	y			1499*
103	n_41372	((M.cor,(C.gra,C)),B)	(Ari,(Cor,Euc))	n			y	1385
104	n_41395	((A,B),C)	((Ari,Cor),Euc)	n	y			1044*
105	n_41455D3	((A,C),B)	((Ari,Cor),Euc)	n	y			1942*
106	n_41627	(A, (B,C))	((Ari,Cor),Euc)	n	y			1498
107	n_41712	(((Cal.gra,C),M.cor),B)	((Ari,Cor),Euc)	n	y			1581
108	n_41838	(((B,M.cor),Cal.gra),C)	(Ari,(Cor,Euc))	n			y	398*
109	n_41845	((A,B),C)	((Ari,E.ste,Cor),Eucs)	n	y			268

110	n_41898	((A,C),B)	((Ari,Euc),Cor)	n		y		2781
111	n_41911	(A,(B,C))	((Ari,Euc),Cor)	n		y		1017*
112	n_41930	((A,B),C)	(Ari,(Cor,Euc))	n	y			1990*
113	n_41999	((A,B),C)	((Ari,Cor),Euc)	n	y			1261*
114	n_42036	((A,B),C)	((Ari,Euc),Cor)	n		y		3707
115	n_42060	((A,B),C)	((Ari,Cor),Euc)	n	y			1157*
116	n_42126	(A,(B,C))	((Ari,Euc),Cor)	n		y		409*
117	n_42166	((A,B),C)	(Ari,(Cor,Euc))	n			y	1581
118	n_42230	((A,C),B)	((Ari,Cor),Euc)	n	y			1121
119	n_42393	(((Cal.gra,(M.cor,C)),B)	(Ari,(Cor,Euc))	n			y	673*
120	n_42490	((A,B),C)	((Ari,Cor),Euc)	n	y			828
121	n_42561	((A,C),B)	((Ari,Euc),Cor)	n		y		1187
122	n_42584	((A,C),B)	((Ari,Cor),Euc)	n	y			1227*
123	n_42721	(M.cor,(Cal.gra(B,C)))	(Ari,(Cor,Euc))	n			y	810
124	n_42754	((A,B),C))	(Ari,(Cor,Euc))	n			y	1549
125	n_dxs2	((A,B),C)	((Ari,Cor),Euc)	n	y			2210*
126	n_idi2	(M.cor,((Cal.gra,B),C))	((Ari,Cor),Euc)	n	y			1884
127	n_2	((A,B),C)	((Ari,Cor),Euc)	n	y			1969
128	n_208	((A,B),C)	((Ari,Cor),Euc)	n	y			2011*
129	n_1873	((Cal.gra,(M.cor,B)),C)	(Ari,(Cor,Euc))	n			y	1298*
130	n_1931	((Cal.gra,(M.cor,C)),B)	((Ari,Cor),Euc)	y	y			1334
131	n_3214	((A,B),C)	(Ari,(Cor,Euc))	n			y	2628
132	n_3712	(A,(B,C))	((Ari,Euc),Cor)	n		y		1117*
133	n_4206	(A,(B,C))	((Ari,Euc),Cor)	n		y		458
134	n_4706	(M.cor,(B,(Cal.gra,C)))	((Ari,Cor),Euc)	n	y			516
135	n_6015	((Cal.gra,C),(M.cor,B))	((Ari,Cor),Euc)	n	y			2331
136	n_6199	((A,C),B)	((Ari,Euc),Cor)	n		y		1691
137	n_22854	(A,(B,C))	(Ari,(Cor,Euc))	n			y	1666
138	n_23033	((A,C),B)	(Ari,(Cor,Euc))	n			y	1317*
139	n_25721	((M.cor,(Cal.gra,C)),B)	(Ari,(Cor,Euc))	n			y	584*

140	n_30471	(A,(B,C))	((Ari,Euc),Cor)	n		y			698*
141	n_30562	(A,(B,C))	((Ari,Cor),Euc)	n	y				1299*
142	n_30839	((A,B),C)	((Ari,Cor),Euc)	n	y				5091*
143	n_31043	((A,C),B)	((Ari,Euc),Cor)	n		y			1854
144	n_31075	((A,B),C)	((Ari,Cor),Euc)	n	y				1152
145	n_31209	(M.cor,(Cal.gra,(B,C)))	(Ari,(Cor,Euc))	n			y		1809
146	n_31501	((A,C),B)	(Ari,(Cor,Euc))	n			y		1982
147	n_31750	(M.cor,(Cal.gra,(B,C)))	((Ari,Cor),Euc)	n	y				1439
148	n_31806	(M.cor,(Cal.gra,(B,C)))	((Ari,Cor),Euc)	n	y				1081
149	n_32269	(A,(B,C))	((Ari,Cor),Euc)	n	y				2692
150	n_32608	((A,C),B)	(Ari,(Cor,Euc))	n			y		629*
151	n_32708	(A,(B,C))	(Ari,(Cor,Euc))	n			y		2143*
152	n_33152	((A,B),C)	(Ari,(Cor,Euc))	n			y		3560
153	n_33440	(A,(B,C))	((Ari,(Cor,Euc)),E.gon)	n			y		1579*
154	n_33442	((A,C),B)	((Ari,Cor),Euc)	n	y				2547
155	n_33580	(A,(B,C))	((Ari,Euc),Cor)	n		y			1684*
156	n_33688	((A,B),C)	((Ari,Euc),Cor)	n		y			1643
157	n_33888	(A,(B,C))	(Ari,(Cor,Euc))	n			y		1210
158	n_34100	((A,B),C)	((Ari,Cor),Euc)	n	y				1650
159	n_34416	(A,(B,C))	((Ari,Cor),Euc)	n	y				544
160	n_34636	(M.cor,((B,Cal.gra),C))	((Ari,Cor),Euc)	n	y				1226*
161	n_34751	(A,(B,C))	((Ari,Cor),Euc)	n	y		y		2219*
162	n_35801	((A,B),C)	((Ari,Cor),Euc)	n		y			556*
163	n_40263	(M.cor,(Cal.gra,(B,C)))	(Ari,(Cor,Euc))	n			y		1027
164	n_49139	(A,(B,C))	((Ari,Euc),Cor)	n		y			1767
165	n_BI096866	(M.cor,(Cal.gra(B,C)))	(Ari,(Cor,((Euc,Cor.cal),Euc)))	n			y		1418
166	n_BI096871	((A,B),C)	(Ari,(Cor,((Euc,Cor,cal),Euc)))	n			y		594*
167	n_BI096904	(M.cor,((C,Cal.gra),B))	(Ari,(Cor,((Euc,Cor.cal),Euc))	y			y		1217
168	n_BI097026	(M.cor,((C,Cal.gra),B))	(Cor.cal,(Ari,(Cor,Euc)))	n			y	y	801
169	n_3075	(M.cor,(Cal.gra,(B,C)))	((Ari,Euc),Cor)	y		y			997

170	n_3659	((A,C),B)	((Ari,Cor),Euc)	n	y				1342
171	n_4167	((A,C),B)	((Ari,Cor),Euc)	n	y				1672
172	n_4178	(A,(B,C))	((Ari,Cor),Euc)	n	y				2385
173	n_4460	((A,C),B)	((Ari,Euc),Cor)	n		y			1386
174	n_4589	((A,B),C)	((Ari,Cor),Euc)	n	y				1267
175	n_5190	(A,(B,C))	((Ari,Cor),Euc)	n	y				1557
176	n_5930	((A,B),C))	((Ari,Cor),Euc)	n	y				1878
177	n_27210	((Cal.gra,(C,(M.cor,B))	(Cor,(Ari,(Euc,Cor.cal))	y			y	y	1241
178	n_40152	((A,B),C)	((Ari,Euc),Cor)	n		y			1220*
179	n_40685	(A,(B,C))	(Ari,(Cor,Euc))	n			y		3934*
180	n_40951	(M.cor,((Cal.gra,C),B))	((Ari,Cor),Euc)	n	y				1760
181	n_41116	(A,(B,C))	((Ari,Euc),Cor)	n		y			926
182	n_41171	(A,(B,C))	(Ari,(Cor,Euc))	n			y		619
183	n_41451	((Cal.gra,(M.cor,B)),C)	((Ari,Euc),Cor)	y		y			3298
184	n_41541	((A,B),C)	((Ari,Euc),Cor)	n		y			1359*
185	n_41647	(A,(B,C))	((Ari,Euc),Cor)	n		y			1020*
186	n_41657	(M.cor,((C,Cal.gra),B))	(Ari,(Cor,Euc))	n			y		819*
187	n_42256	(B,(Cal.gra,(M.cor,C)))	((Ari,Euc),Cor)	n		y			1334*
188	n_42353	((A,B),C)	((Ari,Euc),Cor)	n		y			1345*
189	n_42511	(A,(B,C))	((Ari,Cor),Euc)	n	y				1400
190	n_42576	(A,(B,C))	(Ari,(Cor,Euc))	n			y		1249*
191	n_42717	(Cal.gra,(M.cor,(B,C)))	((Ari,Cor),Euc)	y	y				6588
192	n_42723	(Cal.gra,(M.cor,(B,C)))	((Ari,Cor),Euc)	y	y				1689*
193	n_hdr	(A,(B,C))	((Ari,Cor),Euc)	n	y				1421
194	n_idi1	((Cal.gra,(M.cor,A)),B)	(Ari,(Cor,((Euc,Cor.cal),Euc)))	n	y				646
195	n_BI096903	((A,B),C)	((Ari,Euc),Cor)	n		y			600*
196	n_4744	(A,(B,C))	((Ari,Cor),Euc)	n	y				1752

Table 2. A summary of relationships among *Melaleuca* in 196 gene trees. A, B, and C represent the three main clades identified in Edwards *et al*. (2010). *Calothamnus* is in clade A.

Loci	(A,(B,C))	((A,C),B)	((A,B),C)	The other	Total
Cpl	10	7	5	10	32
Ncl	48	23	41	52	164

Table 3. A summary of relationships among Eucalypteae in 196 gene trees. The abbreviations for the generic names in Eucalypteae are as the following: Ari (*Arillastrum*), Cor (*Corymbia*), Euc (*Eucalyptus*).

Loci	((Ari,Cor),Euc)	(Ari,(Cor,Euc))	((Ari,Euc),Cor)	The other	Total
Cpl	5	3	5	19	32
Ncl	72	40	40	12	164

Table 4. A summary of relationships within *Melaleuca* and Eucalypteae in phylogenies of concatenated loci. The abbreviations for the generic names in Eucalypteae are as the following: Ari (*Arillastrum*), Cor (*Corymbia*), Euc (*Eucalyptus*). A, B, and C represent the three main clades identified in Edwards *et al*. (2010). *Calothamnus gracilis*, which was transferred to *Melaleuca*, is in clade A.

	Concatenated loci	Relationships among the taxa	
		Eucalypteae	*Melaleuca* clades (A,B,C)
1	Cpl_all	((Aril,Cor),Euc)	(A,(B,C))
2	Cpl_ncl_all	((Aril,Cor),Euc)	(A,(B,C))
3	Ncl_46taxa	((Aril,Cor),Euc)	(A,(B,C))
4	Ncl_all	((Aril,Cor),Euc)	(A,(B,C))
5	Cpl_best	(Ari,(Cor,Euc))	(A,(B,C))
6	Ncl_best	((Ari,Cor),Euc)	(A,(B,C))

Figures.

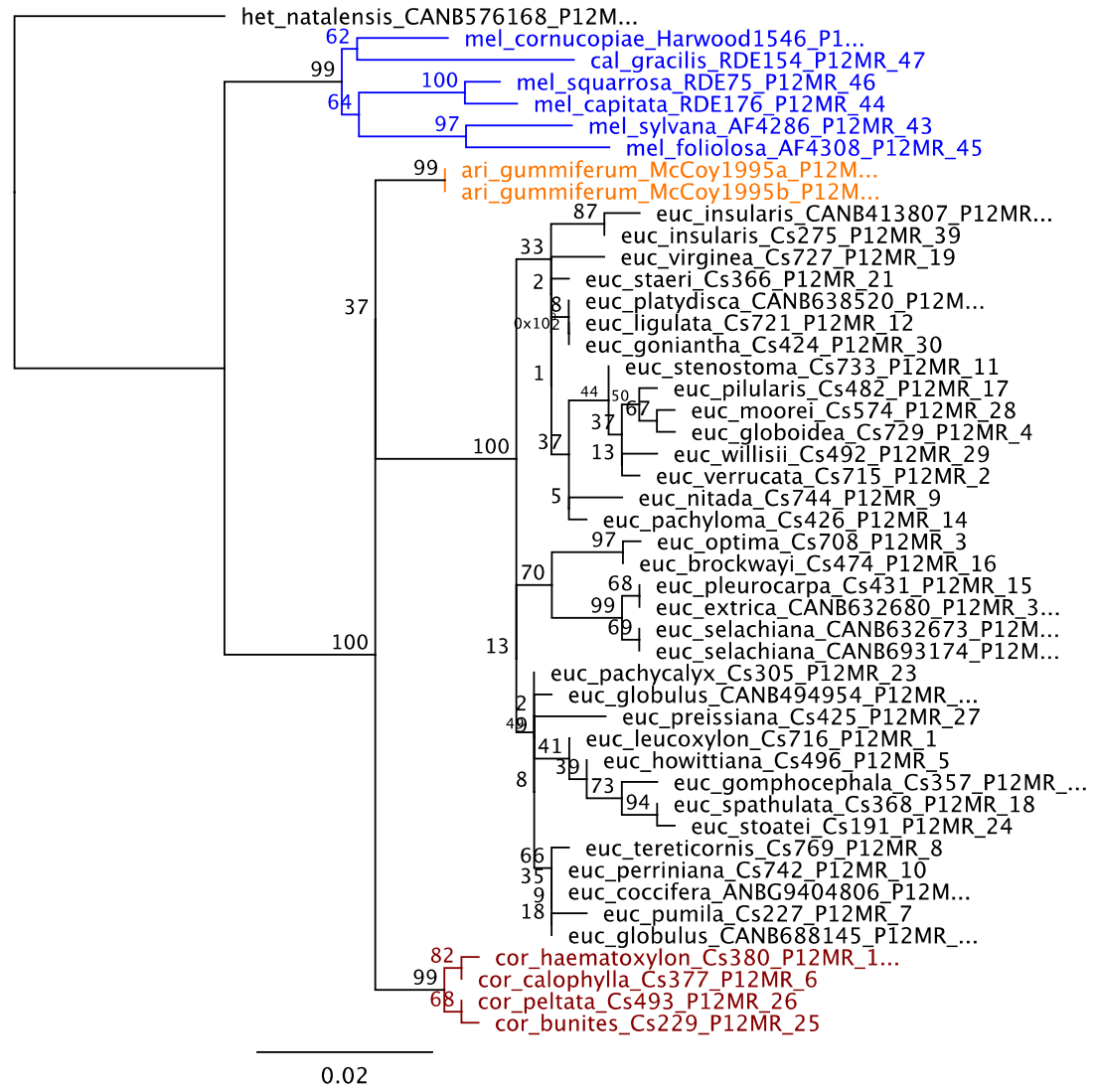

FIGURE 1. An example of maximum likelihood gene tree of a chloroplast locus c_165301165800 for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus name as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number consecutively. *Heteropyxis* was used to root the tree.

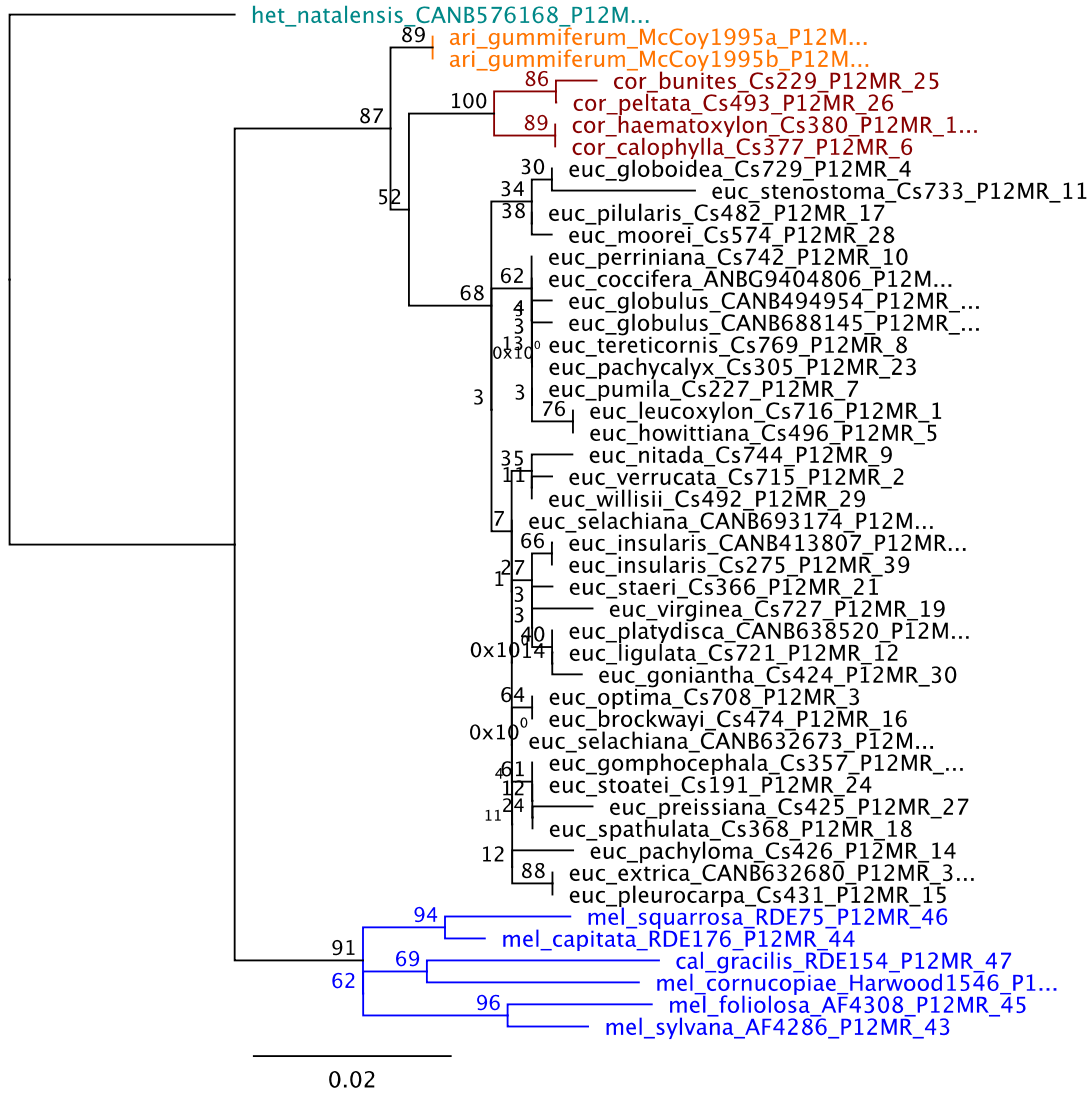

FIGURE 2. An example of maximum likelihood gene tree of chloroplast locus c_164501165000 for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number consecutively. *Heteropyxis* was used to root the tree.

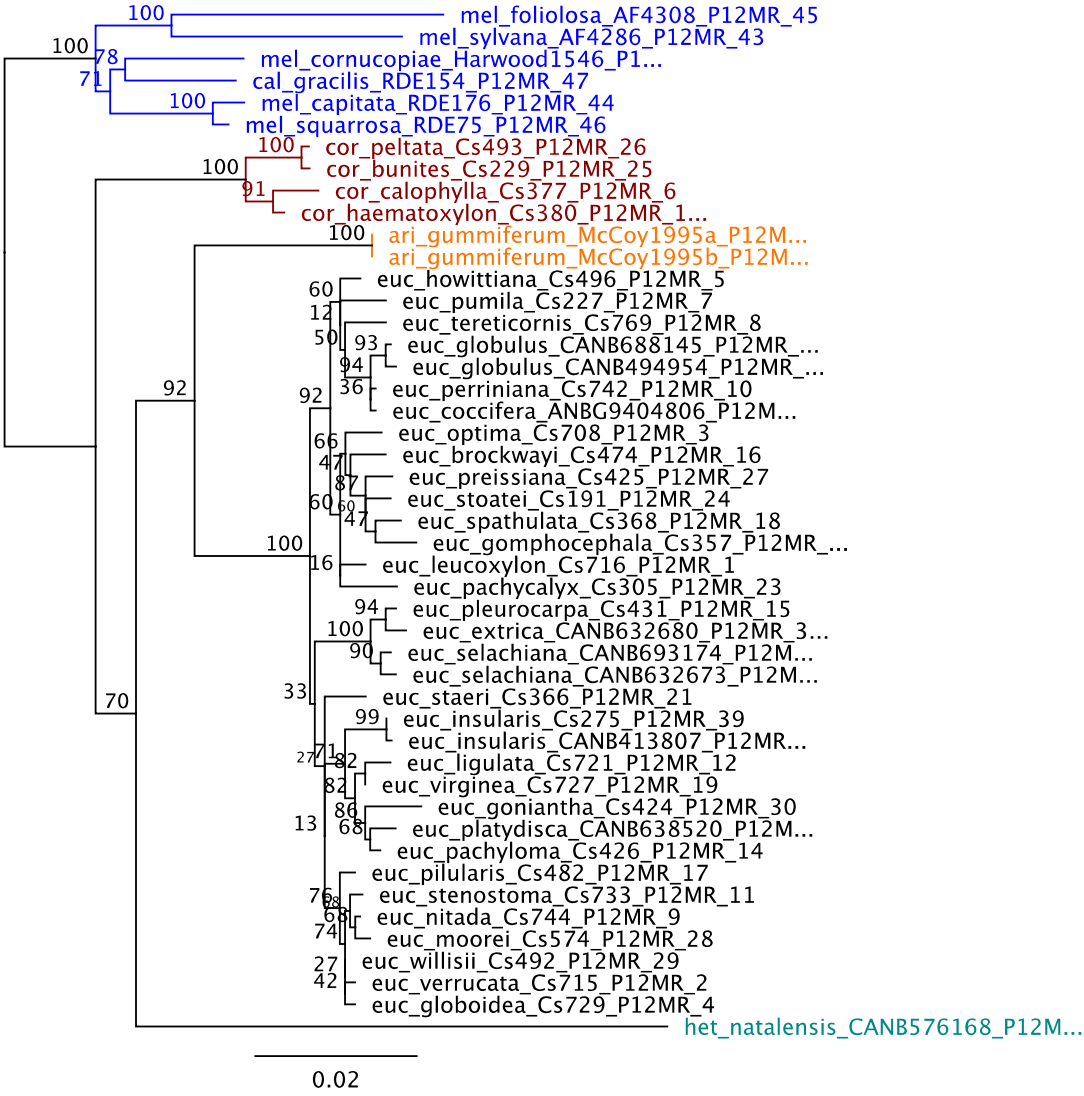

FIGURE 3. An example of maximum likelihood gene tree of a nuclear locus n_33521 for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows species name and a collection number consecutively.

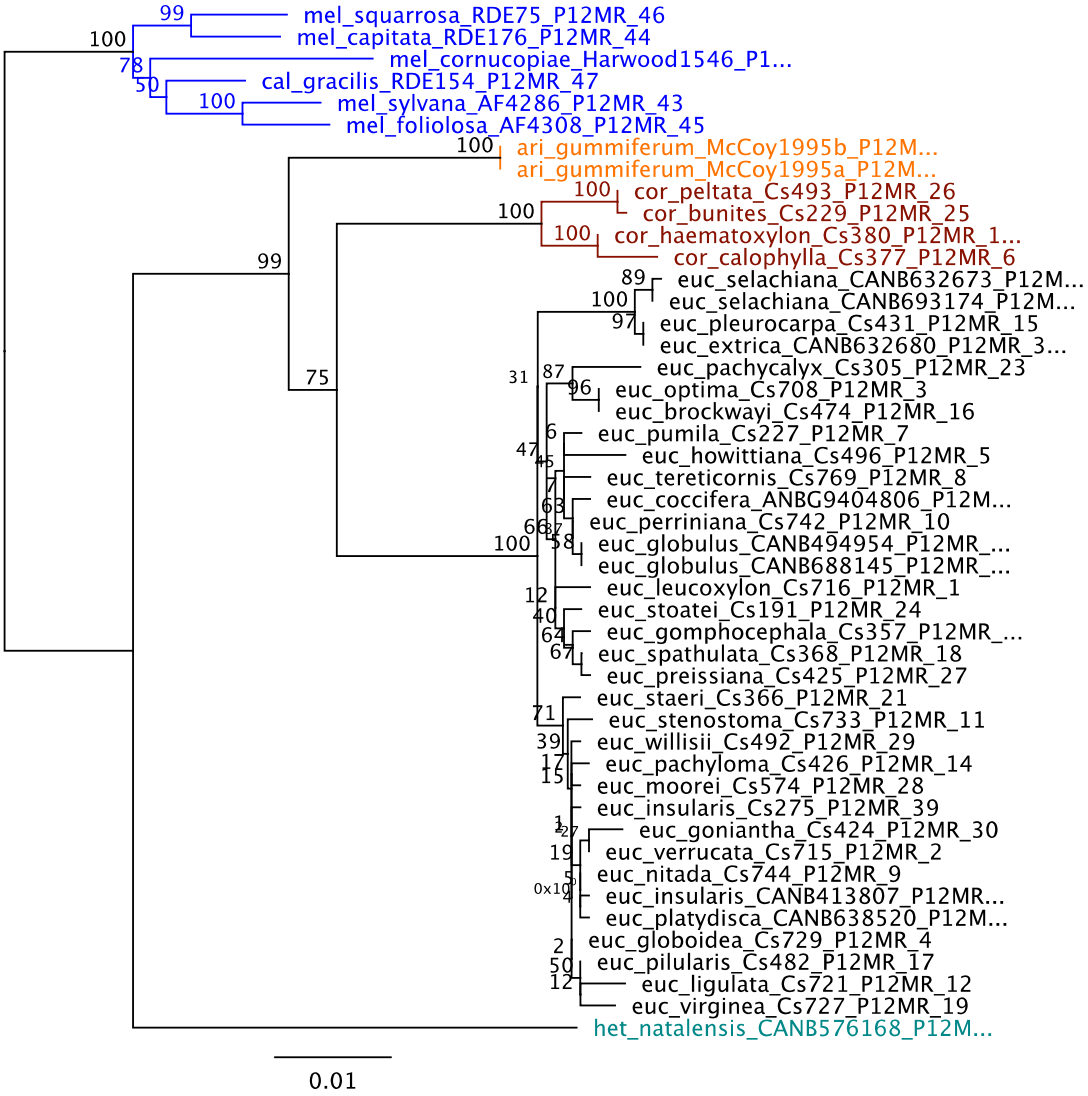

FIGURE 4. An example of maximum likelihood gene tree of nuclear locus n_41372 for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number consecutively.

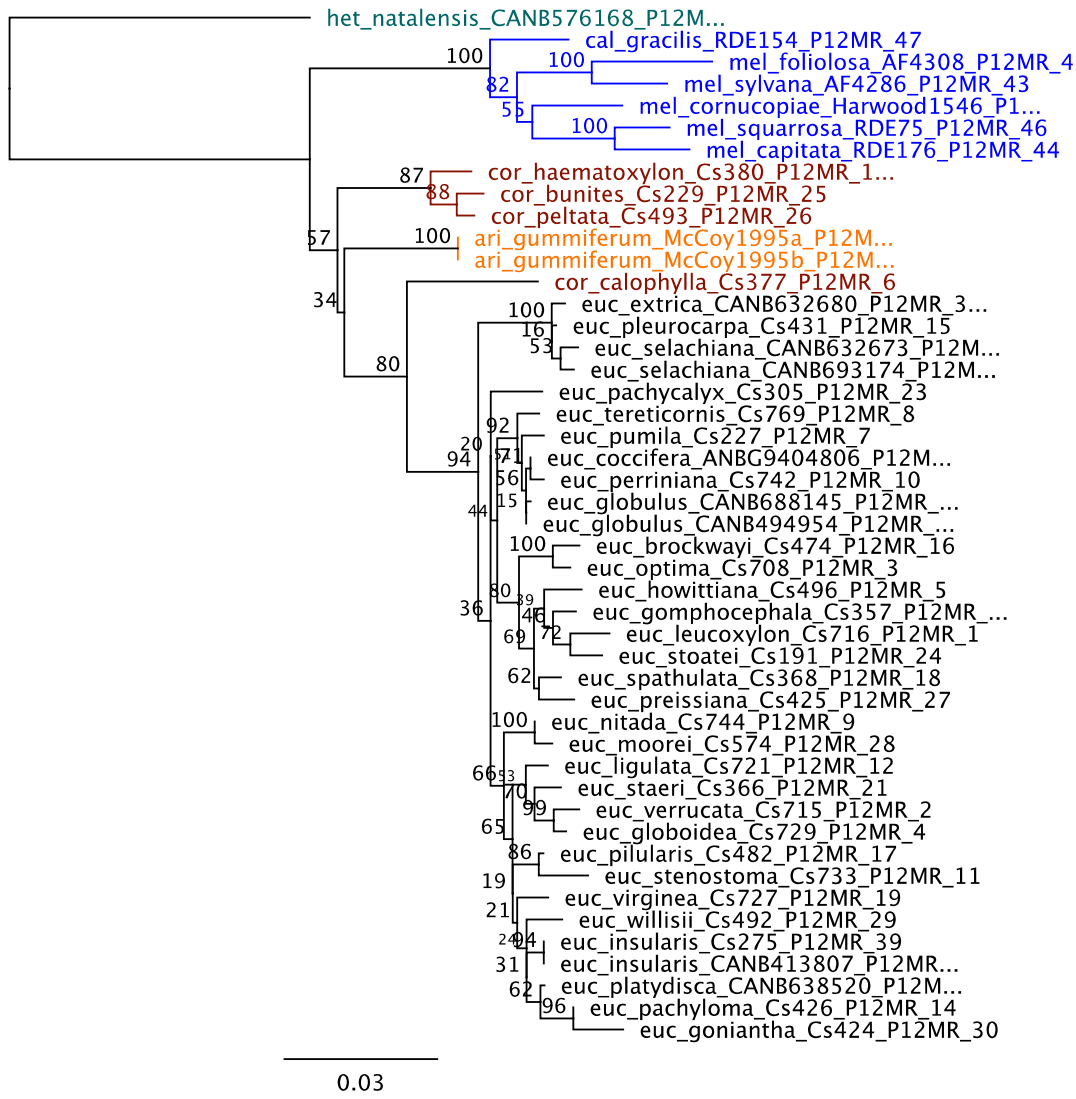

FIGURE 5. An example of maximum likelihood gene tree of nuclear locus n_27210 for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number consecutively. *Heteropyxis* was used to root the tree.

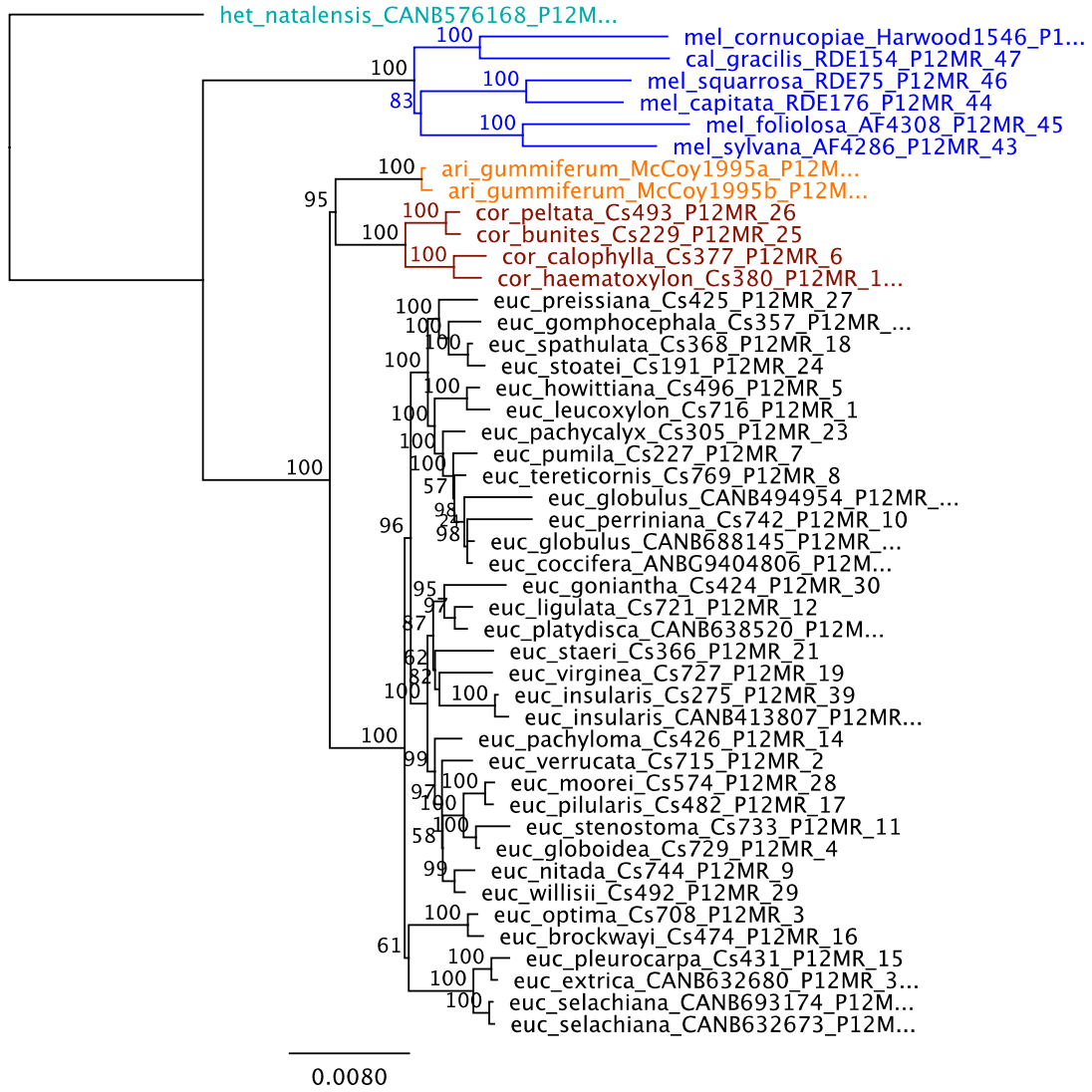

FIGURE 6. Maximum likelihood tree of 32 concatenated chloroplast loci (Cpl_all) for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus*, currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number consecutively. *Heteropyxis* was used to root the tree.

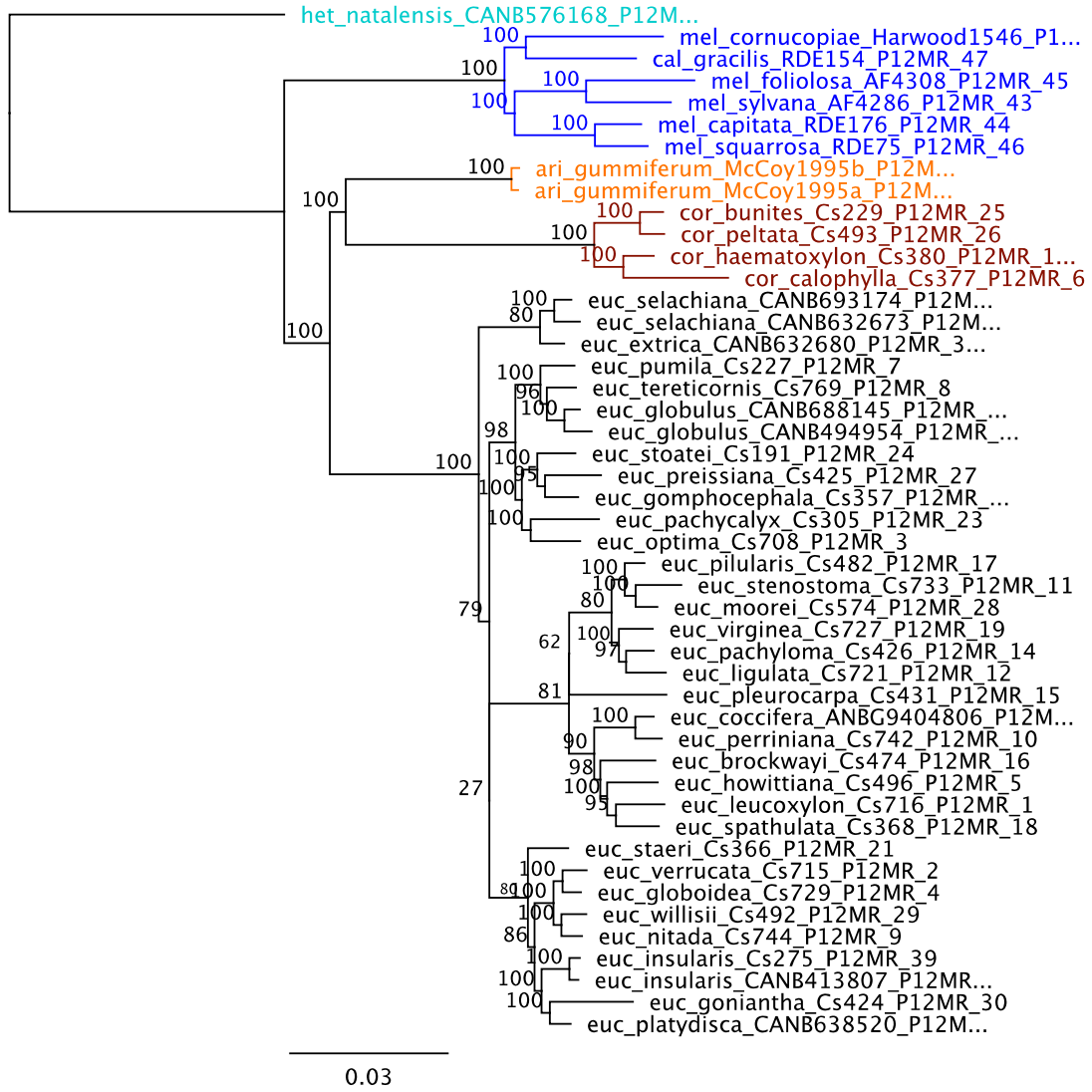

FIGURE 7. Maximum likelihood tree of 164 concatenated nuclear loci (Ncl_all) for 47 Myrtaceous samples with 100 bootstrap estimates. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name an underscore follows a species name and a collection number consecutively. *Heteropyxis* was used to root the tree.

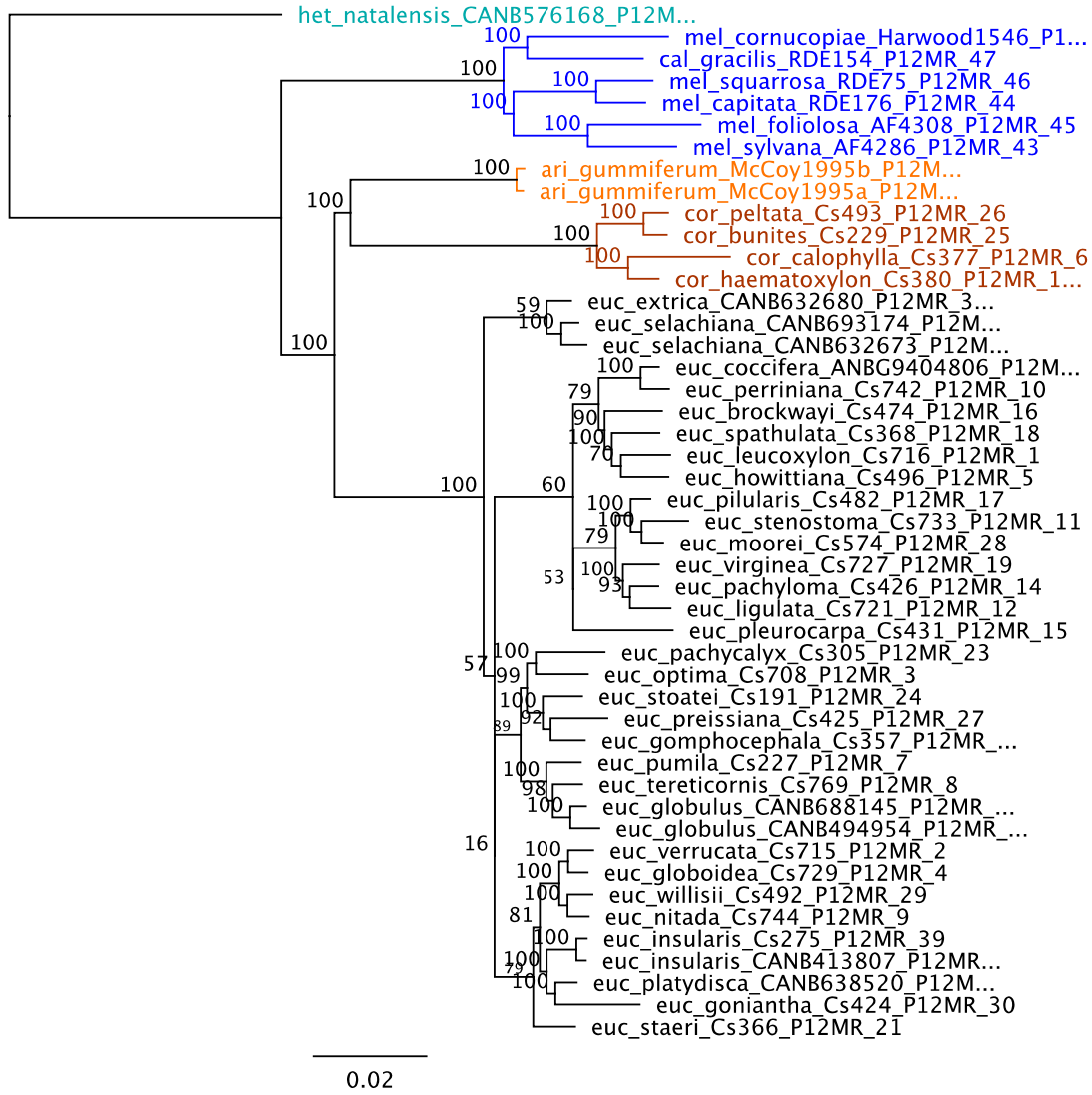

FIGURE 8. Maximum likelihood tree of concatenated 164 nuclear and 32 chloroplast loci (Cpl_ncl_all), with 100 bootstraps. 47 samples across the five genera of Myrtaceae were used. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus*, currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number. *Heteropyxis* was used to root the tree.

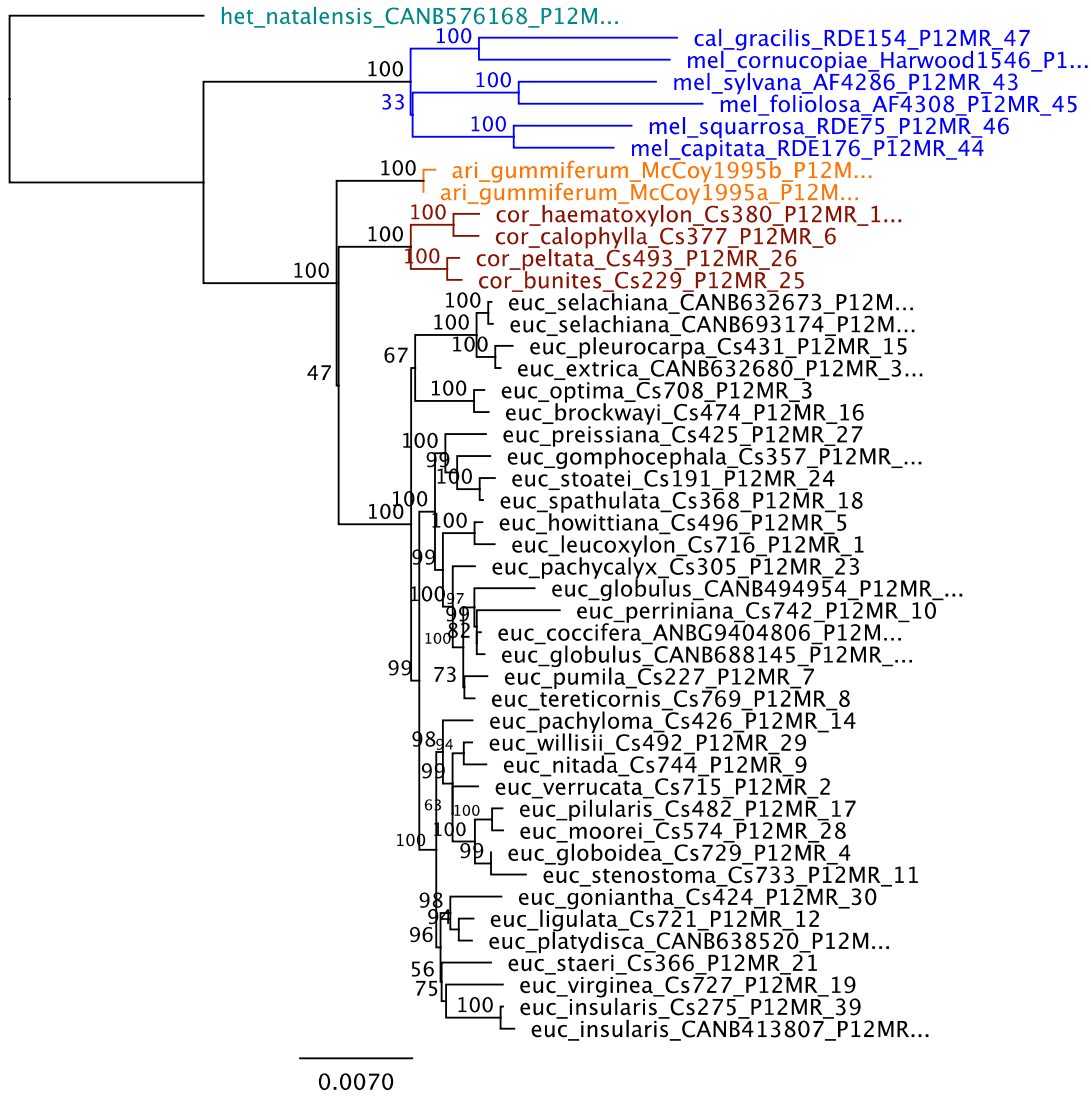

FIGURE 9. Maximum likelihood tree of concatenated 25 chloroplast loci (Cpl_best) for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as the following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus*, currently transferred to *Melaleuca*), het (*Heteropyxis*). After the generic name follows a species name and a collection number. *Heteropyxis* was used as an outgroup taxon in this study.

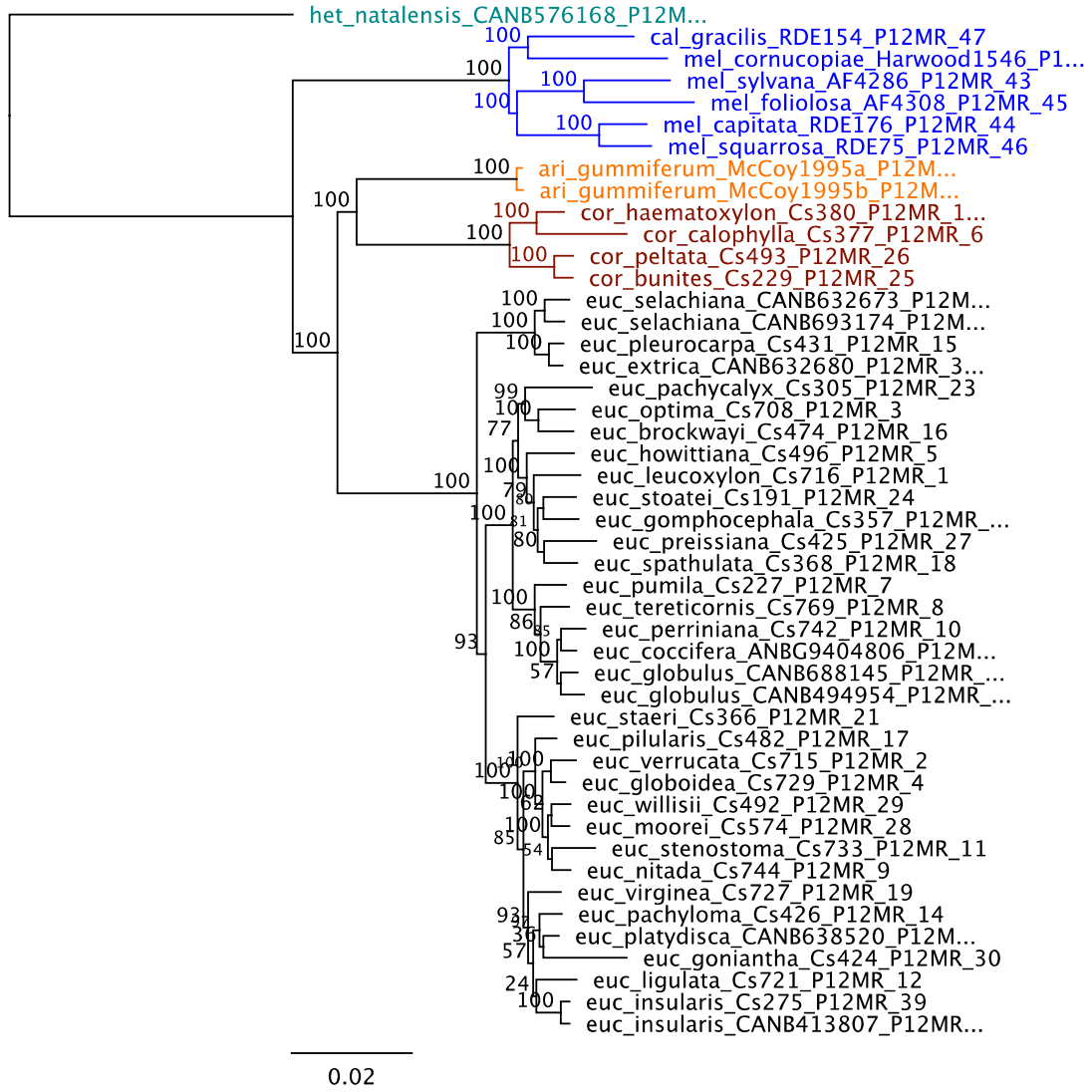

FIGURE 10. Maximum likelihood tree of concatenated 47 nuclear loci (Ncl_best) for 47 Myrtaceous samples with 100 bootstraps. The three digits at the beginning of each sample name are abbreviations for the genus names as following: euc (*Eucalpytus*), ari (*Arillastrum*), cor (*Corymbia*), mel (*Melaleuca*), cal (*Calothamnus,* currently transferred to *Melaleuca*), het (*Heteropyxis*). After a generic name follows a species name and a collection number consecutively. *Heteropyxis* was used as an outgroup taxon in this study.

Chapter 5

Discussion and concluding remarks

Taxonomy of Melaleuceae *sensu* Wilson *et al*. (2005; *Melaleuca sensu lato*) and *Melaleuca* in the strict sense (e.g. Cook *et al*. 2008; Craven 2006; Craven *et al.* 2014; Edwards *et al.* 2010; Udovicic & Spencer 2012) has been investigated in the past but a number of questions surrounding the evolution and phylogeny of this group remain unanswered. This study set out with an ultimate aim of providing a more robust molecular phylogeny of *Melaleuca sensu lato* with better representations of taxa using more genetic markers. In the following paragraphs, I will summarise and discuss the main findings of Chapters 2, 3 and 4. Finally, I will discuss future directions for phylogenetic and evolutionary studies in *Melaleuca*.

Chapter 2 was to provide a snapshot of the problematic taxonomy within the genus *Melaleuca*, specifically the *Melaleuca leucadendra* complex, focusing on resolving the taxonomy of a unique biotype within this complex. Resolving the taxonomy of the *M. leucadendra* complex has been problematic because most species in this group lack morphological autapomorphies, and molecular phylogenies to date have not been sufficiently comprehensive to resolve the taxonomy (Cook *et al.* 2008). In this study, the genetic and morphological diversity of *Melaleuca argentea* (within the *M. leucadendra* complex) was explored to assess the taxonomic status of the *M. argentea* "Ashburton biotype" in the Pilbara region of Western Australia. This biotype is a potentially isolated population of *M. argentea* that was reported to possess a unique chloroplast haplotype as well as a number of private nuclear ITS alleles (Edwards *et al.* 2012). While previous description of the taxon detailed some unique morphological and genetic characteristics the data were not sufficient to resolve the taxonomic status of the biotype. However, the new morphological results presented in this Chapter suggest that, contrary to previous assumptions, the Ashburton biotype is more closely related to *M. leucadendra* than to *M. argentea*. To provide a more definitive taxonomic status of the *M. argentea* "Ashburton biotype", it was concluded that further testing for gene flow between *M. argentea* "Ashburton biotype" and *M. leucadendra* using additional nuclear molecular markers would help elucidate this problem.

Given the taxonomical complexity of the genus *Melaleuca*, illustrated by case study presented in Chapter 2, the primary aim of this thesis was to provide a more robust molecular phylogeny of *Melaleuca sensu lato* with better representations of taxa and many more genetic markers, to help resolve these outstanding taxonomic dilemmas within the genus. To achieve the goal, in Chapter 3, a workflow used to identify

orthologous and low copy genetic markers for *Melaleuca* phylogenetic analyses was developed. In Chapter 4, the genetic markers identified in Chapter 3 were evaluated and assessed using a gene tree approach.

In Chapter 3, the annotated genome of *Eucalyptus grandis* combined with transcriptomes of *Melaleuca quinquenervia* and *M. alternifolia,* whole genome shotgun sequences of *Melaleuca leucadendra* and *M. bracteata* were used to locate genetic marker sets covering the whole family Myrtaceae with a focus on target genera *Melaleuca* and *Eucalyptus*. To identify low copy number and orthologous nuclear marker sets we first mapped randomly selected *Melaleuca quinquenervia* transcriptome contigs and assembled them against the whole shotgun *Melaleuca* genome sequences. After removing the candidate contigs that are high in coverage, we blasted selected contigs against the *Eucalyptus grandis* genome on the Phytozome website to attain the matching *E. grandis* coding DNA sequences and select loci with low copy number. Then, a gene tree approach was followed to filter out potential paralogous loci. In order to locate chloroplast markers, we aligned the chloroplast genomes from two *Melaleuca* species (*M. bracteata, M. leucadendra*) and *Eucalyptus grandis.* In total, we identified over 260 kbp of 212 nuclear and chloroplast loci to target. Of those, 209 loci were mostly successful in sequencing after performing on exon capture on 43 species of five genera of the Myrtaceae family.

In Chapter 4, we assessed gene trees of 196 loci that were identified in Chapter 3 after excluding 13 loci due to low average coverage. We used 47 samples from five genera in Myrtaceae. We found 134 loci that are useful for *Melaleuca* and 174 loci for *Eucalyptus, Corymbia* and *Arillastrum.* Cases where gene trees of *eucalypts* did not cluster by genus, were found to be due to low coverage in few *Eucalyptus* and *Corymbia* samples. In 92% of gene trees, *Calothamnus gracilis* was nested within the rest of the *Melaleuca* species, this result supports the inclusion of *Calothamnus* to *Melaleuca* (Craven *et al*. 2014). Maximum likelihood species trees of concatenated sequence datasets from *Melaleuca* consistently had a topology of (A,(B,C)) into the same three clades as described by Edwards *et al*. (2010). The relationship within Eucalypteae was ((*Arillastrum*,*Corymbia*),*Eucalyptus*) in the concatenated datasets except in a tree based on concatenated 25 chloroplast loci (see Chapter 4 discussion). However, to elucidate the genus and species relationships within *Melaleuca* and *eucalypts* other methods, such as coalescence methods, should be investigated that takes individual gene histories

into account. Combining the marker sets described in this work will provide sufficient information to explore genus to species level phylogenies in *Melaleuca*. Although some of the internal branches remained unresolved further information available from the intronic regions, which are also captured by our methods but not looked at in this study, can provide valuable data to resolve some of the unresolved relationships. The genetic markers are applicable not only for *Melaleuca* and *Eucalyptus* but also other groups in Myrtaceae.

With the scarcity of genetic markers that are applicable across *Melaleuca sensu lato* and eucalypts, exploration of phylogenetic relationshps at different taxonomic levels have been challenging. The marker sets for Myrtaceae discovered in this study are applicable to both *Melaleuca* as well as Eucalyts in Mytaceae, and potentially to other Myrtaceous genera. Considering that only exon sequences were taken into account for this thesis, overhanging introns of the exons may also be a useful source to provide finer resolution in the phylogeny.

Future work applying these loci sets (from Chapters 3, 4) to as many species as possible in *Melaleuca* will provide a robust backbone to explore taxonomic and evolutionary questions in the groups. The genus boundary of *Melaleuca* continues to be controversial (e.g. Craven 2006; Craven *et al.* 2014; Edwards *et al.* 2010; Udovicic & Spencer 2012). In addition, a new phylogeny based on the marker sets could provide an opportunity to further explore the problematic relationships and taxonomy within *Melaleuca* in the strict sense (e.g. *M. leucadendra* complex), such as those described in Chapter 2. By testing the monophyly of *Melaleuca* more rigorously the generic boundary can be verified and the new phylogeny will be able to reflect the evolutionary history of *Melaleuca sensu lato* more accurately. Apart from the generic boundary of *Melaleuca*, many other taxonomic issues remain within the genus. The phylogenetic analyses by Edwards et al. (2010) showed non-monophyly of many species-groups (e.g. *M. pungens, M. uncinata, M. cuticularis, M. ordinifolia, M. fulgens, M. laxiflora, M. lanceolata, M. huegelii* groups) as well as paraphyly at species level in *M. fulgens and M. bracteata*. Phylogenetic studies of the groups using the markersets might be useful to eludicate the species relationships and the evolutionary histories of the groups. Species boundary problems as in the case of *M. argentea* "Ashburton biotype" can be explored. In order to achieve this selecting appropriate genetic markers to target different taxonomic level is inevitable. Inclusion of overhanging intronic regions could be

important to resolve some of the problems. As to how many loci could be informative for estimations of robust phylogeny for which classification rank (e.g. species, genus level) in *Melaleuca* or eucalypts from the marker sets (in Chapter 3), remains to be tested. Valderrama *et al.* (2018) introduced a method to visualise nodal supports to identify level of inclusion of locus, sites for different taxonomic level of phylogeny and their approach can be applied to answer the question.

A more robust phylogeny will provide opportunities to investigate on the evolution of the species diversity within *Melaleuca sensu lato* and the rates of diversification and historical changes in environmental parameters such as, climate. In addition, a robust phylogeny will allow us to do rigorous hypothesis testing on eg. adaptation of this group into different soil types or habitats. *Melaleuca sensu lato* is distributed in diverse habitats across Australia being found in all biomes (southwestern temperate sclerophyll biome, southeastern temperate sclerophyll, monsoonal, arid biome, everwet biome), and is well represented in diversity hotspots. In the southwestern temperate sclerophyll biome (South Western Floristic Region; Crisp *et al.* 2004; Crisp & Cook 2013; Kershaw *et al*. 1994), Myrtaceae are the most diverse angiosperm family, and the *Melaleuca sensu lato* is one of the largest genera. Interestingly, the species diversity and the distribution of the species in morphologically diverse *Melaleuca* clade A (Edwards *et al*. 2010), which in the past included eight genera (*Beaufortia*, *Calothamnus*, *Conothamnus*, *Eremaea*, *Lamarchea*, *Petraeomyrtus*, *Phymatocarpus*, *Regelia*; Wilson *et al*. 2005), is centered in the South Western Floristic Region. The high biodiversity of *Melaleuca sensu lato* in the southwest (SW) temperate Australia contrasts to its relatively low biodiversity in the southeast (SE) temperate biome. Whether *Melaleuca sensu lato* is more diverse than expected in the Southwest Autralian Floristic Region (SWAFR) remains to be tested, but seems likely. Investigation into the reasons for the occurrence of high endemism in the SWAFR, whether it is by a unique combination of ecological factors that restrict the distribution of *Melaleuca sensu lato,* or whether it is by soils and topography (e.g. isolated outcrops, fine scale patterning of soil), climate change, or fire, could be further understood by using a robust phylogeny for the genus. Whether the SW clades went through high speciation, or the SE clades experienced high extinction or a combination of both has happened could also be tested. Testing if the contrasting diversity is because of recent radiation and how immigrants have affected the diversity will help understanding the distribution pattern.

There are a number of future research paths that can be investigated using our newly developed loci. For example, the evolutionary relationships of gall-inducing scale insects with their host, *Melaleuca* can be studied. Most of the galling insects on this group have not been described or studied. Shapes of galls are species-specific and they sometimes look like a fruit or an organ not found on the plant (Cook & Gullan 2004). By matching the phylogenies of host and parasite, the occurrence of cospeciation could be tested (e.g. Hafner & Page 1995). A molecular phylogeny is a useful framework for studying cospeciation as molecular data allow comparisons of host and parasite divergence (Page & Holmes 1998). Molecular divergence might give answers to two main questions of host and parasite phylogeny: 1) how often host switching has occurred, 2) presence of multiple lineages of parasites on each host (Page & Holmes, 1998). Given estimates of the amount of divergence between Melaleuceae and gall insects, rates of evolution can be compared (Page and Holmes, 1998).

References

Cook, L.G. & Gullan, P. (2004) Gall induction has evolved multiple times among the eriococcid scale insects (Sternorrhyncha: Coccoidea: Eriococcidae). *Biological Journal of Linnean Society* 83: 441-452.

Cook, L.G., Morris, D.C., Edwards, R.D. & Crisp, M.D. (2008) Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. *Molecular Phylogenetics and Evolution* 47: 506-522.

Craven, L.A. (2006) New combinations in *Melaleuca* for Australian species of *Callistemon* (Myrtaceae). *Novon: A Journal for Botanical Nomenclature* 16: 468-475.

Craven, L.A., Edwards, R.D. & Cowley, K.J. (2014) New combinations and names in *Melaleuca* (Myrtaceae). *Taxon* 63: 663-670.

Crisp, M., Cook, L. & Steane, D. (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 359: 1551-1571.

Crisp, M.D. & Cook, L.G. (2013) How was the Australian flora assembled over the last 65 million years? A molecular phylogenetic perspective. *Annual Review of Ecology, Evolution, and Systematics* 44: 303-324.

Edwards, R.D., Craven, L.A., Crisp, M.D. & Cook, L.G. (2010) *Melaleuca* revisited: cpDNA and morphological data confirm that *Melaleuca* L. (Myrtaceae) is not monophyletic. *Taxon* 59: 744-754.

Hanfer, M.S. & Page, R.D. (1995) Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 349: 77-83.

Kershaw, A.P., Martin, H.A. & Mason, J.M. (1994) The Neogene: a period of transition. History of the Australian vegetation: Cretaceous to Recent. Cambridge Press, Camgridge.

Page, R.D.M. & Holmes, E.C. (1998) Molecular evolution: a phylogenetic approach. Blackwell Publishing, Malden.

Udovicic, F. & Spencer, R.D. (2012) New combinations in Callistemon (Myrtaceae). *Muelleria* 30: 23-25.

Valderrama, E., Richardson, J.E., Kidner, C.A., Madriñán, S. & Stone, G.N. (2018) Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L.f., Zingiberaceae). *Molecular phylogenetics and evolution* 119: 13-24.

Wilson, P.G., O'Brien, M.M., Heslewood, M.M. & Quinn, C.J. (2005) Relationships within Myrtaceae *sensu lato* based on *matK* phylogeny. *Plant Systematics and Evolution* 251: 3-19.