Genomic diversity of *Leptospermum*

Australian Manuka development

December 2021

CRCHBP
FOR HONEY BEE PRODUCTS

ISBN

ISSN

TITLE: Genetic diversity of *Leptospermum:* Australian Manuka development

Publication No.

Project: Program 1, Project 8

Researcher Contact Details

Rachel Binks
Department of Biodiversity, Conservation and Attractions
Locked Bag 104, Bentley Delivery Centre, WA 6983
Email: rachel.binks@dbca.wa.gov.au

Electronically published by CRC for Honey Bee Products at www.CRChoneybeeproducts.com

FORWARD

The basis of any plant improvement program is to understand the genetic structure of the plant material. Not only will this give an understanding of the gains that could be made but also save time with knowing what natural crosses can be made between plant material to produce the greatest benefit.

What was not expected was the total re-structuring of the Leptospermum complex as originally defined by Joy Thompson in 1989. Suggested is to split the Leptospermum into five genera, and the finalisation of these new genera, and species allocation, has been left with the taxonomists.

What this information immediately provided was groupings within which to focus for plant improvement. Crossing between the new genera is likely to be difficult and require more challenging scientific techniques for success. So, where it was hoped that the drought-tolerance and high dihydroxyacetone (DHA) production in the nectar of *L. nitens* could be crossed with *L. scoparium*, we now know this would be an unlikely natural hybrid.

When the Western Australian *L. Nitens* complex was further investigated, interesting was the lack variation within a provenance. Whilst the mechanism for this lack of variation is still to be unravelled, this has an impact on seed collection for the establishment of breeding trials and their conversion into seed orchards.

Of interest were the genera interspersed between the *Leptospermum* complex, such as *Kunzea, Agonis, Pericalymma, Astreromyrtus* and *Neofabricia*. This suggested that the production of DHA in the nectar may not necessarily be confined to this *Leptospermum* complex but be more widely produced across a number of genera. This has since been shown to occur (Project report: Unravelling the secrets of nectar production for bioactive honey).

This discovery has major implications for bioactive honey production in Australia. We know we have two very different mechanisms by which our honey can attain high antibacterial activity, that is either through the enzymatic peroxide activity or through this mechanism of the non-enzymatic conversion of DHA to methylglyoxal, which has the antibacterial properties. The regular testing of honey for antibacterial activity will become more common until there is a better understanding of the extent of our bioactive honeys.

For the honey bee, this work also has implications. Whereas plantation establishment for high value medicinal Manuka honey was focussed on a single species, *L. scoparium*, these plantings can now become more biodiverse, and hopefully include better pollen-producing species than *L. scoparium* to sustain bee health.

Dr Liz Barbour

CEO

CONTENTS

FIGURES

ABOUT THE AUTHORS

Dr Rachel Binks is a Senior Research Scientist in plant genetics at the Department of Biodiversity, Conservation and Attractions (DBCA). Her research uses population genomic and phylogenomic data to investigate the evolutionary processes occurring across populations, species and landscapes. This research directly informs the conservation management of Western Australian flora, including the resolution of taxonomic issues, identification of management units, impact assessments and informing translocation activities.

Dr Margaret Byrne is recognised as a leading plant geneticist in Australia with over 300 publications. Her research has focused on plant genetics to inform conservation strategies for rare and threatened species, as well as biodiversity conservation at landscape scales in relation to remnant viability, revegetation and adaptation to climate change. Her phylogeographic studies have provided a greater understanding of the evolutionary history of Australian flora, and its influence on current distributions, genetic diversity and location of refugia.

ACKNOWLEDGMENTS

We are grateful to Margaret Heslewood and Peter Wilson at the Royal Botanical Gardens, Sydney for their support with data analyses and sharing their taxonomic knowledge of the Leptospermeae. We thank Karina Knight (Western Australian Herbarium) and Brendan Lepschi (Australian National Herbarium) for access to herbarium specimens, and Bronwyn Macdonald (DBCA) and Linda Broadhurst (CSIRO, Canberra) for facilitating DNA extractions. We also thank Sarah Matthews at CSIRO, Canberra for bioinformatic advice, as well as Michael Hislop and Rob Davis at the Western Australian Herbarium for lengthy discussions and advice regarding taxonomic identifications of Western Australian *Leptospermum*.

ABBREVIATIONS

bp	base pair
BLAST	basic local alignment search tool
cpDNA	DNA from the chloroplast genome
DArT	diversity arrays technology
DBCA	Department of Biodiversity, Conservation and Attractions
DHA	dihydroxyacetone
DNA	deoxyribonucleic acid
ETS	external transcribed spacer
F1	first generation hybrid
F2	second generation hybrid
ITS	internal transcribed spacer
nDNA	DNA from the nuclear genome
rRNA	ribosomal ribonucleic acid
tRNA	transfer ribonucleic acid

Executive Summary

Australia is in a unique position to expand the bioactive honey industry across the growing number of *Leptospermum* species identified to produce dihydroxyacetone (DHA). However, there are a number of outstanding taxonomic issues across the genus, and this limits a full exploration of the commercial interests in this group because a sound taxonomic foundation is necessary to establish breeding programs, to provide consumer confidence in the source of products, and to maximise the quality of product produced. Thus, the resolution of taxonomic issues in *Leptospermum* is vital for the growth of the Australian bioactive honey industry. This report describes the results of two molecular studies that aimed to improve knowledge of the genetic diversity and taxonomic status of Australian *Leptospermum*.

The first study aimed to resolve the longstanding issue of polyphyly and determine appropriate generic boundaries for *Leptospermum*. In doing so, genome skimming was used to compile the chloroplast genome, as well as the nuclear ribosomal region, across 110 accessions of 38 *Leptospermum* species, six closely allied genera and five outgroup genera. Maximum likelihood and Bayesian analyses resolved congruent clades for the chloroplast (132, 143 bp) and nuclear alignments (1,219 bp) to provide a robust interpretation of evolutionary relationships. Together, these data confirmed extensive polyphyly of *Leptospermum*, separating the genus into five monophyletic clades that were spread amongst clades representing six closely allied genera: *Agonis*, *Asteromyrtus*, *Homalospermum*, *Kunzea*, *Neofabricia* and *Pericalymma*. These data provide a clear pathway for taxonomic revision; the *Leptospermum* genus needs to be split into five genera representing each of the monophyletic clades. Moreover, representatives from four of the five putative genera are known to produce DHA in high concentrations, and this suggests that species from other, currently untested genera may also be promising for bioactive honey exploration.

The second study aimed to delimit species boundaries within one of the Western Australian clades identified in the first study. The species in this clade are very closely related and are difficult to identify in the field, including two species of interest to the bioactive honey industry, *L. nitens* and *L. roei*. In this study, 955 plants were sampled across 90 locations and nine species, and genome-wide SNP markers were used to assess the genetic variation across them. These data recovered clear separation of *L. maxwellii*, *L. sericeum* and *L. inelegans* as they are currently circumscribed. *Leptospermum nitens* and *L. roei*, which only differ by a single morphological character that is variable within other species, were genetically indistinct and therefore, represent the same species. There was also no morphological or genetic distinction between *L. incanum* and *L.* sp. Peak Charles/Norseman. *Leptospermum erubescens* and *L. oligandrum* were very closely related and represent up to three species. Widespread hybridisation was detected between most pairs of species that geographically co-occur. Unexpectedly, however, extensive clonality within, and sometimes between, populations was detected for all species except *L. maxwellii*, such that the genotypic diversity within the populations of most species is very low and is suspected to be the result of apomixis.

The results of these studies resolve some of the long-standing taxonomic issues in *Leptospermum* and have important implications for the bioactive honey industry. (1) The primary species currently used in bioactive honey production (*L. scoparium* and *L. polygalifolium*) will retain the name *Leptospermum* but

other species, including some that are of interest or are in development for bioactive honey production (e.g. *L. nitens*, *L. whitei*, *L. speciosum*) will be re-named as new genera. This raises the possibility that other, currently untested genera may also yield valuable species for bioactive honey exploration and opens a branding opportunity to establish a broader market for exclusively Australian bioactive honey. (2) The synonymisation of *L. nitens* and *L. roei* sharpens the focus for bioactive honey production in Western Australian and provides a wider range of genotypes for plant selections. (3) The low genotypic diversity detected within populations of *L. nitens* has major implications for seed collections and breeding trials but the suspected cause, apomixis, may be of downstream benefit to maintain and propagate highly valued genotypes in cultivation.

The studies reported here have demonstrated the value of using genomic data to resolve taxonomic issues in *Leptospermum* and we recommend that further studies are undertaken to resolve other known issues in the genus. We also recommend that future research continues testing DHA production across a range of other species and closely related genera to broaden the pool of flora for bioactive honey production. Finally, further work is needed to assess the extent of clonality and confirm the presence of apomixis in *L. nitens* to inform the expansion of the bioactive honey industry in Western Australia using local species.

Figure 1: *Leptospermum nitens* growing in the wheatbelt of Western Australia.

Academic outputs

Two scientific papers arer published in peer-reviewed, international journals:

Binks RM, Heslewood M, Wilson PG & Byrne M (2022) **Phylogenomic analysis confirms polyphyly of *Leptospermum* and delineates five major clades that warrant generic recognition**. *Taxon* 71.2: 348-359. https://doi.org/10.1002/tax.12650

ABSTRACT: *Leptospermum* is an ecologically and economically important genus with a long unresolved taxonomic issue concerning polyphyly, as indicated from early molecular analysis on two chloroplast regions. To resolve this, we use genome skimming to obtain high copy chloroplast and nuclear ribosomal DNA for a comprehensive phylogenomic analysis of 110 accessions of 38 *Leptospermum* taxa, six closely allied genera and five outgroup genera. Maximum likelihood and Bayesian analyses resolved congruent clades for the chloroplast (132,143 bp: 80 CDSs, four rRNA genes, 29 tRNA genes, 17 introns and 97 IGSs) and nuclear alignments (1,219 bp: ITS1, ITS2, ETS, 5.8S) to provide a robust interpretation of evolutionary relationships. Together, these data confirmed extensive polyphyly of *Leptospermum* that separated the genus into five monophyletic clades spread amongst clades representing six closely allied genera: *Agonis*, *Asteromyrtus*, *Homalospermum*, *Kunzea*, *Neofabricia* and *Pericalymma*. These five *Leptospermum* clades share some similarities with morphological and genetic groupings identified previously but provide greater resolution to inform a clear pathway to taxonomic revision. The evidence presented here provides support for resolution of the current polyphyly of *Leptospermum* through recognition of five genera, while retaining all other genera of Leptospermeae in their current circumscription.

Binks RM & Byrne M (2022) **Species delimitation, hybridisation and possible apomixis in a rapid radiation of Western Australian Leptospermum (Myrtaceae)**. Botanical Journal of the Linnean Society 1-17 https://doi.org/10.1093/botlinnean/boac022

ABSTRACT: Species delimitation is challenging in rapid radiations because the typical markers of speciation are often obscured. Here, we use comprehensive sampling and genome-wide SNPs to assess species boundaries in a radiation of nine morphologically similar Leptospermum taxa that failed to discriminate in previous phylogenomic analyses. Our data recovered clear separation of L.

maxwellii, L. sericeum and L. inelegans as currently circumscribed. A phrase-named taxon, L. sp. Peak Charles/Norseman, was not distinct from L. incanum and we recommend their synonymisation. Another pair, L. nitens and L. roei, were also indistinct and differ by a single morphological character that also varies within L. inelegans without taxonomic recognition. We recommend synonymisation of L. nitens and L. roei and consistent treatment of this character as a non-diagnostic, variable trait. Difficulty arose in discriminating L. erubescens and L. oligandrum; we make three suggestions and recommend further morphological investigation to determine the most appropriate taxonomic outcome. As expected, hybridisation was common across the complex, but unexpectedly, many individual plants were genetically identical within, and sometimes between, populations of most species. We hypothesise that this represents apomixis. Overall, this study demonstrates the value of population genomics in the integrative taxonomy toolbox for disentangling species in rapid radiations, while also offering insight to the evolution of this poorly known group of Australian Leptospermum.

Introduction

Leptospermum is an economically important genus of evergreen shrubs and small trees in the Myrtaceae family that are largely endemic to Australia. The genus has commercial value in the cut-flower and horticultural industries (Slater *et al.*, 2001; Dawson, 2010) but is highly valued for premium essential oil and honey products (e.g. Porter & Wilkins, 1998; Old, 2013; Carter *et al.*, 2016; Essien *et al.*, 2019). While *L. scoparium* is globally recognised as a source of bioactive honey, it is now known that many *Leptospermum* species produce DHA (dihydroxyacetone), a chemical compound that facilitates bioactive honey production (Williams *et al.*, 2018), and this places Australia in a unique position to expand the bioactive honey industry across the genus.

The genus *Leptospermum* current currently consists of 88 species, 85 of which are endemic to Australia, where they are found across a diverse range of environments including coastal, grassland, wetland and arid regions (Thompson, 1989). The diversification and wide distribution of *Leptospermum* across Australia, in conjunction with the increasing number of species being discovered to produce DHA, has opened wider opportunities for establishing bioactive honey production across the country. However, the genus is taxonomically challenging, with a long history of uncertainty in both species and generic boundaries. This limits a full exploration of the economic interests in this group because a sound taxonomic foundation is necessary to establish breeding programs, to provide consumer confidence in the source of products and to maximise the quality of product produced. Thus, the resolution of outstanding taxonomic issues in *Leptospermum* is critical for the growth of the Australian bioactive honey industry.

At the generic level, *Leptospermum* currently sits in the tribe Leptospermeae with nine other genera (Wilson *et al.*, 2005; Wilson, 2011). Over the last 150 years, the boundaries across these genera have been re-arranged through several major taxonomic revisions (Bentham, 1867; Niedenzu, 1898; Briggs & Johnson, 1979; Thompson, 1983, 1989; Wilson *et al.*, 2005; Wilson, 2011). Generic boundaries in this tribe have been particularly difficult to define using morphological approaches because while there is extensive morphological diversity and specialisation across the Leptospermeae, there are also many widely shared characters (Thompson, 1989). More recently, molecular data have provided great insight to phylogenetic relationships and have been able to resolve stable taxonomic genera across much of the Leptospermeae (O'Brien, Quinn, & Wilson, 2000; Wilson *et al.*, 2001, 2005; Lam *et al.*, 2002; De Lange *et al.*, 2010). However, early molecular work by O'Brien *et al.* (2000) revealed that *Leptospermum* is not monophyletic and therefore does not form a cohesive genus. Despite this knowledge, this major taxonomic issue has remained unresolved for the last 21 years. Resolution of this polyphyly requires extensive molecular sequencing across many species of the genus, as well as species of closely related genera, to be able to delimit appropriate generic boundaries. Such a study may result in *Leptospermum* being split into multiple new genera, the amalgamation of *Leptospermum* with other genera, or some combination of both.

In addition to broad taxonomic issues, there is also taxonomic confusion among select *Leptospermum* species that are of interest to the Australian bioactive honey industry. Within Western Australia, there is particular interest in the bioactive potential of *L. nitens* and *L. roei* (Williams *et al.*, 2018). In combination with their apparent tolerance for a wide range of climatic conditions across their relatively

large geographic distributions, these species may provide an opportunity to expand the bioactive honey industry in Western Australia using local species. However, these species are otherwise poorly known and there is much taxonomic confusion in field identification, such that they can be difficult to separate from morphologically similar species, such as *L. erubescens*, *L. oligandrum* and *L. fastigiatum*, that do not produce DHA in high concentrations or at all. Another species of interest, *L. polygalifolium*, is widely distributed along the east coast of Australia and is currently recognised as six subspecies (Thompson, 1989). There is taxonomic uncertainty in these subspecies, with evidence that some of them may warrant recognition as distinct species (Buys *et al.*, 2019) but also evidence of intergradation, raising issues of subspecific validity. Similarly, Australian mainland *L. scoparium* is known to form a challenging complex with *L. continentale* and *L. juniperinum* because the features of these species intergrade and are often difficult to separate. Natural hybridisation among species is known to be common throughout *Leptospermum* (Thompson, 1989) and is likely to be a major contributing factor in the difficulty of using morphology to identify species across these groups. The application of molecular data will allow the identification of hybrids and clarification of species boundaries, which may result in splitting species and/or amalgamating species, to provide a stable taxonomy on which to establish breeding programs for key species of interest.

In this report, we present the results of two molecular studies that aimed to resolve some of the taxonomic issues in *Leptospermum*. The first study used genome skimming to compile the chloroplast genome, as well as the nuclear ribosomal region, across *Leptospermum* and closely related genera to investigate the relationships among *Leptospermum* species and resolve generic boundaries. The second study employed reduced representation sequencing to generate SNP markers that have high resolution to identify hybrids and species boundaries in the *L. nitens* species complex. The findings of these molecular studies will inform opportunities and obstacles to expansion of the bioactive honey industry in Australia.

OBJECTIVES

This project had two primary objectives that are detailed within this report:

1. Understanding the phylogenetic relationships among *Leptospermum* species across Australia
2. Determine the genomic diversity of selected bioactive *Leptospermum* species from Western Australia

In addition, there were two secondary objectives that are not detailed in this report:

3. Determine the genomic diversity of selected bioactive *Leptospermum* species from the Gather By collection. This study was dependent on the provision of samples of *L. polygalifolium* by external parties. Due to logistical difficulties, these samples were not provided and the study, therefore, did not proceed.
4. Support Project 10 to attain a genetic diversity understanding of *L. scoparium* selections. Support was provided in the form of co-supervision of a PhD candidate based at the University of Adelaide; the results of this molecular study will be reported in the final report for Project 10.

KEY ACTIVITIES

1. Produce a phylogeny and determine genetic relationships of Australian *Leptospermum*.
2. Determine the genomic diversity of selected *Leptospermum* species from Western Australia.

IMPACTS

This project provides significant benefits for the Australian bioactive honey industry because a sound taxonomic understanding provides foundational knowledge that can highlight opportunities and identify obstacles for establishing and expanding bioactive honey production. This includes informing selections for breeding, improving productivity, improving optimisation of breeding, while also providing a basis for plant breeders rights, and a means of evaluation of risk assessment. The impact of these factors is difficult to evaluate in direct dollar terms, but all are fundamental components of an effective breeding and deployment program.

OUTPUTS

Two peer reviewed scientific papers, one published in *Taxon* and one submitted to the *Botanical Journal of the Linnean Society*. One conference presentation at the Australasian Honey Bee 2021 Research Conference.

Study 1: Phylogenetic relationships among *Leptospermum* species across Australia

BACKGROUND

Leptospermum is an ecologically and economically important genus with a long history of taxonomic uncertainty. The tribe Leptospermeae presently includes the genera *Agonis*, *Asteromyrtus*, *Homalospermum*, *Kunzea*, *Leptospermum*, *Neofabricia*, *Paragonis*, *Pericalymma* and *Taxandria* (Wilson *et al.*, 2005; Wilson, 2011). However, relationships among *Leptospermum* and these other genera have been re-arranged through several major taxonomic revisions over many years (Bentham, 1867; Niedenzu, 1898; Briggs & Johnson, 1979; Thompson, 1983, 1989; Wilson *et al.*, 2005; Wilson, 2011). More recently, molecular data have provided great insight to phylogenetic relationships across this group, however a major taxonomic issue remains outstanding: *Leptospermum* is not monophyletic and therefore does not represent a single, cohesive genus (O'Brien *et al.*, 2000). For such a large group of economically important species, a detailed investigation to resolve this problematic taxonomy is urgently needed.

The current circumscription of *Leptospermum* is largely based on an extensive morphological examination by Thompson (1989), describing 79 species. Across these species, Thompson recognised two morphological groups, each defined by the degree of persistence and woodiness of the fruit, as well as the shape and surface pattern of the seeds, with a further seven subgroups defined by differing suites of shared characters. Thompson (1989) retained the two morphological groups (hereafter referred to as morphological group one and two) as a single genus, while acknowledging that they may represent separate genera. Since that major revision, the genus has remained relatively stable, with some new species added and only minor revisions to existing species (Bean, 1992; Lyne, 1993; Lyne & Crisp, 1996; Bean, 2004). An early molecular study by O'Brien *et al.* (2000) identified the genus as polyphyletic, forming four genetic groups (hereafter referred to as genetic groups I-IV) that were more closely related to other genera than they were to each other: (I) an eastern Australian group with persistent fruit that aligned with morphological group two and was associated with the monotypic *Homalospermum*; (II) an eastern Australian group with non-persistent fruit (i.e. morphological group one) that was associated with *Neofabricia*; (III) a western Australian group, also in morphological group one with non-persistent fruit, that was associated with *Kunzea*; and (IV) a distinct Western Australian species, *L. spinescens*, that was also recognised by Thompson (1989) as an unusual species that did not neatly fit into either morphological group but was placed within morphological group one. While the results of O'Brien *et al.* (2000) clearly indicate that *Leptospermum* needs substantial revision into multiple, possibly four, genera, the study was based on limited molecular data and few species, and this prevented a clear interpretation. Despite this knowledge, the problematic taxonomy of *Leptospermum* has remained unresolved for the last 21 years. Comprehensive analysis with a larger volume of sequence data and greater species sampling is needed to fully assess the relationships among *Leptospermum* and closely allied genera. In this study, we apply genome skimming to obtain high copy chloroplast and nuclear ribosomal DNA for phylogenomic analysis to resolve the relationships among Australian *Leptospermum* species. A summary of the study is presented below; for a more detailed

description of the methodology and broader discussion points, please refer to the published paper for this study (Appendix 1).

METHODOLOGY

All the samples used in this study were sourced from herbarium specimens housed within the Western Australian Herbarium (PERTH) or the Australian National Herbarium (CANB). A total of 38 *Leptospermum* species were sampled that represent all currently described western Australian species and a representative subset of central and eastern Australian species. Additional specimens were sampled for 20 species across the following genera: *Agonis*, *Asteromyrtus*, *Calothamnus*, *Chamelaucium*, *Homalospermum*, *Kunzea*, *Lophostemon*, *Melaleuca*, *Neofabricia*, *Pericalymma* and *Thryptomene*. Where possible, each species was represented by two specimens. Total genomic DNA was extracted from the dried leaf material of these specimens using a modified CTAB protocol (Doyle & Doyle, 1987).

The final set of 111 DNA samples, as well as 16 technical replicates, were processed at the Australian Genome Research Facility (AGRF; Melbourne, Australia) for shotgun sequencing. A genome skimming approach was used to extract and assemble genomic regions in high copy: chloroplast DNA (hereafter, cpDNA) and nuclear ribosomal units (hereafter, nDNA). Libraries were prepared using the Illumina Truseq DNA-nano kit. Following quality control checks, samples were pooled and sequenced using Illumina sequencing platforms (HiSeq 2500 and Novaseq 6000, Illumina Inc.).

Read quality was assessed using FastQC software v0.11.6 and sequence data assembly was performed using Geneious v11.1 (Kearse *et al.*, 2012). To assemble the chloroplast and nuclear sequence data across the broad range of genera sampled, reads were mapped against the reference genomes available for *Eucalyptus grandis* (GenBank accessions NC_014570.1 and AF390472) and *Eucalyptus globulus* (GenBank accession KC180787.1, AY615680, KM064772). The gene identity of each annotated region was verified through BLAST, reading frames were checked by manual alignment in AliView v1.21 (Larsson, 2014) and the sequences for each genome were concatenated separately using FASCONCAT v1.11 (Kück & Meusemann, 2010).

Phylogenetic reconstruction on both alignments was performed using Maximum likelihood (ML) and Bayesian Inference (BI) methods. For ML analysis, the best-fit nucleotide substitution model for each partition was assessed using ModelFinder (Kalyaanamoorthy *et al.*, 2017) and IQ-TREE (Nguyen *et al.*, 2015) was run for ML tree inference using ultrafast bootstrap approximation (Hoang *et al.*, 2018). For the BI analysis, PartitionFinder 2 v2.1 (Lanfear *et al.*, 2017) was used to determine the best-fit substitution models and the analysis was run using MRBAYES v3.2.6 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). FigTree v1.4.2 was used to visualise the ML and BI consensus trees.

RESULTS

All technical replicates showed high similarity to their sampled counterpart. Of the 111 unique samples, one sample was omitted from analysis due to low read recovery. For the remaining 110 samples, the final chloroplast alignment reached a total length of 132, 143 bp, with an average depth of 148.8 ± 15.3x and consisted of 80 unique coding DNA sequences, four rRNA genes, 29 tRNA genes, 17 introns

and 97 intergenic spacers. The nuclear alignment (ETS, ITS1, ITS2, 5.8S) was recovered with a total length of 1,219 bp and average depth of 1289.7 ± 138.3x. These sequence data are available on GenBank for the chloroplast (Accessions MZ144837-MZ144946) and nuclear (Accessions MW809082-MW809191) alignments.

The ML and BI trees for the chloroplast alignment were well-supported, with 100% bootstrap support and 1.0 posterior probabilities at all deep nodes, as well as most shallow nodes. The two methods returned identical cpDNA tree topologies and therefore we only present the ML tree here. The focal group of Leptospermeae genera formed two clades (Clades A and B) in the cpDNA tree, which confirmed polyphyly in *Leptospermum*. *Leptospermum* species occurred in both clades, and within Clade B *Leptospermum* species were further split into four subclades, with most of the remaining Leptospermeae genera falling between clades and subclades (Figure 1). Clade A consisted of eastern Australian species of *Leptospermum* with persistent fruit, sister to the monotypic *Homalospermum*. This placement of *Leptospermum* species is consistent with both morphological group two and with genetic group I, while the remaining *Leptospermum* species in Clade B aligned with morphological group one and the remaining three genetic groups of O'Brien *et al*. (2000). In Clade B, *Asteromyrtus* and the monophyletic *Agonis* + *Pericalymma* clade were successive sisters to four subclades that included *Leptospermum* species: B1, composed of six western Australian species, including *L. spinescens*, which was singled out morphologically by Thompson (1989) and as genetic group IV; B2, comprising a suite of eastern Australian species sister to one western Australian species (*L. confertum*); B3, comprised of the remaining eastern Australian species as sister to *Neofabricia*, where subclades B2 and B3 provide enhanced resolution of genetic group II; and B4, which aligned with genetic group III and was composed of all remaining Western Australian species as sister to *Kunzea*. Across the tree, most paired specimens for each species were resolved in adjacent tip positions; however, this was not the case for some species from western Australian subclades B1 and B4, and these species were also associated with very short branch lengths relative to those of other clades.

The phylogenetic trees derived from the nuclear data resolved similar groupings to those derived from the chloroplast data, although there were some key differences in tree topology. Both the ML (Figure 2) and BI trees were well-supported (90-100% bs and 0.95-1.0 pp) at the deepest nodes, as well as most shallow nodes, but there was weak support (<80% bs and <0.75 pp) at the nodes separating the lineages of Leptospermeae genera. In the ML tree, this resulted in the Leptospermeae forming two clades, similar to those of the chloroplast tree, but with three key differences. The first is the placement of *Homalospermum* outside of Clades A and B, as sister to the Leptospermeae. Second is the placement of *Kunzea*, which moved from Clade B to Clade A, as sister to the persistent-fruited eastern Australian *Leptospermum* species. And thirdly, *L. subtenue* was separated from subclade B1 to form a distinct lineage. However, given the low support on these branches, these relationships should be regarded with caution. Indeed, the BI tree on the nuclear alignment replaced these weakly supported nodes with polytomies, putting some uncertainty on the relationship among these closely allied genera, as well as the placement of *L. subtenue*. Nevertheless, with the exception of *L. subtenue*, the membership of species within genera and *Leptospermum* subclades was consistent with that of the chloroplast tree. Also, similar to the chloroplast results, species represented by multiple accessions were typically resolved together except for those on very short branches in subclades B1 and B4, although more of these species were resolved in the nuclear tree.

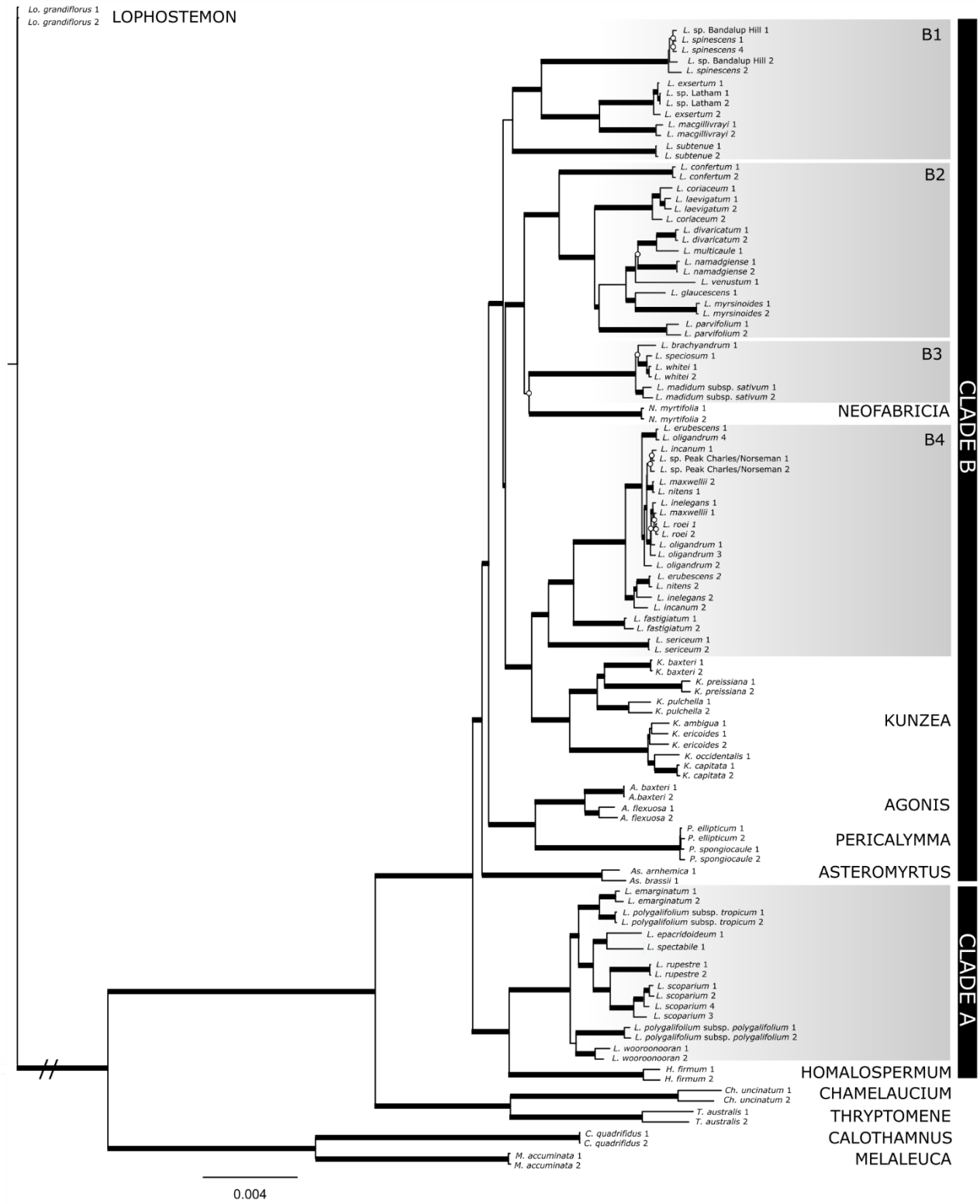

Figure 2: Phylogeny resulting from the cpDNA alignment. Consensus maximum likelihood tree for the chloroplast genome alignment (132, 143 bp) consisting of 110 accessions representing *Leptospermum* and closely allied genera of tribe *Leptospermeae*, as well as representatives of tribes Chamelaucieae and Melaleuceae, with *Lophostemon* as the outgroup. Bootstrap support values are indicated by weighted branches (100% support) or white circles (>90% support) at each node. Support values < 90% are omitted. Each discrete *Leptospermum* clade or subclade is highlighted in grey.

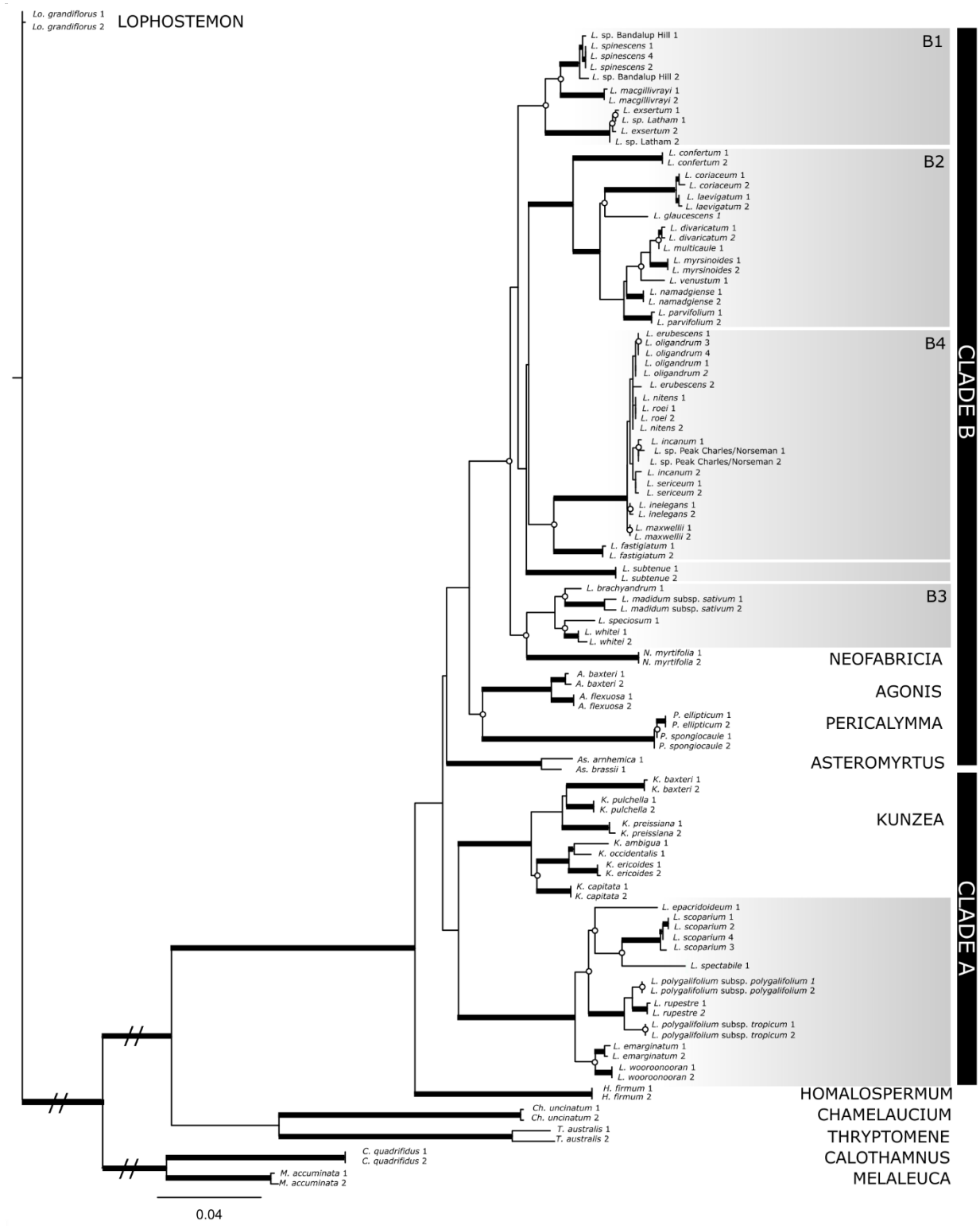

Figure 3: Phylogeny resulting from the nDNA alignment. Consensus maximum likelihood tree for the nuclear genome alignment (1,219 bp) consisting of 110 accessions representing *Leptospermum* and closely allied genera of tribe *Leptospermeae*, as well as representatives of tribes Chamelaucieae and Melaleuceae, with *Lophostemon* as the outgroup. Bootstrap support values are indicated by weighted branches (100% support) or white circles (>90% support) at each node. Support values <90% are omitted. The *Leptospermum* clades and subclades are indicated in grey.

SUMMARY

Our data confirm extensive polyphyly of *Leptospermum* as currently circumscribed and the need for major taxonomic revision. While there were some topological differences between the chloroplast and nuclear trees, and some branches with low support on the latter, the main groupings of species were congruent between the two datasets to provide a robust interpretation of the broad relationships present within the genus *Leptospermum*. Both trees resolved five monophyletic clades of *Leptospermum* species that were spread amongst clades that contained other genera of Leptospermeae. In resolving this extensive polyphyly, we provide evidence that supports division of *Leptospermum*, as currently recognised, into five genera. All species of morphological group two fall within Clade A and, given that the type for *Leptospermum*, *L. scoparium*, falls within this clade, these species should retain the name *Leptospermum*. Within Clade B, the nesting of *Neofabricia* and *Kunzea* among *Leptospermum* subclades presents two main options for taxonomic resolution: either the combination of all *Leptospermum* species of Clade B with *Neofabricia* and *Kunzea* (which would take the oldest name, *Kunzea*), or the distinction of the four *Leptospermum* subclades into separate genera. Given the wide morphological diversity across this group, incongruence in the placement of *Kunzea* in the chloroplast (Clade B) and nuclear (Clade A) trees, as well as the relative molecular divergence of other genera of Leptospermeae in our data generally, we recommend that the most appropriate treatment would be to recognise four new genera for each of the *Leptospermum* subclades in Clade B, while retaining *Kunzea* and *Neofabricia* as separate genera. This is consistent with previous discussion on the resolution of polyphyly in *Leptospermum* (O'Brien *et al.*, 2000), and our associated colleagues, Peter G. Wilson and Margaret Heslewood who have taxonomic expertise in *Leptospermum*, will publish a full taxonomic revision in line with these recommendations in the near future.

The strong morphological and geographic signals in our dataset provide a high level of confidence for assigning unsampled species to each major group. The two major clades, Clade A and B, separated *Leptospermum* species in alignment with the two morphological groupings recognised by Thompson (1989); the species in Clade A all belong to morphological group two and the species across the four subclades of Clade B are all members of morphological group one. Thus, the diagnostic characters of Clade A are evident and the unsampled members of this group can be confidently assigned based on the morphological attributes of morphological group two, as outlined by Thompson (1989). Within Clade B, although there were some consistent patterns, the four subclades do not align with the morphological subgroups proposed by Thompson (1989). However, they do largely align with the genetic groups identified by O'Brien *et al.* (2000) and show strong geographic signals, consisting of exclusively western or central/eastern Australian species, with only a single exception (Western Australian *L. confertum* in eastern subclade B2). Given that all western Australian species were sampled, the western subclades, B1 and B4, appear to be complete. However, membership of subclades B2 and B3 requires further consideration. Collectively, the members of these subclades were previously recognised as an eastern Australian genetic group with non-persistent fruit, albeit as a polytomy with low branch support (O'Brien *et al.*, 2000). The increased resolution provided by this study presents a more complex scenario with these species split into two monophyletic subclades. Thorough morphological assessment by the above-mentioned colleagues in their taxonomic revision will identify a suite of diagnostic characters for each subclade to facilitate species assignment.

Resolution of the phylogenetic relationships within *Leptospermum* has highlighted great insight, as well as gaps in knowledge, regarding species of current and future interest for the production of bioactive honey. Firstly, we note that species reported to produce high levels of DHA were present across four of the five distinctive clades identified in this study, for example, *L. scoparium* and *L. polygalifolium* in Clade A, *L. subtenue* in subclade B1, *L. whitei* and *L. speciosum* in subclade B3, and *L. nitens* and *L. incanum* in subclade B4 (Williams, 2018; Williams *et al.*, 2018). Nectar testing has not yet been undertaken for most of the species in subclade B2 and so DHA production in this group is largely unknown. This widespread occurrence demonstrates that DHA production is not unique to *L. scoparium* and its closest relatives but has evolved widely across the Leptospermeae. Therefore, it may be worth investigating other closely related genera, such as Kunzea or Agonis, for their potential to produce bioactive honey. Second, the placement of *L. scoparium* specimens in the phylogenetic trees indicated that New Zealand *L. scoparium* is more closely related to Tasmanian *L. scoparium* than it is to mainland Australian populations. This is consistent with previous studies that have hypothesised that *L. scoparium* dispersed to New Zealand from Tasmania and not from mainland Australia (Thornhill *et al.*, 2015; Buys *et al.*, 2019). Third, short branch lengths and poor tip resolution of the ten western Australian species in subclade B4 indicate recent and rapid speciation in this group, such that the chloroplast and nuclear regions sequenced in this study were not sufficient to delineate species. Study 2 continues the molecular exploration of this group using a population-genomic approach for higher resolution at the species level. Finally, the two sampled subspecies of *L. polygalifolium* in this study were not resolved as closest relatives in either the chloroplast or nuclear trees. This is also consistent with the findings of Buys *et al.* (2019) and suggests that the six subspecies of *L. polygalifolium* need taxonomic re-assessment.

Study 2: Genomic diversity of selected bioactive *Leptospermum* species from Western Australia

BACKGROUND

In addition to taxonomic confusion at the generic level, *Leptospermum* also has a long history of taxonomic confusion among closely related species (e.g. Bentham, 1867; Briggs & Johnson, 1979; Thompson, 1983; 1989; Wilson *et al.*, 2005). One of the putative new genera identified in Study 1 (i.e., subclade B4), consists of ten Western Australian species that exhibited very short branch lengths, low branch support and poor tip resolution in both the chloroplast and nuclear phylogenetic trees. These features are indicative of a recent, rapid radiation of this group, such that individual species are very closely related and consequently, are difficult to delineate. The morphological characters used to distinguish these species were described some decades ago (Thompson, 1989) and while they are readily recognised on select herbarium specimens, taxonomic confusion quickly becomes evident when attempting to identify these species in the field. The difficulty of morphological identification in this group is threefold: (1) there are few diagnostic characters between species pairs; (2) many of these characters are floral or fruiting and can only be observed at specific times of the year; (3) there is extensive variation within species such that continuous characters often intergrade between species, or the same plant can exhibit both binary states of some categorical characters. Hybridisation is a likely explanation for the latter given that there is much overlap in geographical range, species often co-occur at the same sites and flowering times overlap. There is also some inconsistency in the treatment of some characters between species, such that diagnostic characters in one species are considered non-diagnostic for other species. Thus, the available morphological and molecular data are not sufficient to resolve these species. And despite their wide distribution across the southwest of Western Australia, these species are otherwise poorly known. There has been recent scientific interest in members of this group (*L. nitens* and *L. roei*) for their potential in bioactive honey production (Williams *et al.*, 2018; Santos *et al.*, 2021) and a solid taxonomic foundation is needed to explore the commercial opportunities in this group.

The ten species in the study system are all perennial shrubs that are endemic to the southwest region of Western Australia, except for *L. fastigiatum*, which also occurs in South Australia. Some species, particularly *L. nitens* and *L. erubescens*, are widespread and common across a wide range of substrates and climates, while others are more habitat-specific, such as the granite-associated species, *L. incanum* and *L. sericeum*, or the arid-associated *L. fastigiatum*. Notably, *L. oligandrum* has a disjunct distribution between the northern and southern parts of the southwest. There is some variation in flowering time but all species flower through the Southern Hemisphere spring. The reproductive biology and mating systems of these species are not known but like most Myrtaceae, these species are expected to be generalist insect-pollinated, and the very small seed is likely to be gravity and wind dispersed (Beardsell

et al., 1993). The eastern Australian *L. scoparium* is known to be self-compatible (Burrell, 1965; Bennick, 2009) but self-compatibility is not known for the study species.

Morphologically, the most distinct species is *L. sericeum*, which is easily recognised by large, silver-grey leaves and large flowers. The remaining species exhibit small, green leaves and small flowers, and all appear superficially similar but can be distinguished by various combinations of specific characters, including growth habit, leaf length, pedicel length, hypanthium shape and hair arrangement, presence or absence of bracts and stem tubercles (Thompson, 1989). Note that there are no documented morphological characters separating *L. incanum* and *L.* sp. Peak Charles/Norseman and it is expected that they are synonymous.

In this study we use a population genomic approach to improve our understanding of genetic diversity and relationships in this clade of recently radiated *Leptospermum*, with particular focus on the bioactive species, *L. nitens* and *L. roei*. Extensive sampling across southwestern Australia in combination with genome-wide SNP markers were used to assess the genetic patterns of variation among the ten currently circumscribed species. A summary of the study is presented below; for a more detailed description of the methodology, results and broader discussion points, please refer to the published paper for this study (Appendix 2).

METHODOLOGY

Sampling was undertaken over three flowering seasons (2018-2020) to cover the geographic ranges of all ten species across southwestern Australia. For each species, leaves from up to eight, widely spaced plants were collected from each population for a total collection of 955 samples from 90 populations. Many of these collections included multiple species from the same location. To achieve a genetic baseline for each taxon, samples were largely collected from plants that could be confidently identified to species as they are currently circumscribed, and in some cases, this included different morphotypes. For example, *L. nitens* and *L. roei* differ by a single character, having appressed or spreading hairs on the hypanthium, and yet *L. inelegans* also occurs with both appressed and spreading hairs but without taxonomic distinction. We, therefore, collected both forms of *L. inelegans* to test the taxonomic value of this character. Some collections were challenging to identify to species and exhibited intermediate characters that were suspected to represent hybrid forms. Some of these were included to test their placement against the genetic background of the main species collections. Voucher specimens were collected from each population of each species and lodged with the Western Australian Herbarium.

Genomic DNA was extracted from freeze-dried leaf material using a modified 2% CTAB method (Doyle & Doyle, 1987). Samples were sent to Diversity Arrays Technology Pty Ltd (DArT, Canberra, Australia) for DArTseqT™ analysis as per Sansaloni *et al.* (2010). Briefly, library preparation involved DNA digestion using two methylation-sensitive restriction enzymes, *Pst*I and *Hpa*II, and fragments were ligated with uniquely barcoded adaptors. Following PCR and quantification, the samples were standardised and pooled for sequencing in a HiSeq 2500 (Illumina). Read assembly, initial quality control and SNP calling were undertaken by DArT using their proprietary DArTsoft14 software.

Preliminary exploration of the dataset revealed that many individual plants were genetically identical, which we refer to as 'clonality' from hereon. To eliminate bias from the overrepresentation of clonal genotypes, samples that were identified as clonemates were removed to leave a single representative per genotype. Preliminary analysis also found that *L. fastigiatum* was highly divergent from all other species and its inclusion limited the number of SNP loci that could adequately cover all ten species. Consequently, *L. fastigiatum* was removed from the study and the DArTsoft14 pipeline was repeated to call SNPs for the nine remaining species. Further filtering was then applied using the *dartR* package v1.8.3 (Gruber *et al.* 2018) in *R* v4.0.4 (R Development Core Team, 2021) to produce a high-quality dataset for analysis that consisted of 1,435 SNP loci with 8.95% missing data and 22.78x mean read depth.

Several individual-based analytical approaches were used to assess the genetic relationships among the nine study species: (1) Principal Coordinates Analysis (PCoA) was performed using the *adegenet* package v2.1.3 (Jombart & Ahmed, 2011) in *R*; (2) STRUCTURE v2.3.4 (Pritchard, *et al.*, 2000) was used to identify the most likely number of genetic clusters in the data; and (3) Splitstree v4.15.1 (Huson & Bryant, 2006) was used to generate a phylogenetic network. While the results of all three analyses formed strong species groupings, there were a number of outlier samples that indicated substantial admixture and were identified as hybrids. To clarify species boundaries without the noise brought by hybridisation, these hybrid individuals were removed and the PCoA, STRUCTURE and Splitstree analyses were repeated.

Hybridisation was assessed in more detail for select individuals using NewHybrids v1.1 (Anderson & Thompson, 2002). Individuals that were identified as intermediate, both morphologically and genetically, were run in NewHybrids to assess their assignment to various hybrid categories (i.e. F1, F2 or simple backcrosses).

RESULTS

All species except *L. maxwellii* exhibited a high level of apparent clonality within populations. While this was an unexpected result and the sampling was not undertaken to specifically assess clonality, there are some interesting findings to report. Across these species, where the average number of samples collected per population was 7.37 ± 0.147, the average number of unique genetic individuals per population was just 1.96 ± 0.154. In six cases involving *L. nitens*, *L. roei*, *L. erubescens*, *L. inelegans* and *L. sericeum*, identical genotypes were shared across populations that ranged between 4 km and 46 km apart. Interestingly, for *L. maxwellii*, clonality was only detected in the three populations that were identified as hybrids (*L. maxwellii* × *oligandrum*) and not in the five remaining populations that represented "pure" *L. maxwellii*.

Clustering analyses supported the presence of six to eight genetic entities. The PCoA identified six genetic clusters across the first three axes (Figure 4A-C). The first two axes (27.8% and 18.7% of the total variation, respectively) separated (1) *L. maxwellii,* (2) *L. sericeum* and (3) *L. incanum* + *L.* sp. Peak Charles/Norseman into distinct clusters. The remaining species clustered together on axes one and two but separated on the third axis (10.1%) into (4) *L. inelegans* (with both appressed and spreading hairs), (5) *L. nitens* + *roei* and (6) *L. erubescens* + *L. oligandrum*. This hierarchical structuring was also reflected in STRUCTURE, where *K* = 8 was identified as the most biologically meaningful structure

(Figure 5A). These eight clusters were consistent with clusters 1-5 of the PCoA with almost no admixture, such that *L.* sp. Peak Charles/Norseman was indistinguishable from *L. incanum*, and similarly, *L. nitens* from *L. roei*. However, STRUCTURE split *L. erubescens*, northern *L. oligandrum* and southern *L. oligandrum* into three separate but admixed clusters. Finally, the Splitstree network resolved six clearly divergent clusters (Figure 5B), consistent with the six PCoA clusters. Similar to the PCoA and STRUCTURE, there was no delineation between *L. incanum* and *L.* sp. Peak Charles individuals, nor between *L. nitens* and *L. roei*; however, the cluster for *L. erubescens* and *L. oligandrum*, while clearly stemming from the one main branch, showed clear divergence of northern *L. oligandrum*, and to a lesser extent, southern *L. oligandrum*, within this cluster.

A total of 47 individuals from 34 populations were identified as hybrids based on intermediate genetic placement in the PCoA, STRUCTURE and Splitstree analyses. These hybrid individuals exhibited various levels of admixture between genetic clusters, indicating complex hybrid structure beyond the first generation. This was supported by the NewHybrids analysis of select individuals that confirmed F1 and F2 hybrids of *L. maxwellii* × *oligandrum*, F2 hybrids of *L. inelegans* × *nitens*, and F1, F2 and uniparental backcrosses of *L. erubescens* × *nitens*.

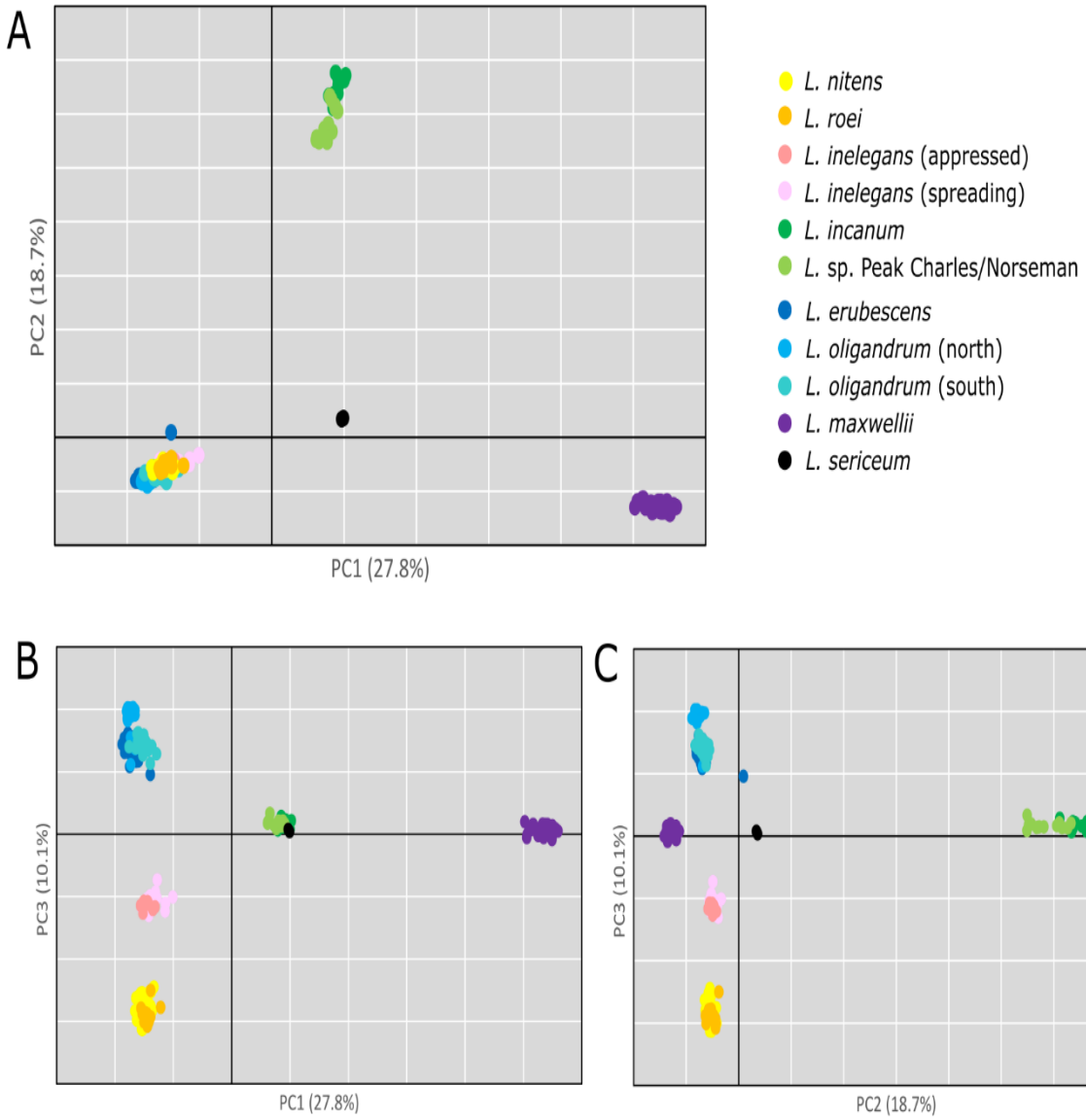

Figure 4: Principal coordinates analysis, where each point represents a unique genetic individual, coloured according to the nine study species, as well as relevant morphological (spreading and appressed forms of *Leptospermum inelegans*) and geographic (northern and southern populations of *L. oligandrum*) groups.

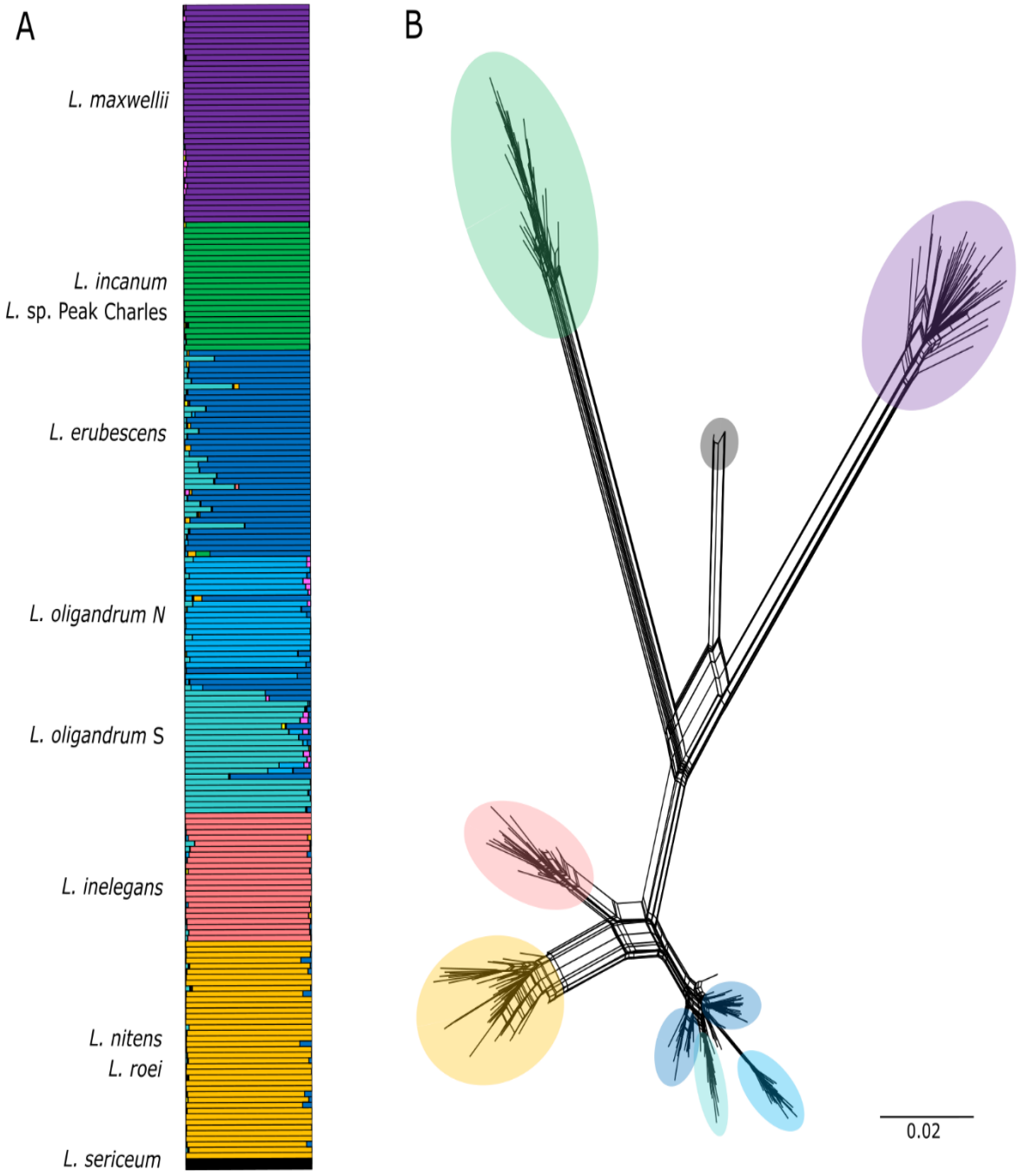

Figure 5. A) Bar chart of individual admixture proportions based on K=8 clusters from STRUCTURE analysis. B) Splitstree network, where coloured regions correspond to those of the eight STRUCTURE clusters.

SUMMARY

Our study provides a straightforward taxonomic interpretation for five of the genetic entities identified in this study system. Three entities correspond with current species as they are already circumscribed: *L. sericeum*, *L. maxwellii* and *L. inelegans*. In regard to morphological variation within these species, there was no indication of genomic divergence between prostrate and erect forms of *L. maxwellii*, nor between forms with spreading or appressed hypanthium hairs in *L. inelegans*, such that these characters do not appear to be taxonomically informative. There was no genomic divergence between *L. nitens* and *L. roei* and as these species only differ morphologically by hypanthium hair arrangement, we recommend synonymisation to the oldest name, *L. nitens*, and allowing this morphological character to be treated consistently as a non-informative character across the group. The fifth genetic entity, comprising *L. incanum* and *L.* sp. Peak Charles/Norseman, was also consistent with expectations that the latter phrase-named taxon was erroneously raised following a clerical error. With no record of morphological distinction and now a lack of genetic distinction between these taxa, we recommend that *L.* sp. Peak Charles/Norseman is sunk into *L. incanum*. These recommendations would reduce the seven current species to five species.

Interpreting the sixth genetic entity comprising *L. erubescens*, northern *L. oligandrum* and southern *L. oligandrum* is more complex. From the genetic results, is clear that *L. erubescens* and *L. oligandrum* are very closely related and it could be argued that they represent a single, geographically structured and morphologically variable species. However, there is evidence for molecular, morphological, phenological and reproductive divergence that may warrant the recognition of multiple species. If they do not represent one species, the PCoA and hierarchical STRUCTURE results clearly indicate that southern *L. oligandrum* is more closely related to *L. erubescens* than it is to northern *L. oligandrum*. Thus, regardless of the interpretation, these species should not remain as they are currently circumscribed. Here, we suggest that there are three alternative taxonomic pathways: (1) a one species solution that sinks *L. oligandrum* into *L. erubescens*, the older name of the two; (2) a two species solution that sinks southern *L. oligandrum* into *L. erubescens*, leaving northern *L. oligandrum* as a distinct species; or (3) a three species solution that splits *L. oligandrum* into a northern and southern species, as distinct from *L. erubescens*. Given the wide morphological variation seen within *L. erubescens*, as well as the absence of a morphological assessment between northern and southern *L. oligandrum*, we recommend that further morphological examination is needed to guide the most appropriate taxonomic pathway forward for these species.

As expected for a rapid radiation of species, we found hybridisation to be widespread across the study group. Despite the sampling intention of this study to primarily collect unambiguous morphological representatives of each species and only a select number of morphologically intermediate samples to test for hybridisation, a much wider range of samples were identified as hybrids based on conservative genetic criteria. Indeed, the morphologically intermediate samples were all confirmed as recent hybrids (F1, F2 and simple backcrosses) but the presence of hybrids among the morphologically representative samples suggests that more complex backcrossing and introgression is also occurring, and this will continue to make field identifications difficult.

The detection of apparent clonality across most of the study group was unanticipated and this had several consequences for our analysis and interpretation of population dynamics. Firstly, it reduced

sample sizes such that populations were often represented by a single genotype and typical population-level parameters of genetic diversity, inbreeding and differentiation could not be estimated within species. Secondly, our sampling was not designed for a dedicated spatial investigation of clonal structure within populations. Nevertheless, our sampling was widely spaced to cover the spatial spread of the plants at each site, which demonstrates that these clonal genotypes are spatially extensive within populations (up to at least 100 m). Given the evidence for extensive hybridisation, it is also clear that both sexual and asexual reproductive strategies are operating in this study system. As such, the wide arrangement of clonemates, combined with massive floral displays, indicates that, if self-compatible, inbreeding *via* selfing may be prevalent in these species. While genotypic diversity appears to be very low within populations, sampling just eight plants at each site may have missed rare genotypes or more complex arrangements of spatially distributed genotypes (Arnaud-Haond *et al.*, 2007) so a dedicated study is needed to assess the spatial extent of clonality within these populations.

Whilst knowledge of clonality in the study system is currently limited, the revelation of clonality itself is a major finding. Asexual reproduction is not uncommon in the Myrtaceae family, with vegetative growth documented in a number of genera, including *Eucalyptus* (e.g. Walker *et al.*, 2009; Bradbury *et al.*, 2021) and *Melaleuca* (e.g. Chong *et al.*, 2013; Hewitt *et al.*, 2019), and apomixis so far reported in *Callistemon* (Rye, 1979), *Psidium* (Souza-Pérez & Speroni, 2017) and *Syzygium* (Thurlby *et al.*, 2012). However, to our knowledge, neither of these reproductive strategies have been reported in *Leptospermum*. We hypothesize that the apparent clonality detected in the study group is indicative of apomixis, largely because the shared clonal genotypes between geographically distant populations could only be explained by apomictic seed dispersal. There is also extensive hybridisation occurring across the group, which is recognised as a major trigger for apomixis (Hojsgaard & Hörandl, 2019), and finally, there is documentation of polyploidy in *L. roei* (Rye, 1979), which is often associated with hybridisation and apomixis (Hojsgaard & Hörandl, 2019). Given that clonality was not detected in the most divergent species in the group, *L. maxwellii*, it seems probable that this propagation strategy is unique to this specific group and may not occur more widely across other *Leptospermum* species.

Implications and recommendations

A solid taxonomic foundation in *Leptospermum* is vital for successful expansion of the bioactive honey industry in Australia. The two molecular studies reported here resolve some of the taxonomic issues in the genus and provide a wealth of information to support already existing and newly establishing breeding programs and plantations of *Leptospermum*.

The resolution of polyphyletic relationships for Study 1 requires the genus to be split into five genera. This study provides answers to a long-standing taxonomic problem that has been known for 21 years and finally resolved, now provides clarity for consumer confidence in the source of any commercial products produced from these species. More specifically, this result has implications for the bioactive honey industry, the most obvious being that many species will lose the *Leptospermum* name. *Leptospermum scoparium*, the primary species from which bioactive honey is currently produced, is the type for the genus, which means that it and all species with persistent fruit as per Thompson (1989), including *L. polygalifolium*, will retain the name *Leptospermum*. A very interesting finding was that bioactive species are not limited to this group, with a number of species of interest occurring across at least three of the four remaining groups. For example, *L. nitens*, *L. whitei*, *L. subtenue* and *L. speciosum*, all produce DHA in their nectar despite occurring across different genera. This provides a branding opportunity to establish a broader market for exclusively Australian bioactive honey. Indeed, the revelation of DHA production outside of what we now know as true *Leptospermum* also presents the possibility that species from other, currently untested genera may also yield valuable species for bioactive honey exploration.

The results of Study 1 also provide guidance to breeding trials of *Leptospermum*. Breeding trials often use hybridisation as a mechanism to cultivate improved varieties that combine the desirable traits of multiple species. Being different genera, it is likely that the species from each of the five groups identified in this study will have evolved reproductive incompatibilities, such that any hybridisation attempts between groups are less likely to be successful, either at F1 or later generations. Thus, intentional hybridisation in breeding trials should avoid crossing species from the different groups and instead focus on crossing species within each of the groups to maximise the likelihood of success. In addition, this study also highlighted taxonomic issues regarding *L. polygalifolium*, which has an established bioactive honey operation in eastern Australia. While our study only included two of the six subspecies, the results revealed that they are not each other's closest relative. This indicates that they should not be considered subspecies of a single species, and likely represent distinct species. This has implications for existing and establishing operators if these subspecies are used interchangeably in breeding programs or plantations when they may be entirely different species and should be utilised differently. Further molecular work with a focus on these subspecies, using methods similar to those of the study for Study 2, is needed to resolve the taxonomic status of *L. polygalifolium*.

The findings of Study 2 provide clarity regarding species boundaries within the Western Australian subclade B4, identified in Study 1. The primary reason for this study was interest in exploring *L. nitens* and *L. roei* for bioactive honey production in Western Australia and therefore, the revelation that they are the same species is an important finding for the industry to be aware of. This is consistent with the fact that both produce DHA in similarly high concentrations and their amalgamation provides a wider

range of genotypes in which to make selections for breeding trials. The resolution of species boundaries across the remaining species in this group also provides clarity in the morphological characters that are useful for distinguishing species, and this will aid in correct field identifications for seed collections or nectar testing. Unfortunately, the confirmation of extensive natural hybridisation means that some field identifications will still be difficult and therefore it is recommended that care is taken to not utilise seed, leaf or nectar that cannot be confidently identified to a given species, particularly if multiple species occur at the site. It is also unknown how hybridisation between species with contrasting DHA productivity, for example *L. erubescens* × *nitens*, impacts DHA production and so the detection of extensive hybridisation where these species co-occur has implications for the choice of locations to place commercial hives or plantations to avoid any negative impacts of natural hybridisation on DHA production.

An unexpected finding in Study 2 was that of extensive clonality within populations of most species, including *L. nitens* and *L. roei*. This also has major implications for seed collections in that seeds harvested from a single population are highly likely to be genetically identical. Moreover, that identical genotypes were found to occur among populations indicates that genotypic diversity also cannot be guaranteed with collections from geographically proximate sites. This is critical information to be aware of when searching for diverse genotypes to select for breeding trials. However, that these species occur naturally in dense, often apparently monoclonal stands and that clonality may be the result of apomixis, will be of downstream benefit to interested growers because highly valued genotypes may be relatively easy to maintain and propagate in cultivation.

Future research opportunities

While the molecular studies reported here have provided much needed clarity and insight regarding *Leptospermum* to support the growing bioactive honey industry in Australia, gaps in knowledge remain. Thus, to further inform and refine knowledge for industry benefits, we suggest ongoing research is needed in the following areas:

- Continue testing DHA in other species and genera to broaden the pool of flora for bioactive honey exploration and marketing opportunities.
- Resolve other taxonomic issues in key bioactive species, notably the six *L. polygalifolium* subspecies and the *L. scoparium/continentale/juniperinum* complex.
- Assess how DHA production in *L. nitens* is impacted by hybridisation with low DHA producing species, such as *L. erubescens.*
- Spatially intensive surveys of clonality in *L. nitens* populations to thoroughly document the mating system, polyploidy and the spatial extent of clonality in this species.
- Seed progeny arrays in *L. nitens* to confirm that clonality is the result of apomixis.

REFERENCES

Anderson EC & Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217–1229.

Arnaud-Haond S, Duarte CM, Alberto F & Serrão EA. 2007. Standardizing methods to address clonality in population studies. *Molecular Ecology* 16: 5115–5139.

Bean AR. 1992. The genus *Leptospermum* Forst. et Forst. F. (Myrtaceae) in northern Australia and Malesia. *Austrobaileya* 3: 643–659.

Bean AR. 2004. Three new species of *Leptospermum* (Myrtaceae) from Queensland and northern New South Wales. *Telopea* 10: 831–838.

Beardsell D V, O'Brien SP, Williams EG, Knox RB & Calder DM. 1993. Reproductive biology of Australian Myrtaceae. *Australian Journal of Botany* 41: 511–526.

Bennick RM. 2009. The effects of honeybeees (*Apis mellifera*) on the biodiversity of manuka (*Leptospermum scoparium*) patches. *Master of Science*, Massey University, New Zealand.

Bentham G. 1867. *Flora Australiensis: A description of the plants of the Australian territory*. London: Lovell Reeve & Co.

Bradbury D, Binks RM & Byrne M. 2021. Genomic data inform conservation of rare tree species: clonality, diversity and hybridity in *Eucalyptus* series in a global biodiversity hotspot. *Biodiversity and Conservation* 30: 619–641.

Briggs BG & Johnson LAS. 1979. Evolution in the Myrtaceae - evidence from inflorescence structure. *Proceedings of the Linnean Society of New South Wales* 102: 157–256.

Burrell J. 1965. Ecology of Leptospermum in Otago. *New Zealand Journal of Botany* 3: 3–16.

Buys MH, Winkworth RC, De Lange PJ, Wilson PG, Mitchell N, Lemmon AR, Moriarty Lemmon E, Holland S, Cherry JR & Klápště J. 2019. The phylogenomics of diversification on an island: Applying anchored hybrid enrichment to New Zealand *Leptospermum scoparium* (Myrtaceae). *Botanical Journal of the Linnean Society* 191: 1–17.

Carter DA, Blair SE, Cokcetin NN, Bouzo D, Brooks P, Schothauer R & Harry EJ. 2016. Therapeutic manuka honey: No longer so alternative. *Frontiers in Microbiology* 7: 1–11.

Chong C, Edwards W, Pearson RG & Waycott M. 2013. Sprouting and genetic structure vary with flood disturbance in the tropical riverine paperbark tree, *Melaleuca leucadendra* (Myrtaceae). *American Journal of Botany* 100: 2250–2260.

Dawson M. 2010. A history of *Leptospermum scoparium* in cultivation: Garden selections. *New Zealand Garden Journal* 13: 2–9.

Doyle JJ & Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Essien SO, Baroutian S, Dell K & Young B. 2019. Value-added potential of New Zealand mānuka and kānuka products: A review. *Industrial Crops and Products* 130: 198–207.

Hewitt A, Rymer P, Holford P, Morris EC & Renshaw A. 2019. Evidence for clonality, breeding system, genetic diversity and genetic structure in large and small populations of *Melaleuca deanei* (Myrtaceae). *Australian Journal of Botany* 67: 36–45.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ & Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Hojsgaard D & Hörandl E. 2019. The rise of apomixis in natural plant populations. *Frontiers in Plant Science* 10: 358.

Huelsenbeck JP & Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Huson DH & Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Jombart T & Ahmed I. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A & Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

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, Meintjes P & Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kück P & Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution* 56: 1115–1118.

Lam N, Wilson PG, Heslewood MM & Quinn CJ. 2002. A phylogenetic analysis of the *Chamelaucium* alliance (Myrtaceae). *Australian Systematic Botany* 15: 535–543.

Lanfear R, Frandsen PB, Wright AM, Senfeld T & Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

De Lange PJ, Smissen RD, Wagstaff SJ, Keeling DJ, Murray BG & Toelken HR. 2010. A molecular phylogeny and infrageneric classification for *Kunzea* (Myrtaceae) inferred from rDNA ITS and ETS sequences. *Australian Systematic Botany* 23: 309–319.

Larsson A. 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30: 3276–3278.

Lyne AM. 1993. *Leptospermum namadgiensis* (Myrtaceae), a new species from the Australian Capital Territory-New South Wales border area. *Telopea* 5: 319–324.

Lyne AM & Crisp MD. 1996. *Leptospermum jingera* (Myrtaceae-Leptospermoideae): A new species from north-eastern Victoria. *Australian Systematic Botany* 9: 301–306.

Nguyen LT, Schmidt HA, von Haeseler A & Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Niedenzu F. 1898. Myrtaceae. In: *Die natürlichen Pflanzenfamilien, Volume 3*. Leipzig: Engelmann, 57–105.

O'Brien MM, Quinn CJ & Wilson PG. 2000. Molecular systematics of the *Leptospermum* suballiance (Myrtaceae). *Australian Journal of Botany* 48: 621–628.

Old N. 2013. The medicine of the manuka: an investigation of the usages and methods for utilization of honey derived from the pollen of *Leptospermum scoparium* in holistic nursing practice. *Journal of Holistic Nursing* 31: 200–203.

Porter NG & Wilkins AL. 1998. Chemical, physical and antimicrobial properties of essential oils of *Leptospermum scoparium* and *Kunzea ericoides*. *Phytochemistry* 50: 407–415.

Pritchard JK, Stephens M & Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

R Development Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ronquist F & Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Rye BL. 1979. Chromosome number variation in the Myrtaceae and its taxonomic implications. *Australian Journal of Botany* 27: 547–573.

Sansaloni C, Petroli C, Carling J, Hudson C, Steane D, Myburg A, Grattapaglia D, Vaillancourt R & Kilian A. 2010. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* 6: 16.

Santos AA, Hancox TJJ, Picanço MC, Delaporte K & Hogendoorn K. 2021. Potential distribution of *Leptospermum* species (Myrtaceae) in Australia for bioactive honey production purposes. *New Zealand Journal of Crop and Horticultural Science*.

Slater A, Blakemore MC, Faragher JD, Franz PR, Henderson B & Green K. 2001. *Leptospermum as an export cut flower crop*. Report for the Rural Industries Research and Development Corporation, Australian Capital Territory.

Souza-Pérez M & Speroni G. 2017. New apomictic pathway in Myrtaceae inferred from *Psidium cattleyanum* female gametophyte ontogeny. *Flora* 234: 34–40.

Thompson J. 1983. Redefinitions and nomenclatural changes within the *Leptospermum* suballiance of Myrtaceae. *Telopea* 2: 379–383.

Thompson J. 1989. A revision of the genus *Leptospermum* (Myrtaceae). *Telopea* 3: 301–449.

Thornhill AH, Ho SYW, Külheim C & Crisp MD. 2015. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Molecular Phylogenetics and Evolution* 93: 29–43.

Thurlby KAG, Wilson PG, Sherwin WB, Connelly C & Rossetto M. 2012. Reproductive bet-hedging in a rare yet widespread rainforest tree, *Syzygium paniculatum* (Myrtaceae). *Austral Ecology* 37: 936–944.

Walker EA, Byrne MA, Macdonald BB, Nicolle DC & Mccomb JA. 2009. Clonality and hybrid origin of the rare *Eucalyptus bennettiae* (Myrtaceae) in Western Australia. *Australian Journal of Botany* 57: 180–188.

Williams SD, Pappalardo L, Bishop J & Brooks PR. 2018. Dihydroxyacetone production in the nectar of Australian *Leptospermum* is species dependent. *Journal of Agricultural and Food Chemistry* 66: 11133–11140.

Williams S. 2018. *A Beekeeper's Guide to Australian Leptospermum Trees and Honey* (P Brooks, Ed.). Sunshine Coast, Australia: University of the Sunshine Coast.

Wilson PG, O'Brien MM, Gadek PA & Quinn CJ. 2001. Myrtaceae revisited: A reassessment of infrafamilial groups. *American Journal of Botany* 88: 2013–2025.

Wilson PG, O'Brien MM, Heslewood MM & Quinn CJ. 2005. Relationships within Myrtaceae *sensu lato* based on a matK phylogeny. *Plant Systematics and Evolution* 251: 3–19.

Wilson PG. 2011. Myrtaceae. In: Kubitzki K, ed. *The Families and Genera of Vascular Plants. Volume X. Sapindales, Cucurbitales, Myrtaceae*. Heidelberg: Springer-Verlag, 212–271.

Australian Government
Department of Industry, Science, Energy and Resources

AusIndustry
Cooperative Research Centres Program