http://researchcommons.waikato.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Elucidating the molecular basis of dihydroxyacetone (DHA) production in *Leptospermum* nectar

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Ella Ruby Purangi Grierson

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Leptospermum scoparium, Nancy Adams 1949. Accessed from Te Papa collections.

Abstract

Mānuka (*Leptospermum scoparium*) nectar contains variable amounts of dihydroxyacetone (DHA), a triose sugar that is the precursor to methylglyoxal (MGO), the antimicrobial compound underpinning the high value of mānuka honey and forming the basis of the Aotearoa-New Zealand honey industry. This trait is unique to some species within the Myrtaceae. The molecular basis of this high-value trait is unknown and identifying it would allow development of gene-based tools to identify high value germplasm to inform replanting and breeding programmes. Despite the fundamental importance of nectar to crop pollination and honey industries, the genetic control of nectary function is poorly understood, especially in non-model species. This thesis aims to elucidate the molecular basis of DHA production in *Leptospermum* nectar, through identifying associated genes and genomic regions, followed by further exploration of gene regulation and enzyme characterisation *in vitro* and *in vivo*.

RNAseq analysis identified nectary-associated genes differentially expressed between high and low nectar-DHA genotypes of *L. scoparium*, and a mānuka high-density linkage map and quantitative trait loci (QTL) mapping population revealed genetic regions associated with nectar DHA content. Expression and QTL analyses both pointed to the involvement of a phosphatase gene, *LsSgpp2*, as its expression correlated with nectar DHA accumulation and it co-located with a QTL on chromosome 4. To investigate *Sgpp2* and its complex locus further, we produced high-quality genomes and hypanthium transcriptomes for two further Leptospermeae species with contrasting DHA phenotypes – *Leptospermum morrisonii* and *Gaudium laevigatum*. Expression patterns of *Sgpp2* again correlated with nectar DHA in these species, and comparative analyses of *Sgpp* sequences within the Myrtaceae indicated that high *Sgpp2* expression was likely ancestral in DHA producing taxa, followed by repeated loss of the trait. Comparison within and between *Sgpp* genes with differing expression profiles identified regions unique to the highly expressed *Sgpp2* genes, which contained two C-box motifs, and a bZIP11 transcription factor predicted to bind to these motifs was identified as significantly differentially expressed in the RNAseq dataset. The bZIP11 gene was subsequently found to co-locate with another of the QTLs identified, further supporting its involvement. The *L. morrisonii* promoter drove strong nectary-specific expression in transgenic lines of *Petunia* and *Nicotiana*, showing elements essential for *Sgpp2* expression are within that region. However, the promoter in 13 DHA producing *L. scoparium* genotypes was found to be similar, indicating variation in nectar DHA amounts may be conferred by further complex transcriptional regulation fine tuning *Sgpp2* expression from beyond the promoter region analysed here. Functional analyses demonstrated that LmSGPP2 can dephosphorylate dihydroxyacetone-phosphate (DHAP) to produce DHA and phosphate *in vitro*. Transgenic *Petunia* and *Nicotiana* lines were created to

characterise LmSGPP2 activity *in vivo*, but results were inconclusive due to low and mis-located expression of the transgene.

Together our results suggest *Sgpp2* may contribute to maintaining phosphate homeostasis in a photosynthesising nectary, potentially evolving due to low phosphate availability – which is common in Australia where this tribe originates. This work advances our understanding of nectar biology and reports some of the first QTLs and genes linked to a low abundance nectar compound. The genes and genomic regions identified here provide a foundation for developing tools to identify high-value germplasm – ensuring biodiversity can be maintained while increasing high-value honey production.

Keywords: dihydroxyacetone, *Leptospermum, Leptospermum scoparium, Leptospermum morrisonii, Gaudium laevigatum,* Myrtaceae, Leptospermeae, mānuka, nectar, phosphor-sugar phosphatase, QTL, *Sgpp.*

Acknowledgements

I sincerely thank my supervisors Mike Clearwater, Kevin Davies, Kathy Schwinn and David Chagné. I am so grateful to have learnt so much from working with you all, for your trust in me to lead this research, and for your help developing me as a scientist. Thank you Mike, for your valuable insights and support; Kathy, for leading our mānuka work forming the foundation for this project, and for being so supportive of me and my career; Kevin for your writing and editing ability and being such a trusted source of advice, and David, thank you for your infectious enthusiasm and for co-leading this project.

I am grateful for our MBIE Smart Ideas funding, and the partnership of Ngati Porou Miere Rangahau. Thank you to Plant and Food Research for your continued support and employment, this PhD would not have been possible without it.

Thank you to the Metabolite Traits in Plants team. Specifically, David Lewis for guidance and nectar sampling; Chethi Waite and Sarah Moss for your friendship and help nectar sampling; Nick Albert for being such a great sounding board and source of advice, and Rebecca Yorker for your summer working on the Smart Idea – it was so appreciated. Thank you to Ian King and Julie Ryan for taking care of all the plants, and Steve Arathoon for nectar sampling and enthusiasm for all things.

I am also grateful to Amali Thrimawithana for helping me navigate bioinformatics; Sarah Bailey for assistance with sequencing analysis, and Duncan Hedderley and Andrew McLachlan for help with statistics. Thank you to Robert Simpson for your patience and help with enzyme assays; Cyril Hamiaux for answering so many questions on protein purification, as well as John van Klink and Catherine Sansom for nectar analysis. Thank you to Ed Morgan and Zac Hanley for your support.

Thank you to my friends and family for providing so much fun outside of work. Thank you Laurie and Sebastien for so many delicious feasts and great walks. Thank you to my parents, for continuing to be a wonderful and continued source of support and fruit. Thank you to Rio, for sometimes letting me sleep, and being such a delightful distraction – ensuring I have work life balance.

The biggest thank you to my wife Nayer, for your endless support in every aspect of my life. Joining me on sampling trips, so many amazing adventures, and now being such a wonderful parent. You have made it possible for me to juggle parenthood and a PhD, and this thesis would absolutely not have been possible without you.

Table of Contents

List of Figures

List of Tables

List of Supplementary Figures

List of Supplementary Tables

Chapter 1: Introduction

Mānuka nectar contains variable amounts of dihydroxyacetone (DHA), the precursor to the antimicrobial compound underpinning the high-value of the manuka honey industry in Aotearoa-New Zealand. Mānuka honey sells for a significant premium over other honeys Aotearoa-New Zealand (MPI, 2023), and export prices are strongly linked to methylglyoxal (MGO) content and therefore nectar DHA. Increasing nectar DHA could create additional value for the Aotearoa-New Zealand honey industry, however the basis of this valuable and variable trait of nectar DHA in mānuka is currently unknown. Recent work has proposed that the origin of nectar DHA is through dihydroxyacetone-phosphate (DHAP), but exactly how is unknown. Understanding the molecular basis of this trait would allow development of gene-based tools to identify high value germplasm and inform replanting and breeding programmes Partnering with Māori is an important consideration for work in this area (Morgan *et al.*, 2019), as mānuka is a taonga species and Māori are key stakeholders in this industry. The present research was developed with our partners Ngati Porou Miere Rangahau, and we established that there was a need to understand the variation in honey value between their blocks, and to be able to produce a consistently high value product while still eco-sourcing plant lines from their rohe (territory or boundary of an iwi/tribe), respecting whakapapa and regional provenance. Despite the fundamental importance of nectar, the genetic control of nectary function is poorly understood. This thesis aims to elucidate the molecular basis of the relatively unique nectar trait of dihydroxyacetone (DHA) production in *Leptospermum*. Here we review literature surrounding nectar and nectaries, as well as mānuka and nectar DHA.

1.1 Nectar and nectaries

Nectar is a sugar rich liquid produced by glands called nectaries, and it has essential ecological and evolutionary functions (Carter *et al.*, 2019; Nicolson, 2022). Floral nectar is the key reward mediating the interface between plants and their pollinators (Liao *et al.*, 2025). Extrafloral nectar can also be provided as a reward to predatory insects for physical defence of the plant (Roy *et al.*, 2017). Nectaries have evolved multiple times, resulting in diversity in their structure (Bernardello, 2007). Nectar traits including volume and composition are variable, which impacts plant-animal interactions (Roy *et al.*, 2017). Approximately 70% of all plant species and 75% of domesticated crop species benefit from nectar mediated pollination (Carter *et al.*, 2019).

The main solutes in nectar are sucrose, glucose and fructose. The ratio of sucrose to hexose (glucose and fructose) is often consistent within a species, but can vary (Roy *et al.*, 2017; Parachnowitsch *et al.*, 2019). Nectar also contains a range of other components, including water, other sugars, amino

acids, proteins, lipids, vitamins and specialised metabolites (Pacini & Nepi, 2007; Roy *et al.*, 2017). The biological functions of these non-sugar components are not fully understood, but can contribute to flavour (e.g. Gardener & Gillman, 2002) or scent (e.g. Raguso, 2004), can attract pollinators or deter nectar robbers (Heil, 2011; Roy *et al.*, 2017), provide nutrition to pollinators (e.g. Kram *et al.*, 2008), and prevent microbial growth (Carter & Thornburg, 2004; Roy *et al.*, 2017). Nectar can also contain microbial communities, which can themselves influence nectar composition (Roy *et al.*, 2017; Vannette & Fukami, 2018; Quevedo-Caraballo *et al.*, 2025).

Despite the fundamental importance of nectar, nectary research has only recently emerged as a significant field. Studies on nectar secretion have often been conducted in *Arabidopsis* (Kram & Carter, 2009), *Nicotiana* (Ren *et al.*, 2007; Silva *et al.*, 2018) and more recently squash (Solhaug *et al.*, 2019). This research has resulted in multiple models of nectar synthesis and secretion, the most prevalent of which are the eccrine and merocrine based models (Chatt *et al.*, 2021). These models involve sugars accumulating as starch in the nectary tissue which provides precursors for nectar synthesis. The eccrine model then involves sucrose export facilitated by transmembrane transporters, and the merocrine model proposes that sucrose and other metabolites are packaged into vesicles and fuse with the plasma membrane to be released (Roy *et al.*, 2017; Chatt *et al.*, 2021).

Many nectaries do not have photosynthetic capacity, so nectar carbohydrates are derived from photosynthesis elsewhere in the plant and delivered through the phloem (Pacini & Nepi, 2007). These sugars can be directly transported into nectar (Ren *et al.*, 2007), but they are often stored as starch which is later degraded (Lin *et al.*, 2014; Solhaug *et al.*, 2019). In many plants, nectaries are green and contain plastids (Lüttge, 2013) and it had been hypothesised that nectar sugars can also be derived from photosynthesis within the nectary (Pacini & Nepi, 2007; Lüttge, 2013). It has recently been demonstrated in mānuka, which have long lived flowers with green nectaries, that photosynthesis can directly contribute to nectar carbohydrates (Clearwater *et al.*, 2021).

Nectar studies can be challenging, with many environmental and physiological factors to consider and minimise (Mitchell, 2004; Pacini & Nepi, 2007). Nectar traits can change throughout the day, during the flowers life span, and across the plants lifetime (Mitchell, 2004; Clearwater *et al.*, 2021). Nectar traits are very responsive to environmental variation, so many studies looking at genotypic differences in nectar traits have used controlled environments – as even substantial genetic variation can be hard to measure in the field (Mitchell, 2004). Galetto and Bernardello (2005) proposed some recommendations to minimise variability in assessing nectar, including sampling from defined locations across the plant, sampling flowers of the same age, and sampling at the same time of day.

1.2 Molecular genetics of nectaries

The genetic controls of nectary form and function are variable due to the multiple evolutionary origins of nectaries (Roy *et al.*, 2017), but surprisingly often share genes across plant lineages (Liao *et al.*, 2025). Most work has been done in the model species *Arabidopsis*, *Nicotiana*, cotton and squash (Ren *et al.*, 2007; Kram *et al.*, 2009; Solhaug *et al.*, 2019; Chatt *et al.*, 2021), as well as some others (e.g Carey *et al.*, 2023; Katzer *et al.*, 2025). Within angiosperms, the relative order of floral organs (sepals, petals, stamens and carpels) is conserved, and specified by the overlapping actions of various transcription factors – referred to as the ABC(E) genes (Slavković *et al.*, 2021). The location of nectaries, however, is variable among species (Baum *et al.*, 2001). In *Arabidopsis*, *CRABS CLAW* (*CRC*) was identified as a key regulator of nectary formation (Bowman & Smyth, 1999), and it has since been shown to be widely expressed among angiosperms (Thornburg, 2007; Roy *et al.*, 2017; Obeng-Darko, 2018), however additional regulators are required (Liao *et al.,* 2025), and it isn't always associated with nectary development (e.g. Liu *et al.*, 2019). Studies in lineages beyond the Brassicaceae and Solanaceae have identified other regulators required for nectary development. These include *GoNe* in *Gossypium* (Pei *et al.*, 2021) and *STYLISH* (*STY*) in *Aquilegia* (Min *et al.*, 2019). There has been limited work in non-model species, and there are likely further novel regulators yet to be identified (Liao *et al.*, 2025). See Liao *et al.* (2025) and references within for a review of the developmental basis of nectary diversity.

Key genes that contribute to controlling nectar composition and volume in the eccrine model of nectar secretion have been characterised. In the standard model, starch is accumulated in amyloplasts prior to anthesis, and the starch is then hydrolysed into hexoses. Those hexoses provide the precursors for sucrose synthesis by Sucrose Phosphate Synthase – SPS, which is then exported via the uniporter SWEET9 into the apoplast. The sucrose in the apoplast is then hydrolysed into fructose and glucose by a cell wall invertase CWINV4, creating a sucrose gradient which is required for SWEET9 to continue to export sucrose from the symplast as it is bidirectional (Lin *et al.*, 2014; Minami *et al.*, 2021, Fig. **1.1**). Plants carrying mutations in the genes coding for SPS, SWEET9 or CWINV4 do not produce nectar,

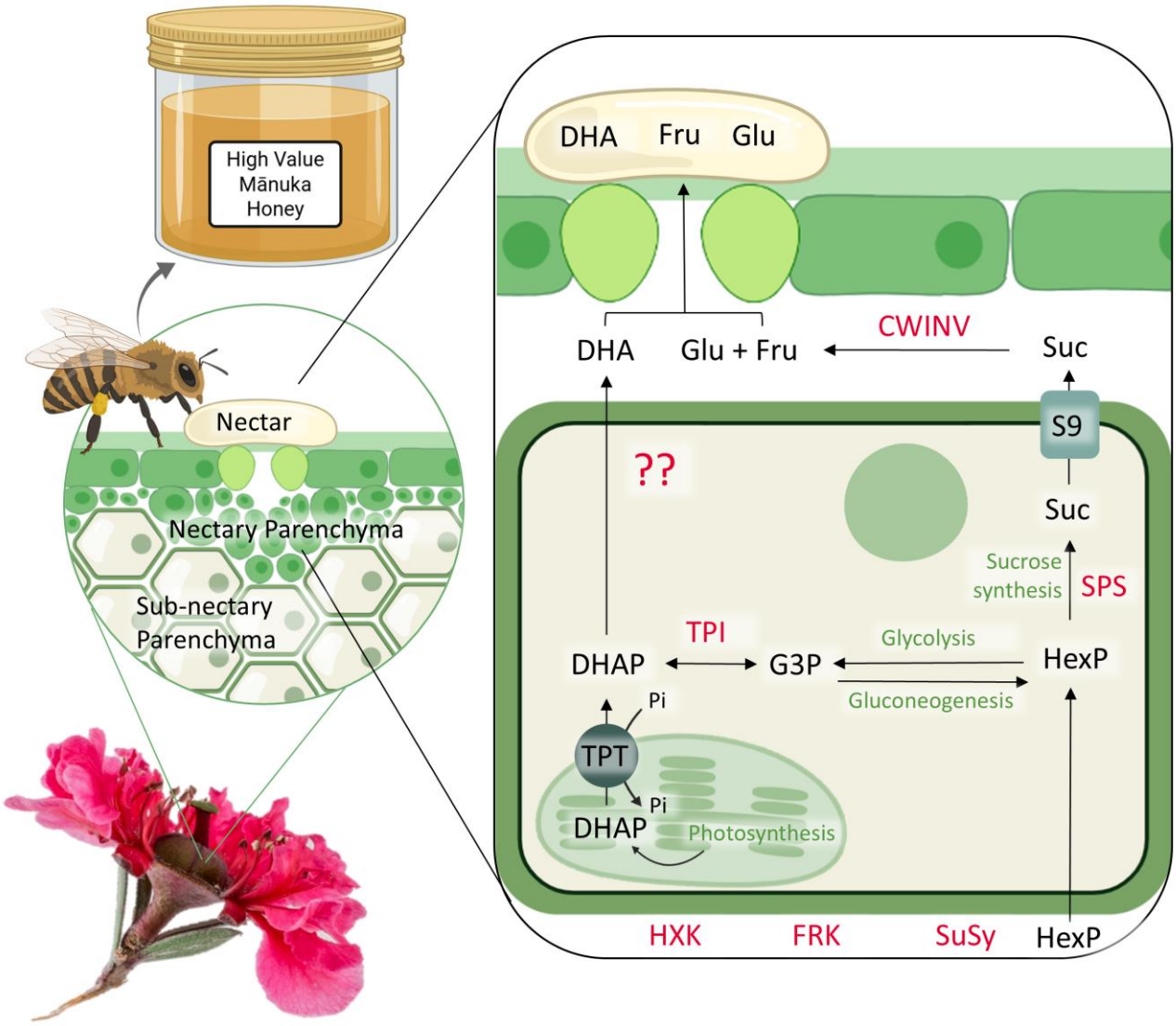


Figure 1.1: Simplified proposed mānuka (*Leptospermum scoparium)* nectar production model. Developed from the proposed nectar production model in Clearwater *et al.* (2021). The sugar pathway through sucrose phosphate synthase (SPS), SWEET9 (S9), and Cell Wall Invertase 4 (CWIN4) represents the standard model of eccrine nectar secretion. In this mānuka model, the nectary sugars are unloaded into the subnectary parenchyma then enter the hexose phosphate (HexP) pool and primary metabolic pathways in nectary parenchyma via sucrose synthase (SuSy), hexokinase (HXK) or fructokinase (FRK). The nectary HexP pool is also contributed to by photosynthesis in nectary chloroplasts. Photosynthesis derived DHAP is exported via the Triose Phosphate Translocator (TPT) and is subsequently isomerised by Triose Phosphate Isomerase (TPI) to form glyceraldehyde-3-phosphate (G3P) and is further converted to HexP by gluconeogenesis. Some of this exported DHAP is dephosphorylated by an unknown enzyme to form DHA, which then diffuses into the apoplast and into the nectar. Figure created with some components from BioRender.

and conversely, overexpression of *Sweet9* leads to an increase in nectar (Ruhlmann *et al.*, 2010; Lin *et al.*, 2014). However, the role these genes play in nectar production in more diverse nectaries in non-model species is unknown (Liao *et al.*, 2025). The production of specialised metabolites in nectar has also started to be characterised (e.g. Roy *et al.*, 2017; Magner *et al.*, 2023, 2024) but relatively few species have been studied in detail.

The mechanisms controlling gene expression and tissue specificity are complex and often involve multiple *cis* (e.g DNA sequence motifs) and *trans* (e.g. transcription factors) acting regulatory elements (see reviews by Schmitz *et al.*, 2022; Marand *et al.*, 2023; Beernink *et al.*, 2025). The genes promoter contains *cis*-regulatory motifs that can enhance or repress gene expression through the binding of transcription factors. Distal *cis* elements include enhancers and repressors further from the target gene that can be multiple megabases away, but that are brought into close proximity through chromatin looping (Beernink *et al.*, 2025). Many other factors can affect gene expression and transcript abundance, such as chromatin accessibility and microRNA-triggered degradation (Bartel, 2009; Schmitz *et al.*, 2022). It has been hypothesised that as multiple nectary types generally use the same set of genes, nectary development and function could involve a core nectar regulatory module (Liao *et al.*, 2025).

1.3 *Leptospermum scoparium* - Mānuka

Mānuka (*Leptospermum scoparium* J.R. Forst et G. Forst) belongs to Myrtaceae family, along with kānuka, *Eucalyptus*, rātā and pōhutukawa. *L. scoparium* is widespread throughout Aotearoa-New Zealand, and potentially our only indigenous *Leptospermum*. The genus has been shown to be polyphyletic (Binks *et al.*, 2022) and has recently been revised and split into five genera – *Aggreflorum*, *Acetospermum*, *Gaudium*, *Leptospermopsis* and *Leptospermum* (Fig. **1.2**, Wilson & Heslewood, 2023). The *Leptospermum* genus is concentrated in Australia, and *L. scoparium* also grows naturally in south-eastern Australia (Thompson, 1989). Recent modelling suggests that within *L. scoparium* two clades diverged ~9-12 million years ago and the Aotearoa-New Zealand mānuka contains 5 geographically distinct gene pools (Koot *et al.*, 2022). Further taxonomic assessment has been called for by numerous authors (Perry *et al.*, 1997; Stephens *et al.*, 2005; Lange & Schmid, 2021; Schmid *et al.*, 2023), but recent genotyping by Chagné *et al.*, (2023) did not support further species segregation within *L. scoparium* in Aotearoa-New Zealand based on significant gene flow occurring between subpopulations in Aotearoa-New Zealand, but did support the division of *L. scoparium* between Australia Aotearoa-New Zealand.

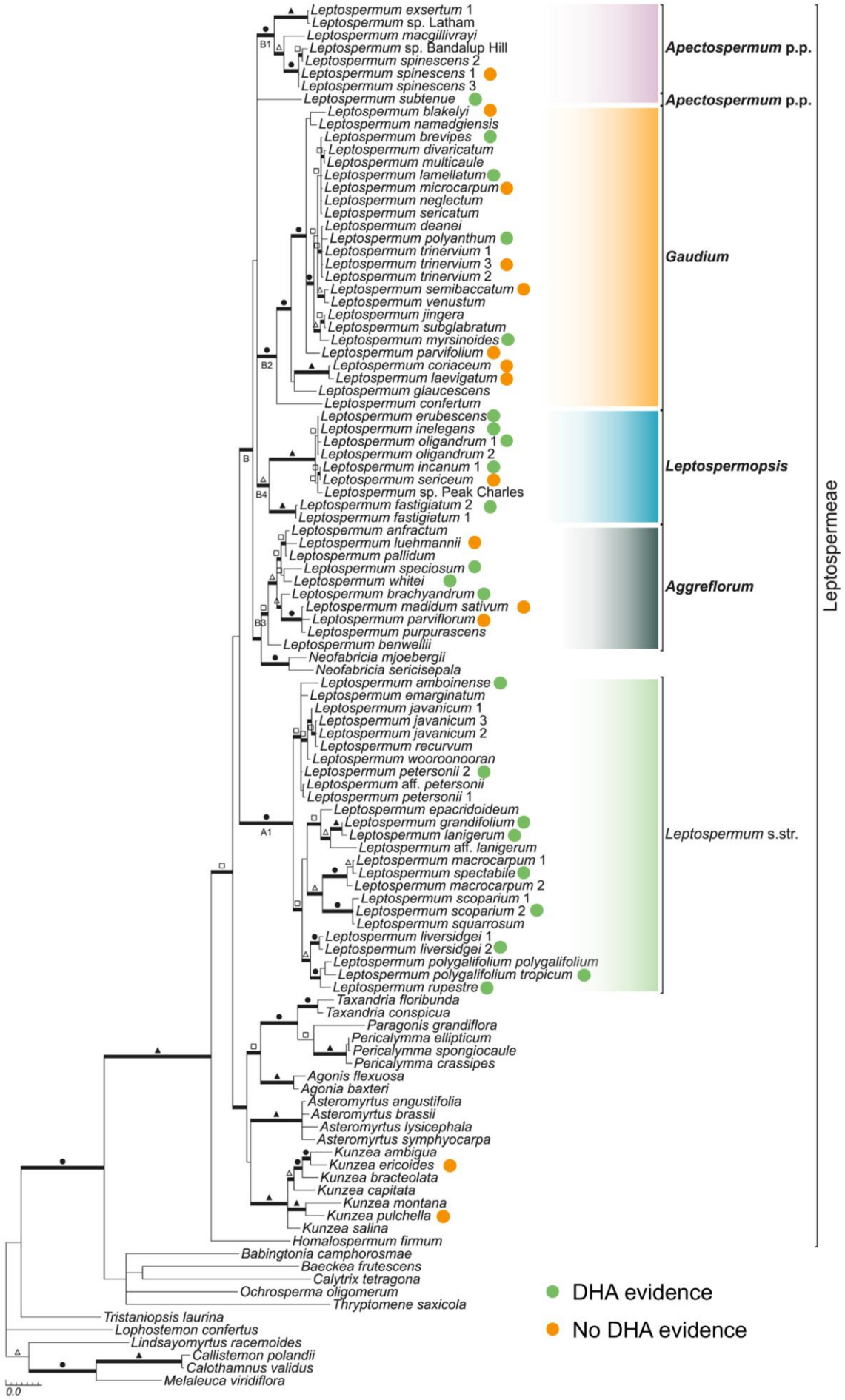

Leptospermum exsertum 1
Leptospermum sp. Latham
Leptospermum macgillivrayi
Leptospermum sp. Bandalup Hill
Leptospermum spinescens 2
Leptospermum spinescens 1
Leptospermum spinescens 3
Leptospermum subtenue
Leptospermum blakelyi
Leptospermum namadgiensis
Leptospermum brevipes
Leptospermum divaricatum
Leptospermum multicaule
Leptospermum lamellatum
Leptospermum microcarpum
Leptospermum neglectum
Leptospermum sericatum
Leptospermum deanei
Leptospermum polyanthum
Leptospermum trinervium 1
Leptospermum trinervium 3
Leptospermum trinervium 2
Leptospermum semibaccatum
Leptospermum venustum
Leptospermum jingera
Leptospermum subglabratum
Leptospermum myrsinoides
Leptospermum parvifolium
Leptospermum coriaceum
Leptospermum laevigatum
Leptospermum glaucescens
Leptospermum confertum
Leptospermum erubescens
Leptospermum inelegans
Leptospermum oligandrum 1
Leptospermum oligandrum 2
Leptospermum incanum 1
Leptospermum sericeum
Leptospermum sp. Peak Charles
Leptospermum fastigiatum 2
Leptospermum fastigiatum 1
Leptospermum anfractum
Leptospermum luehmannii
Leptospermum pallidum
Leptospermum speciosum
Leptospermum whitei
Leptospermum brachyandrum
Leptospermum madidum sativum
Leptospermum parviflorum
Leptospermum purpurascens
Leptospermum benwellii
Neofabricia mjoebergii
Neofabricia sericisepala
Leptospermum amboinense
Leptospermum emarginatum
Leptospermum javanicum 1
Leptospermum javanicum 3
Leptospermum javanicum 2
Leptospermum recurvum
Leptospermum wooroonooran
Leptospermum petersonii 2
Leptospermum aff. petersonii
Leptospermum petersonii 1
Leptospermum epacridoideum
Leptospermum grandifolium
Leptospermum lanigerum
Leptospermum aff. lanigerum
Leptospermum macrocarpum 1
Leptospermum spectabile
Leptospermum macrocarpum 2
Leptospermum scoparium 1
Leptospermum scoparium 2
Leptospermum squarrosum
Leptospermum liversidgei 1
Leptospermum liversidgei 2
Leptospermum polygalifolium polygalifolium
Leptospermum polygalifolium tropicum
Leptospermum rupestre
Taxandria floribunda
Taxandria conspicua
Paragonis grandiflora
Pericalymma ellipticum
Pericalymma spongiocaule
Pericalymma crassipes
Agonis flexuosa
Agonia baxteri
Asteromyrtus angustifolia
Asteromyrtus brassii
Asteromyrtus lysicephala
Asteromyrtus symphyocarpa
Kunzea ambigua
Kunzea ericoides
Kunzea bracteolata
Kunzea capitata
Kunzea montana
Kunzea pulchella
Kunzea salina
Homalospermum firmum
Babingtonia camphorosmae
Baeckea frutescens
Calytrix tetragona
Ochrosperma oligomerum
Thryptomene saxicola
Tristaniopsis laurina
Lophostemon confertus
Lindsayomyrtus racemoides
Callistemon polandii
Calothamnus validus
Melaleuca viridiflora
0.0
B1
B2
B
B4
B3
A1
Apectospermum p.p.
Apectospermum p.p.
Gaudium
Leptospermopsis
Aggreflorum
Leptospermum s.str.
Leptospermeae
DHA evidence
No DHA evidence

Figure 1.2: Bayesian estimate of the phylogeny of tribe Leptospermeae (Myrtaceae). (Page 6). Adapted from Wilson & Heselwood (2023, Figure 4). Phylogeny is based on the combined analysis of two nuclear markers with thick lines showing supported clades (≥0.95 PP). Green circles represent species with evidence of dihydroxyacetone (DHA) in nectar, and orange circles indicate species that have been tested and have been found to have no DHA in their nectar (primarily based on Williams *et al.* 2018, with contributions from Obeng-Darko *et al.* 2023 and McDonald *et al.* 2018). Those with no symbols have not been tested.

Mānuka occupies two main ecological niches – dominating unfavourable sites or as a seral species re-colonising disturbed land (Wardle, 1991; Stephens *et al.*, 2005). Mānuka occupies a wide range of habitats and can tolerate a range of conditions, including low-nutrient soils, harsh environmental conditions, and high or low soil moisture (Stephens *et al.*, 2005; Derraik, 2008). Mānuka can tolerate low phosphate soils (Gutiérrez-Ginés *et al.*, 2019; Noe, 2022), and Noe (2022) found that nectar production and growth was not affected by low phosphate treatments. More recently, the ecological and economic value of mānuka has been recognised. Mānuka can stabilise slopes, prevent erosion, and reduce sediment runoff through their roots binding soils and canopy rain interception (Marden & Phillips, 2015). Mānuka can also reduce runoff of nutrients and microbial contaminants through the production of antimicrobial compounds (Prosser *et al.*, 2016). Mānuka can also be used for carbon sequestration, with mānuka/kānuka shrubland accumulating carbon rapidly, with similarity to plantation forests (Scott *et al.*, 2000; Trotter *et al.*, 2005).

Mānuka is highly polymorphic (Cockayne, 1912; Yin *et al.*, 1984; Dawson, 2009). Variation has been observed in morphology of leaves, growth form, root anatomy, stem and foliage colour, leaf density, and flower colour and size (Stephens *et al.*, 2005; Dawson, 2009). Some variation is due to the environmental effects of the vast ecological range it occupies, but common garden experiments have demonstrated that much is due to genotypic variation (Yin *et al.*, 1984; Dawson, 2009). At least 150 ornamental cultivars have been bred from *L. scoparium*, much of which has been based on wild variation (Dawson, 2009). *L. scoparium* contains bioactive essential oils, and a range of chemotypes has been observed (Fig. **1.3a**, Perry *et al.*, 1997; Porter & Wilkins, 1999; Douglas *et al.*, 2004; van Klink *et al.*, 2005). Perry *et al.* (1997) identified three chemotypes, and a distinction between *L. scoparium* from Aotearoa-New Zealand and Australia. Porter & Wilkins (1999) proposed a fourth chemotype, and Douglas *et al.* (2004) later identified 11 chemotypes, highlighting the variation within and between mānuka genotypes (Fig. **1.3a**). Analysis by Chagné *et al.* (2023) also identified distinct subpopulations (Fig. **1.3b**), but also found significant gene flow occurring between them. Mānuka oils from Tairāwhiti – the East Cape – has shown to have distinctly higher amounts of triketones (leptospermone and flavesone) which have antimicrobial activity (Perry *et al.*, 1997; van Klink *et al.*, 2005).

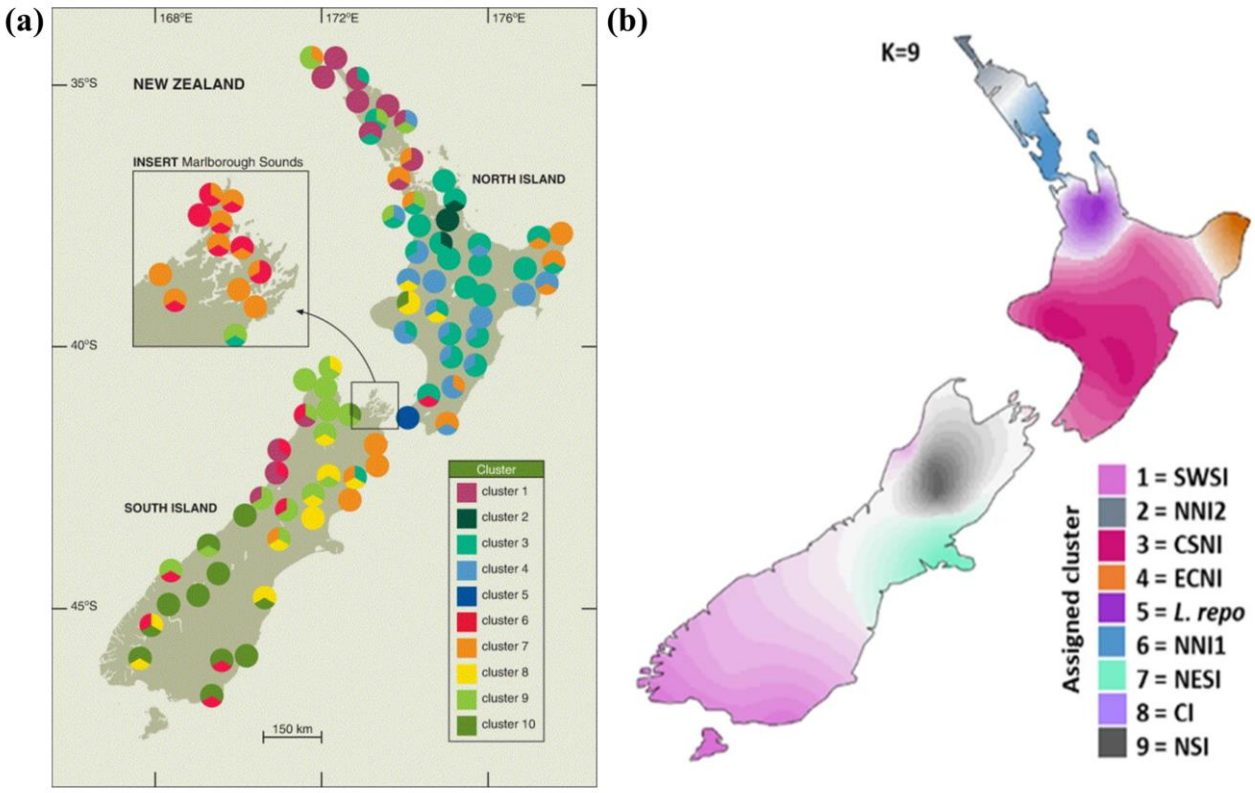


Figure 1.3: Chemotype and Genetic diversity of *Leptospermum scoparium* in Aotearoa-New Zealand. (a) Cluster analysis of essential oil composition of *L. scoparium* in Aotearoa-New Zealand taken from Douglas *et al.* (2004) (b) Geographical distribution of Aotearoa-New Zealand clusters based on SNP array, taken from Chagné *et al.* (2023).

Honey has historically been used to treat infections and wounds, and is effective due to its high osmolarity, acidity, and presence of hydrogen peroxide (Molan, 2006). Mānuka honey was discovered to have additional non-peroxide antimicrobial activity (Molan & Russell, 1988) which gave it increased medicinal value. Mavric *et al.* (2008) discovered that the non-peroxide antibacterial compound in mānuka honey was methylglyoxal (MGO), and Adams *et al.* (2009) demonstrated that the origin of MGO was dihydroxyacetone (DHA) from within in mānuka nectar. The nectar DHA converts to the antimicrobial compound methylglyoxal (MGO) in the derived honey, through a non-enzymatic dehydration that has been well studied (Grainger *et al.*, 2016, Owens *et al*. 2019 and references therein). The discovery of the additional antibacterial effect of mānuka honey made it attractive medicinally, as the MGO is not broken down by the body's catalase as hydrogen peroxide would be (Molan, 2006). The medical market has strong demand for high-MGO mānuka honey, and a range of medical products have been developed (Stephens *et al.*, 2005). The antimicrobial properties of mānuka have also made it increasingly popular in the health food industry (McPherson, 2016).

Honey is a valuable industry in Aotearoa-New Zealand, with mānuka-dominated honey exports valued at almost NZ$420M in 2024-25 (MPI, 2025). Mānuka honey sells for a significant premium

over other honeys Aotearoa-New Zealand (in 2023, $48.40/kg average for monofloral mānuka honey vs $21.40/kg for non-mānuka, MPI, 2023). There has been research interest into how to increase production through mānuka plantations (e.g. the government and industry funded 'High-Performance Mānuka Plantations' Primary Growth Partnership' (PGP), Douglas *et al.* 2019), and cultivar selection through traditional breeding or selection (Nickless, 2015).

1.4 Mānuka – a taonga

Māori, Aotearoa-New Zealand's indigenous people, have a long history of the use of mānuka throughout Aotearoa-New Zealand. There were at least six names for *L. scoparium* – the most widely used being mānuka and kahikatoa (Moorfield, 2005; de Lange, 2021). Mānuka was used by Māori for food, medicine and timber (Stephens *et al.*, 2005). Māori traditional medicinal uses include a decoction of the leaves being drunk, applied as a salve, chewed, or the vapours inhaled (Brooker *et al.*, 1987). The bark was boiled and the liquid drunk, capsules were chewed or boiled (Allan Herbarium, 2000). Pia mānuka (sugary gum exuded from young branches) was used to alleviate coughs (Stephens *et al.*, 2005). Mānuka branches were used to splint broken limbs (Jones, 2007), and the leaves were used as a substitute for tea (Allan Herbarium, 2000; Stephens *et al.*, 2005). More recently, Māori are important participants in the honey industry, and many of the natural mānuka stands used for honey production are on Māori owned land.

1.5 Mānuka nectar and dihydroxyacetone (DHA)

Mānuka flowers are dish shaped with visible nectar, they are open-access and visited by a range of insect pollinators (Fig. **1.4**, Murphy & Robertson, 2000). Flowers within the Myrtaceae family generally produce nectar from the inner surface of the hypanthium, through modified stomata (O'Brien *et al.*, 1996; Davis, 1997). Mānuka nectary anatomy has recently been shown to consist of subnectary parenchyma, nectary parenchyma, and an epidermis with modified stomata (Obeng-Darko, 2018, Fig. **1.1**). This structure is similar to other mesenchymal nectaries upon which the standard eccrine secretion model is based, however Obeng-Darko (2018) also found that mānuka nectaries do not accumulate starch and contain many chloroplasts.

Mānuka nectar is generally hexose rich, and is composed of fructose and glucose, as well as small amounts of sucrose and DHA (Fig. **1.5**, Clearwater *et al.*, 2018; Noe *et al.*, 2019). The DHA amount found in nectar is not directly linked to synthesis of other sugars, and there is also evidence for selective reabsorption of nectar in mānuka as the ratios of fructose to glucose change over the life of the flower (Clearwater *et al.*, 2018). Mānuka nectar also contains other unique phytochemicals, including leptosperin and lepteridine (Smallfield *et al.*, 2018). Nectar traits in mānuka vary with

genotype, floral stage and environmental factors (Clearwater *et al.*, 2018; Noe *et al.*, 2019). Noe *et al*. (2019) in a study across five sites in the North Island found that nectar traits, including nectar DHA, varied more among plants than among sites, suggesting genetic variation was a contributing factor. Mānuka flowers are andromonecious, with both male and hermaphroditic flowers, and the male flowers have been shown to produce nectar with higher ratio of DHA to total sugar (Williams *et al.*, 2018).

DHA is found in nectar of a range of species within the Leptospermeae tribe (Fig. **1.2**), with differences in amount between species (Williams *et al*., 2018). This trait until recently was thought to be confined to the previously defined *Leptospermum* genus – however two more nectar-DHA producing species outside of the Leptospermeae have recently been identified (Obeng-Darko *et al.*,

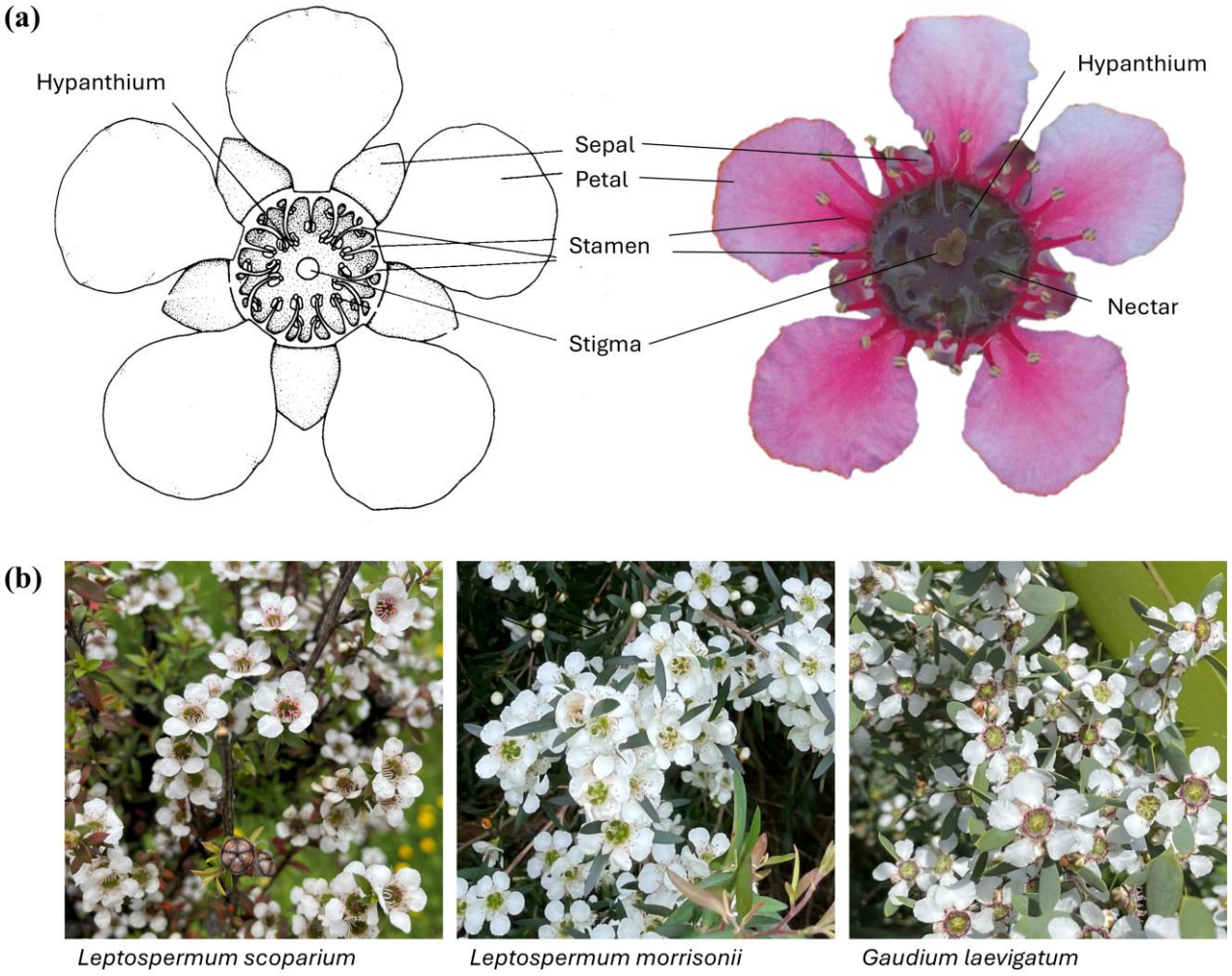


Figure 1.4: *Leptospermum* floral anatomy and some of the species used in the present study. (a) Adapted from Thompson *et al.,* 1989, *Leptospermum* floral morphology. Of specific relevance to the present study are the hypanthium which contains the nectaries and is where the nectar produced accumulates. Mānuka image from Clearwater *et al.* (2018) (b) Three Leptospermeae species discussed within this thesis - *Leptospermum scoparium*, *Leptospermum morrisonii* and *Gaudium laevigatum*.

2023). These species – *Ericomyrtus serpyllifolia* and *Verticordia chrysantha*, belong to genera within the Chamelaucieae – the sister tribe to the Leptospermeae (Wilson *et al.*, 2001). The presence of DHA in nectar is a trait that has been found so far among a total of six genera within the Myrtaceae family and is often observed in some but not all members of the genus – such as presence in *V. chrysantha*, absent in *Verticordia picta* (Obeng-Darko *et al*., 2023), and presence in *Aggreflorum whitei*, and absence in *Aggreflorum parviflorium* (Williams *et al.,* 2018). The recent discovery of DHA producing species outside of the Leptospermeae highlights that there are many species within these tribes with nectar that is so far untested for presence of DHA (Obeng-Darko *et al*., 2023).

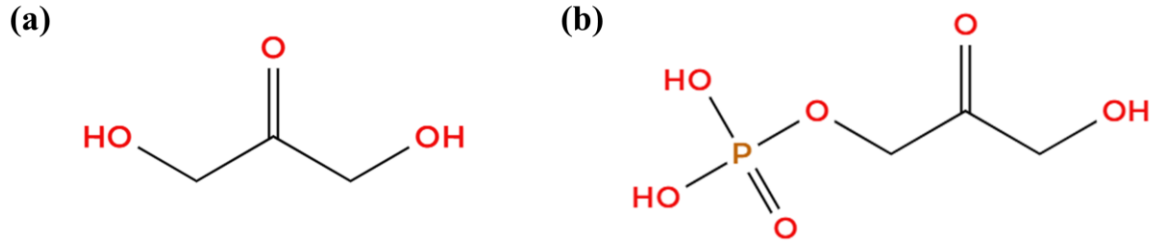


Figure 1.5: Structure of (a) Dihydroxyacetone (DHA) and (b) Dihydroxyacetone phosphate (DHAP).

An alternative hypothesis for the origin of DHA in nectar is through microbial production (Williams *et al.*, 2014). There are microbes that are known to produce DHA (Lidia & Stanisław, 2012), and microbes are known to inhabit nectar and alter nectar chemistry (Vannette & Fukami, 2018). It has been demonstrated that mānuka has a core microbiome (Wicaksono *et al.*, 2016; Noble *et al.*, 2020; Larrouy *et al.*, 2023), and that mānuka essential oil chemistry can be altered by microbes (Wicaksono *et al.*, 2017). When sequencing the mānuka genome, Thrimawithana *et al.* (2019) found many bacterial and fungal sequence reads, indicating an associated endophytic microbiome. Kauffman-Bresciano (2021) identified a ubiquitous microbial taxa found in mānuka nectar (*Pantoea agglomerans*), that was able to significantly increase nectar DHA *in vitro*. Larrouy *et al.* (2023) has characterised the floral microbiome of mānuka and found a wide diversity of taxa both endophytic and epiphytic.

Mānuka and other Myrtaceous nectaries have been hypothesised to be photosynthetic (Nepi, 2007; Clearwater *et al.*, 2018). They fit the green nectary model of open flowers, green nectaries, and a long nectar production phase (Nepi, 2007). Recent anatomical investigation of mānuka nectaries has also observed that they do not accumulate starch (Obeng-Darko, 2018). Clearwater *et al.* (2021) has recently demonstrated that mānuka nectaries photosynthesise, and that photosynthesis within the flower contributes to nectar production. It was hypothesised that photosynthesising nectary chloroplasts could have a role in regulating nectar flow and composition in response to light.

Clearwater *et al*. (2021) proposed a model of nectar production in mānuka and other plants with long lived green nectaries (Fig. **1.1**), combining elements from previous models and integrating a role of nectary photosynthesis. They also hypothesised that the origin of nectar DHA is through pools of triose phosphates modulated by photosynthesis in the nectary.

1.6 Context of the present research

Despite the fundamental importance of nectar to crop pollination and honey industries, the genetic control of nectary function is poorly understood. The basis of this valuable and variable trait of nectar DHA in mānuka is currently unknown. Understanding the molecular basis of this trait would allow development of gene-based tools to identify high value germplasm and inform replanting and breeding programmes. Export prices for monofloral mānuka honey are strongly linked to MGO content and therefore nectar DHA, so increasing nectar DHA could create additional value for the Aotearoa-New Zealand honey industry. Being able to genetically identify high-DHA potential lines more easily than with the usual nectar sampling would allow selection at pre-flowering life stages and higher throughput. Being able to identify many genotypes is preferred as this can ensure resilient populations – for example by preserving biodiversity for disease resistance (e.g. myrtle rust – Frampton *et al.*, 2024). The potential of mānuka plantations for high value honey has been explored through the High-Performance Mānuka Plantations PGP, with recommendations including incorporation of locally adapted wild mānuka, and use of multiple mānuka varieties (Douglas *et al.*, 2019).

Partnering with Māori is an important consideration for work in this area (Morgan *et al.*, 2019), as we are working with a taonga species and Māori are key stakeholders in this industry. In developing the present research with our partners Ngati Porou Miere Rangahau, we established that there was a need to understand the variation in honey value between their blocks, and to be able to produce a consistently high value product while still eco-sourcing plant lines from their rohe (territory or boundary of an iwi/tribe). Sourcing plants from the area in which they will be grown (eco-sourcing) is beneficial as plants are well adapted to the environment, and in the case of this taonga crop species respecting whakapapa (the Māori concept of genealogy and lineage) and maintaining provenance and regional diversity is also important.

Recent work has provided genomic resources for mānuka research. Chagné *et al.* (2017) developed Short Simple Repeat (SSR) markers. Chagné *et al.* (2019) constructed a linkage map for mānuka genotyping by sequencing (GBS) and developed a segregating population using Tairāwhiti - East Cape mānuka accessions. These research projects, and others, have been underpinned by the generation of a complete assembly of the mānuka genome (Thrimawithana *et al*., 2019). Koot *et al.*, (2022)

recently explored the genetic structure and relatedness of mānuka across Aotearoa-New Zealand, and between Aotearoa-New Zealand and Australian populations. They found five geographically distinct mānuka gene pools and found that the Aotearoa-New Zealand and Australian populations are genetically distinct. Chagné *et al*. (2023) conducted further genotyping and found significant gene flow occurring between subpopulations in Aotearoa-New Zealand and further supported the division of *L. scoparium* between Australia and Aotearoa-New Zealand.

We hypothesise that there is a plant genetic basis for nectar DHA. Clearwater *et al*. (2021) demonstrated that photosynthesis is contributing to nectar sugars including DHA in mānuka nectar and described a modified eccrine model of nectar production. Clearwater *et al*. (2021) also hypothesised that photosynthesis derived DHAP is dephosphorylated by an unknown phosphatase enzyme, and diffuses in trace amounts into the apoplast, and from there into the nectar (Fig. **1.1**). Investigating this hypothesis has implications beyond mānuka – especially as this trait is present throughout the Leptospermeae and Chamelaucieae phylogeny.

1.7 Research aims and thesis outline

My thesis aims to elucidate the molecular basis of dihydroxyacetone (DHA) production in *Leptospermum* nectar. It is organised in five research chapters, including one published chapter and one chapter submitted for publication, followed by synthesis and conclusions.

Chapter 2 - Published paper: A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*)

What genes and genomic regions are associated with nectar DHA in *L. scoparium*?

This chapter uses two complementary approaches to identify genes underpinning the nectar DHA trait in *Leptospermum scoparium*. The first was to use RNAseq analysis to identify nectary-associated genes differentially expressed between high and low nectar DHA genotypes. The second approach used a mānuka high-density linkage map and quantitative trait loci (QTL) mapping population, supported by an improved genome assembly, to reveal genetic regions associated with nectar DHA content. Expression and QTL analyses both pointed to the involvement of a phosphatase gene, *LsSgpp2*. The expression pattern of *LsSgpp2* correlated with nectar DHA accumulation, and it co-located with a QTL on chromosome 4. The identification of a total of three QTLs, the first reported for a minor abundance compound of plant nectar, indicates polygenic control of DHA content. We established plant genetics as a key influence on DHA accumulation. The data suggest the hypothesis of LsSGPP2 releasing DHA from DHA-phosphate and variability in *LsSgpp2* gene expression contributing to variation in this nectar trait. This paper is published in New Phytologist and contains contributions from multiple authors (described in co-authorship declaration in appendices) including from myself prior to my PhD start date. The majority of the work however was conducted by myself during my PhD.

https://doi.org/10.1111/nph.19714

Chapter 3 - Submitted paper: Genome and transcriptome assemblies of Australian tea trees *Leptospermum morrisonii* and *Gaudium laevigatum*.

What genetic diversity is present in contrasting DHA phenotypes within the Leptospermeae tribe?

In order to further understand the nectar DHA trait in the wider Leptospermeae tribe, genomes for two Australian species with differing nectar DHA phenotypes were produced. We produced high-quality haplotyped genome assemblies of *Leptospermum morrisonii* (contains nectar DHA)

and *Gaudium laevigatum* (no nectar DHA), using high-fidelity long-read sequencing and synteny-based scaffolding against the *Leptospermum scoparium* genome. These new genome assemblies provide an important resource for further work studying traits related to nectar chemistry, as well as foliage chemistry and understanding the genetic diversity within the Leptospermeae tribe. This paper was primarily written by David Chagné, with contributions from myself and a third author (described in co-authorship declaration in appendices). My main contributions to this paper were developing the idea, acquiring funding, and all field and lab-based work.

Chapter 4 – Draft paper to be submitted: Regulation of *Sgpp2* and the structure of the *Sgpp* locus in the Leptospermeae

Does *Sgpp2* expression correlate with DHA amounts in other Leptospermeae species? How is *Sgpp2* expression regulated?

This chapter aimed to further understand the regulation of the *Sgpp2* gene and the structure of the *Sgpp* locus within the Leptospermeae and wider Myrtaceae family. Chapter 2 identified three *Sgpp* genes in a complex locus, of which *Sgpp2* was highly expressed and correlated with nectar DHA. The *Sgpp* locus, gene sequences, and transcript abundances were compared among the Leptospermeae species – *L. morrisonii*, and *G. laevigatum,* in the context of the previously described *L. scoparium*. To assess the *LmSgpp2* promoter *in vitro* transgenic lines of *Petunia* and *Nicotiana* we generated using the *LmSgpp2* promoter driving *Venus* expression. Conserved and unique regions were identified between and within the *Sgpp*s and regions and binding motifs unique to the highly expressed *Sgpp*s were identified. *Sgpp2* transcript abundances correlated with DHA amounts in two further Leptospermeae species - providing further evidence that this gene is responsible for DHA production. Within the conserved promoter regions unique to the highly expressed *Sgpp2*s were two C-box motifs, and a bZIP11 type transcription factor predicted to bind to them was identified as significantly differentially expressed in the RNAseq dataset presented in Chapter 2, and was subsequently found to co-locate with another of the QTLs identified in Chapter 2. The promoter region across 13 *L. scoparium* genotypes was found to be highly similar, indicating further complex transcriptional regulation fine tuning *Sgpp2* expression. The *L. morrisonii Sgpp2* promoter driving *Venus* lines of *Petunia* and *Nicotiana* were found to have strong nectary specific expression.

Chapter 5 – Draft paper to be submitted: Enzymatic characterisation of the *Leptospermum morrisonii* SGPP2

Can SGPP2 dephosphorylate DHAP to produce DHA and phosphate *in vitro*?

This chapter aims to establish whether SGPP2 was able to dephosphorylate DHAP and compare kinetic parameters of three *Leptospermum morrisonii* SGPPs with the *Arabidopsis thaliana* SGPP. These three LmSGPP enzymes, along with the AtSGPP, were produced and purified in microbial systems, and assayed for phosphatase activity on three substrates including DHAP. I aimed to analyse in detail the activity of SGPP2 on DHAP and compare the phosphatase activity and substrate specificity between the *Leptospermum* SGPPs and AtSGPP. I found that LmSGPP2 was able to dephosphorylate DHAP, and hypothesised that shifts in kinetic parameters between the SGPPs could be due to changes in in the substrate specificity loop.

Chapter 6 – Draft paper to be submitted: Overexpression of *Leptospermum morrisonii Sgpp2* in the nectaries of *Petunia* and *Nicotiana*

Can overexpression of *LmSgpp2* in *Petunia* and *Nicotiana* nectaries produce DHA in nectar?

To assess the effect of *LmSgpp2* expression in nectaries that do not produce DHA, I generated transgenic lines overexpressing the *L. morrisonii Sgpp2* under the nectary specific promoter *Petunia axillaris Nec1*/*Sweet9.* No increase in DHA was found in the nectar of these transgenic lines, however *LmSgpp2* transcript abundance was found to be significantly less than *Venus* transcript abundance under the same promoter. The *PaNec1* promoter was found to be not as strong and nectary specific than previously described in the literature, and it was suggested to use the *LmSgpp2* promoter driving *LmSgpp2* in future work.

Chapter 7 - Synthesis

The results and conclusions from each chapter are synthesised, and the direction of possible future work is described.

1.8 References

Adams CJ, Manley-Harris M, Molan PC. **2009**. The origin of methylglyoxal in New Zealand manuka (*Leptospermum scoparium*) honey. *Carbohydrate Research* **344**: 1050–1053.

Allan Herbarium. **2000**. Ngā Tipu o Aotearoa - New Zealand Plant Names Database. *RECORD NUMBER: 1115*.

Bartel DP. **2009**. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**: 215–233.

Baum SF, Eshed Y, Bowman JL. **2001**. The *Arabidopsis* nectary is an ABC-independent floral structure. *Development* **128**: 4657–67.

Beernink BM, Vogel JP, Lei L. **2025**. Enhancers in Plant Development, Adaptation and Evolution. *Plant and Cell Physiology* **66**: 461–476.

Bernardello G. **2007**. A systematic survey of floral nectaries. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Springer, Dordrecht, 19–128.

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**: 348–359.

Bowman JL, Smyth DR. **1999**. CRABS CLAW, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **126**: 2387–96.

Brooker SG, Cambie RC, Cooper RC. **1987**. *New Zealand medicinal plants*. Raupo Publishing.

Carey S, Zenchyzen B, Deneka AJ, Hall JC. **2023**. Nectary development in *Cleome violacea*. *Frontiers in Plant Science* **13**.

Carter C, Thornburg RW. **2004**. Is the nectar redox cycle a floral defense against microbial attack? *Trends in Plant Science* **9**: 320–324.

Carter C, Thornburg RW, Nepi M (Eds.). **2019**. *New Perspectives on the Biology of Nectaries and Nectars*. Frontiers Media SA.

Chagné D, Montanari S, Kirk C, Mitchell C, Heenan P, Koot E. **2023**. Single nucleotide polymorphism analysis in *Leptospermum scoparium* (Myrtaceae) supports two highly differentiated endemic species in Aotearoa New Zealand and Australia. *Tree Genetics & Genomes* **19**: 31.

Chagné D, Ryan J, Saeed M, Van Stijn T, Brauning R, Clarke S, Jacobs J, Wilcox P, Boursault E, Jaksons P, *et al.* 2019. A high density linkage map and quantitative trait loci for tree growth for New Zealand mānuka (*Leptospermum scoparium*). *New Zealand Journal of Crop and Horticultural Science* **47**: 261–272.

Chagné D, Thrimawithana A, Bowatte D, Hilario E, Crowhurst R, Smallfield B, Boase M, Lewis D, de Lange P, Schwinn K. **2017**. Simple sequence repeat (SSR) markers for New Zealand mānuka (*Leptospermum scoparium*) and transferability to kānuka (Kunzea spp.). *New Zealand Journal of Crop and Horticultural Science* **45**: 216–222.

Chatt EC, Mahalim S-N, Mohd-Fadzil N-A, Roy R, Klinkenberg PM, Horner HT, Hampton M, Carter CJ, Nikolau BJ. **2021**. Nectar biosynthesis is conserved among floral and extrafloral nectaries. *Plant Physiology* **185**: 1595–1616.

Clearwater MJ, Noe ST, Manley-Harris M, Truman G-L, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. **2021**. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

Clearwater MJ, Revell M, Noe S, Manley-Harris M. **2018**. Influence of genotype, floral stage, and water stress on floral nectar yield and composition of manuka (*Leptospermum scoparium*). *Annals of Botany* **121**: 501–512.

Cockayne L. **1912**. Observations concerning evolution, derived from ecological studies in New Zealand. *Transactions of the New Zealand Institute* **44**: 1–50.

Davis A. **1997**. Influence of floral visitation on nectar-sugar composition and nectary surface changes in *Eucalyptus*. *Apidologie* **28**: 27–42.

Dawson M. **2009**. A history of *Leptospermum scoparium* in cultivation: Discoveries from the wild. *New Zealand Garden Journal* **12**: 21–25.

Derraik J. **2008**. New Zealand manuka (*Leptospermum scoparium*; Myrtaceae): a brief account of its natural history and human perceptions. *New Zealand Garden Journal* **11**: 4–8.

Douglas B, Carey A, Archer R, Lee S, Olsen M. **2019**. *HIGH PERFORMANCE MĀNUKA PLANTATIONS PGP PROGRAMME FINAL REPORT*. Prepared by Mānuka Research Partnership (NZ) Limited (MRPL) and Massey University.

Douglas M, Klink J, Smallfield B, Perry N, Anderson R, Johnstone P, Weavers R. **2004**. Essential oils from New Zealand manuka: Triketone and other chemotypes of *Leptospermum scoparium*. *Phytochemistry* **65**: 1255–64.

Frampton RA, Shuey LS, David CC, Pringle GM, Kalamorz F, Pegg GS, Chagné D, Smith GR. **2024**. Analysis of Plant and Fungal Transcripts from Resistant and Susceptible Phenotypes of *Leptospermum scoparium* Challenged by *Austropuccinia psidii*. *Phytopathology®* **114**: 2121–2130.

Galetto L, Bernardello G. **2005**. Rewards in flowers: nectar. In: Dafni A, Kevan PG, Husband BC, eds. *Practical Pollination Biology*. Enviroquest, Cambridge, 261–313.

Gardener MC, Gillman MP. **2002**. The taste of nectar – a neglected area of pollination ecology. *Oikos* **98**: 552–557.

Grainger MNC, Manley-Harris M, Lane JR, Field RJ. 2016. Kinetics of conversion of dihydroxyacetone to methylglyoxal in New Zealand mānuka honey: Part I – Honey systems. Food Chemistry **202**: 484–491.

Gutiérrez-Ginés MJ, Madejón E, Lehto NJ, McLenaghen RD, Horswell J, Dickinson N, Robinson BH. **2019**. Response of a Pioneering Species (*Leptospermum scoparium* J.R.Forst. & G.Forst.) to Heterogeneity in a Low-Fertility Soil. *Frontiers in Plant Science* **10**: 93.

Heil M. **2011**. Nectar: generation, regulation and ecological functions. *Trends in Plant Science* **16**: 191–200.

Jones R. **2007**. Rongoā – medicinal use of plants - Other medicinal plants. *Te Ara – the Encyclopedia of New Zealand*.

Katzer AM, Wessinger CA, Anaya BM, Hileman LC. **2025**. CRABS CLAW-independent floral nectary development in Penstemon barbatus. *American Journal of Botany* **112**: e70058.

Kauffman-Bresciano CJ. **2022**. The role of microorganisms in shaping the nectar chemistry of Mānuka plants (*Leptospermum scoparium*). PhD thesis, Massey University, Palmerston North, New Zealand.

van Klink JW, Larsen L, Perry NB, Weavers RT, Cook GM, Bremer PJ, MacKenzie AD, Kirikae T. **2005**. Triketones active against antibiotic-resistant bacteria: synthesis, structure–activity relationships, and mode of action. *Bioorganic & Medicinal Chemistry* **13**: 6651–6662.

Koot E, Arnst E, Taane M, Goldsmith K, Thrimawithana A, Reihana K, González-Martínez SC, Goldsmith V, Houliston G, Chagné D. **2022**. Genome-wide patterns of genetic diversity, population structure and demographic history in mānuka (*Leptospermum scoparium*) growing on indigenous Māori land. *Horticulture Research* **9**: uhab012.

Kram BW, Bainbridge EA, Perera MADN, Carter C. **2008**. Identification, cloning and characterization of a GDSL lipase secreted into the nectar of *Jacaranda mimosifolia*. *Plant Molecular Biology* **68**: 173.

Kram BW, Carter CJ. **2009**. *Arabidopsis thaliana* as a model for functional nectary analysis. *Sexual Plant Reproduction* **22**: 235–246.

Kram BW, Xu WW, Carter CJ. **2009**. Uncovering the *Arabidopsis thaliana* nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. *BMC Plant Biology* **9**: 92.

de Lange PJ. **2021**. *Leptospermum scoparium* var. *scoparium* Fact Sheet (content continuously updated). *New Zealand Plant Conservation Network*.

Lange PJ de, Schmid LMH. **2021**. *Leptospermum repo* (Myrtaceae), a new species from northern Aotearoa / New Zealand peat bog habitats, segregated from *Leptospermum scoparium* s. l. *Ukrainian Botanical Journal* **78**: 247–265.

Larrouy JL, Dhami MK, Jones EE, Ridgway HJ. **2023**. Physiological stage drives fungal community dynamics and diversity in Leptospermum scoparium (mānuka) flowers. *Environmental Microbiology* **25**: 766–771.

Liao IT, Gong Y, Kramer EM, Nikolov LA. **2025**. The developmental basis of floral nectary diversity and evolution. *New Phytologist* **246**: 2462–2477.

Lidia S-R, Stanisław B. **2012**. Production of dihydroxyacetone from an aqueous solution of glycerol in the reaction catalyzed by an immobilized cell preparation of acetic acid bacteria Gluconobacter oxydans ATCC 621. *European Food Research and Technology* **235**: 1125–1132.

Lin IW, Sosso D, Chen LQ, Gase K, Kim SG, Kessler D, Klinkenberg PM, Gorder MK, Hou BH, Qu XQ, *et al.* 2014. Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**: 546–9.

Liu H, Ma J, Li H. 2019. Transcriptomic and microstructural analyses in Liriodendron tulipifera Linn. reveal candidate genes involved in nectary development and nectar secretion. BMC Plant Biology **19**: 531.

Lüttge U. **2013**. Green nectaries: the role of photosynthesis in secretion. *Botanical Journal of the Linnean Society* **173**: 1–11.

Magner ET, Freund Saxhaug K, Zambre A, Bruns K, Carroll P, Snell-Rood EC, Hegeman AD, Carter CJ. **2024**. A multifunctional role for riboflavin in the yellow nectar of *Capsicum baccatum* and *Capsicum pubescens*. *New Phytologist* **243**: 1991–2007.

Magner ET, Roy R, Freund Saxhaug K, Zambre A, Bruns K, Snell-Rood EC, Hampton M, Hegeman AD, Carter CJ. **2023**. Post-secretory synthesis of a natural analog of iron-gall ink in the black nectar of *Melianthus* spp. *New Phytologist* **239**: 2026–2040.

Marand AP, Eveland AL, Kaufmann K, Springer NM. **2023**. *cis*-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annual Review of Plant Biology* **74**: 111–137.

Marden M, Phillips C. **2015**. *A review of research on the erosion control effectiveness of naturally reverting mānuka (Leptospermum scoparium) and kānuka (Kunzea ericoides complex): implications for erosion mitigation of space-planted mānuka on marginal hill country*. Landcare Research, Hawke's Bay Regional Council.

Mavric E, Wittmann S, Barth G, Henle T. **2008**. Identification and quantification of methylglyoxal as the dominant antibacterial constituent of Manuka (*Leptospermum scoparium*) honeys from New Zealand. *Molecular Nutrition & Food Research* **52**: 483–489.

McDonald CM, Keeling SE, Brewer MJ, Hathaway SC. **2018**. Using chemical and DNA marker analysis to authenticate a high-value food, manuka honey. *npj Science of Food* **2**: 9.

McPherson AJ. **2016**. Manuka—A viable alternative land use for New Zealand's hill country. *NZ Journal of Forestry* **61**: 11–19.

Min Y, Bunn JI, Kramer EM. 2019. Homologs of the STYLISH gene family control nectary development in Aquilegia. New Phytologist **221**: 1090–1100.

Minami A, Kang X, Carter CJ. **2021**. A cell wall invertase controls nectar volume and sugar composition. *The Plant Journal* **107**: 1016–1028.

Ministry for Primary Industries (MPI). **2023**. *New Zealand Honey Exports.* [WWW document] URL https://www.mpi.govt.nz/dmsdocument/42360-New-Zealand-honey-exports/ [Accessed 1 October 2023]

Ministry for Primary Industries (MPI). **2025**. *Situation and Outlook for Primary Industries June 2025*. Wellington, New Zealand

Mitchell RJ. **2004**. Heritability of nectar traits: why do we know so little? *Ecology* **85**: 1527–1533.

Molan PC. **2006**. Using honey in wound care. *International Journal of Clinical Aromatherapy* **3**: 21–24.

Molan PC, Russell KM. **1988**. Non-Peroxide Antibacterial Activity in Some New Zealand Honeys. *Journal of Apicultural Research* **27**: 62–67.

Moorfield JC. **2005**. *Te aka: Māori-English, English-Māori dictionary and index*. Longman.

Morgan ER, Perry NB, Chagné D. **2019**. Science at the intersection of cultures – Māori, Pākehā and mānuka. *New Zealand Journal of Crop and Horticultural Science* **47**: 225–232.

Murphy C, Robertson AW. **2000**. *Preliminary study of the effects of honey bees (*Apis mellifera*) in Tongariro National Park*. Department of Conservation Wellington, New Zealand.

Nepi M. **2007**. Nectary structure and ultrastructure. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Springer, Dordrecht, 129–166.

Nickless EM. **2015**. Influential factors in nectar composition and yield in *Leptospermum scoparium*. PhD thesis, Massey University, Palmerston North, New Zealand.

Nicolson SW. **2022**. Sweet solutions: nectar chemistry and quality. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **377**: 20210163.

Noble AS, Noe S, Clearwater MJ, Lee CK. **2020**. A core phyllosphere microbiome exists across distant populations of a tree species indigenous to New Zealand. *Plos One* **15**: e0237079.

Noe ST. **2022**. The effects of soil nutrient addition on mānuka (*Leptospermum scoparium*) growth, flowering, and nectar production. PhD thesis, University of Waikato, Hamilton, New Zealand.

Noe S, Manley-Harris M, Clearwater MJ. **2019**. Floral nectar of wild mānuka (*Leptospermum scoparium*) varies more among plants than among sites. *New Zealand Journal of Crop and Horticultural Science* **47**: 282–296.

Obeng-Darko S. **2018**. Anatomical and molecular basis of floral nectar secretion by mānuka (*Leptospermum scoparium*). MSc thesis, University of Waikato, Hamilton, New Zealand.

Obeng-Darko SA, Sloan J, Binks RM, Brooks PR, Veneklaas EJ, Finnegan PM. **2023**. Dihydroxyacetone in the Floral Nectar of *Ericomyrtus serpyllifolia* (Turcz.) Rye (Myrtaceae) and *Verticordia chrysantha* Endl. (Myrtaceae) Demonstrates That This Precursor to Bioactive Honey Is Not Restricted to the Genus *Leptospermum* (Myrtaceae). *Journal of Agricultural and Food Chemistry* **71**: 7703–7709.

O'Brien SP, Loveys BR, Grant WJR. **1996**. Ultrastructure and Function of Floral Nectaries of *Chamelaucium uncinatum* (Myrtaceae). *Annals of Botany* **78**: 189–196.

Pacini E, Nepi M. **2007**. Nectar production and presentation. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Dordrecht: Springer Netherlands, 167–214.

Parachnowitsch AL, Manson JS, Sletvold N. **2019**. Evolutionary ecology of nectar. *Annals of Botany* **123**: 247–261.

Pei Y, Zhang J, Wu P, Ye L, Yang D, Chen J, Li J, Hu Y, Zhu X, Guo X, et al. 2021. GoNe encoding a class VIIIb AP2/ERF is required for both extrafloral and floral nectary development in Gossypium. The Plant Journal **106**: 1116–1127.

Perry NB, Brennan NJ, Van Klink JW, Harris W, Douglas MH, McGimpsey JA, Smallfield BM, Anderson RE. **1997**. Essential oils from New Zealand manuka and kanuka: chemotaxonomy of *Leptospermum*. *Phytochemistry* **44**: 1485–1494.

Porter NG, Wilkins AL. **1999**. Chemical, physical and antimicrobial properties of essential oils of *Leptospermum scoparium* and Kunzea ericoides. *Phytochemistry* **50**: 407–415.

Prosser JA, Woods RR, Horswell J, Robinson BH. **2016**. The potential in-situ antimicrobial ability of Myrtaceae plant species on pathogens in soil. *Soil Biology and Biochemistry* **96**: 1–3.

Quevedo-Caraballo S, de Vega C, Lievens B, Fukami T, Álvarez-Pérez S. **2025**. Tiny but mighty? Overview of a decade of research on nectar bacteria. *New Phytologist* **245**: 1897–1910.

Raguso RA. **2004**. Why are some floral nectars scented? *Ecology* **85**: 1486–1494.

Ren G, Healy RA, Klyne AM, Horner HT, James MG, Thornburg RW. **2007**. Transient starch metabolism in ornamental tobacco floral nectaries regulates nectar composition and release. *Plant Science* **173**: 277–290.

Roy R, Schmitt AJ, Thomas JB, Carter CJ. **2017**. Review: Nectar biology: From molecules to ecosystems. *Plant Science* **262**: 148–164.

Ruhlmann JM, Kram BW, Carter CJ. **2010**. CELL WALL INVERTASE 4 is required for nectar production in *Arabidopsis*. *Journal of Experimental Botany* **61**: 395–404.

Schmid LMH, de Lange PJ, Marshall AJ. **2023**. Leptospermum hoipolloi (Myrtaceae), a new species from Aotearoa / New Zealand, segregated from Leptospermum scoparium s. l. *Ukrainian Botanical Journal* **80**: 173–198.

Schmitz RJ, Grotewold E, Stam M. **2022**. *Cis*-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell* **34**: 718–741.

Scott NA, White JD, Townsend JA, Whitehead D, Leathwick JR, Hall GM, Marden M, Rogers GN, Watson AJ, Whaley PT. **2000**. Carbon and nitrogen distribution and accumulation in a New Zealand scrubland ecosystem. *Canadian Journal of Forest Research* **30**: 1246–1255.

Silva FA, Guirgis A, Thornburg R. **2018**. Nectar Analysis Throughout the Genus Nicotiana Suggests Conserved Mechanisms of Nectar Production and Biochemical Action. *Frontiers in Plant Science* **9**: 1100.

Slavković F, Dogimont C, Morin H, Boualem A, Bendahmane A. **2021**. The Genetic Control of Nectary Development. *Trends in Plant Science* **26**: 260–271.

Smallfield BM, Joyce NI, van Klink JW. **2018**. Developmental and compositional changes in *Leptospermum scoparium* nectar and their relevance to mānuka honey bioactives and markers. *New Zealand Journal of Botany* **56**: 183–197.

Solhaug EM, Roy R, Chatt EC, Klinkenberg PM, Mohd-Fadzil N-A, Hampton M, Nikolau BJ, Carter CJ. **2019**. An integrated transcriptomics and metabolomics analysis of the *Cucurbita pepo* nectary implicates key modules of primary metabolism involved in nectar synthesis and secretion. *Plant Direct* **3**: e00120.

Stephens J, Molan P, Clarkson B. **2005**. A review of *Leptospermum scoparium* (Myrtaceae) in New Zealand. *New Zealand Journal of Botany* **43**: 431–449.

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

Thornburg RW. **2007**. Molecular biology of the *Nicotiana* floral nectary. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Springer, Dordrecht, 265–288.

Thrimawithana AH, Jones D, Hilario E, Grierson E, Ngo HM, Liachko I, Sullivan S, Bilton TP, Jacobs JME, Bicknell R, *et al.* 2019. A whole genome assembly of *Leptospermum scoparium* (Myrtaceae) for mānuka research. *New Zealand Journal of Crop and Horticultural Science* **47**: 233–260.

Trotter C, Tate K, Scott N, Townsend J, Wilde H, Lambie S, Marden M, Pinkney T. **2005**. Afforestation/reforestation of New Zealand marginal pasture lands by indigenous shrublands: the potential for Kyoto forest sinks. *Annals of forest science* **62**: 865–871.

Vannette RL, Fukami T. **2018**. Contrasting effects of yeasts and bacteria on floral nectar traits. *Annals of Botany* **121**: 1343–1349.

Wardle P. **1991**. *Vegetation of New Zealand*. Cambridge University Press.

Wicaksono WA, Jones EE, Monk J, Ridgway HJ. **2016**. The Bacterial Signature of Leptospermum scoparium (Mānuka) Reveals Core and Accessory Communities with Bioactive Properties. *Plos One* **11**: e0163717.

Wicaksono WA, Jones EE, Sansom CE, Perry NB, Monk J, Black A, Ridgway HJ. **2017**. Indigenous bacteria enhance growth and modify essential oil content in *Leptospermum scoparium* (mānuka). *New Zealand Journal of Botany* **55**: 306–317.

Williams S, King J, Revell M, Manley-Harris M, Balks M, Janusch F, Kiefer M, Clearwater M, Brooks P, Dawson M. **2014**. Regional, Annual, and Individual Variations in the Dihydroxyacetone Content of the Nectar of Mānuka (*Leptospermum scoparium*) in New Zealand. *Journal of Agricultural and Food Chemistry* **62**: 10332–10340.

Williams SD, Pappalardo L, Bishop J, Brooks PR. **2018**. Dihydroxyacetone Production in the Nectar of Australian *Leptospermum* Is Species Dependent. *Journal of Agriculture and Food Chemistry* **66**: 11133–11140.

Wilson PG, Heslewood MM. **2023**. Revised taxonomy of the tribe Leptospermeae (Myrtaceae) based on morphological and DNA data. *TAXON* **72**: 550–571.

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

Yin R, Mark AF, Wilson JB. **1984**. Aspects of the ecology of the indigenous shrub *Leptospermum scoparium* (Myrtaceae) in New Zealand. *New Zealand Journal of Botany* **22**: 483–5

Chapter 2: A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*)

Ella R.P. Grierson[1,2], Amali H. Thrimawithana[3], John W. van Klink[4], David H. Lewis[1], Ignacio Carvajal[3], Jason Shiller[5], Poppy Miller[5], Simon C. Deroles[3], Michael J. Clearwater[2], Kevin M. Davies[1], David Chagné[1*], Kathy E. Schwinn[1*]

1 The New Zealand Institute for Plant and Food Research Limited (PFR), Palmerston North, 4472, New Zealand

2 Te Aka Mātuatua - School of Science, University of Waikato, Hamilton 3216, New Zealand

3 PFR, Auckland, 1142, New Zealand

4 PFR, Dunedin, 9054, New Zealand

5 PFR, Te Puke, 3182, New Zealand

*Authors for correspondence:

kathy.schwinn@plantandfood.co.nz

david.chagne@plantandfood.co.nz

Paper published in New Phytologist: https://doi.org/10.1111/nph.19714

2.1 Summary

- Floral nectar composition beyond common sugars shows great diversity but contributing genetic factors are generally unknown. Mānuka (*Leptospermum scoparium*) is renowned for the antimicrobial compound methylglyoxal in its derived honey, which originates from the precursor, dihydroxyacetone (DHA), accumulating in the nectar. Although this nectar trait is highly variable, genetic contribution to the trait is unclear. Therefore, we investigated key gene(s) and genomic regions underpinning this trait.
- We used RNAseq analysis to identify nectary-associated genes differentially expressed between high and low nectar-DHA genotypes. We also used a mānuka high-density linkage map and quantitative trait loci (QTL) mapping population, supported by an improved genome assembly, to reveal genetic regions associated with nectar DHA content.
- Expression and QTL analyses both pointed to the involvement of a phosphatase gene, *LsSgpp2*. The expression pattern of *LsSgpp2* correlated with nectar DHA accumulation, and it co-located with a QTL on chromosome 4. The identification of three QTLs, some of the first reported for a plant nectar trait, indicates polygenic control of DHA content
- We have established plant genetics as a key influence on DHA accumulation. The data suggest the hypothesis of LsSGPP2 releasing DHA from DHA-phosphate and variability in *LsSgpp2* gene expression contributing to the trait variability.

Key words: dihydroxyacetone, *Leptospermum, Leptospermum scoparium,* mānuka, nectar, phosphor-sugar phosphatase, QTL

2.2 Introduction

Floral nectar is a reward secreted by specialised nectaries to attract pollinators or other mutualistic animals, with a strong correlation between nectar quantity and quality and the efficacy of the resulting plant-animal mutualism (Roy *et al.*, 2017; Nepi *et al*., 2018). Nectar biology thus underpins the economic performance of many crops, facilitating successful pollination and providing the raw material for honey production (Prasifka *et al.*, 2018). Floral nectar is principally composed of a variable combination of amino acids and glucose, fructose and sucrose, which provide carbon and nitrogen sources to feeding animals. However, many other compounds can be found at lower amounts in nectar, including different carbohydrates, organic acids, inorganic ions, proteins and specialised metabolites such as phenolics, alkaloids and terpenes. At least some of these have been shown to promote a positive interaction with the target organism while discouraging unwanted organisms, such as pathogens (Roy *et al.*, 2017; Nepi *et al.*, 2018).

Even though the great majority of angiosperms produce nectar, many aspects of the biology of nectaries and nectars are relatively poorly understood. Nectaries have evolved several times within the angiosperms leading to variation in their structure and underpinning molecular genetics (Bernardello, 2007; Slavković *et al.*, 2021). Additionally, species specific nectar traits have arisen that contribute to pollinator differentiation and pollination success. Yet studies on the molecular genetics of nectar production and secretion have focussed primarily on a few model species such as *Arabidopsis thaliana* (Kram & Carter, 2009), *Nicotiana* (Ren *et al.*, 2007; Silva *et al.*, 2018) and more recently squash (*Cucurbita pepo*) (Solhaug *et al.*, 2019). Recently, genes or enzymes associated with production of important specialised metabolites in nectar have also started to be characterised (Roy *et al.*, 2022; Magner *et al.*, 2023), but relatively few species have been studied in detail. Moreover, nectar composition can vary markedly even among species in a single genus. Thus, complimentary studies of a wider range of plant types are desirable.

Leptospermum (Myrtaceae) is a genus of woody perennials renowned for many of its species accumulating variable amounts of the triose dihydroxyacetone (DHA) in their nectar. The *Leptospermum* genus is concentrated in Australia and its species are found in a range of habitats, with *L. scoparium* (J.R. Forst et G. Forst) growing naturally in south-eastern Australia (Thompson, 1989). *L. scoparium* is also widespread throughout Aotearoa-NZ and is the only indigenous *Leptospermum* species (Koot *et al.*, 2022; Chagné *et al.*, 2023). Recent modelling suggests that within *L. scoparium*, two clades diverged ~9-12 million years ago and that Aotearoa-NZ *L. scoparium* (mānuka) is genetically distant from the Australia population and contains five geographically distinct gene pools (Koot *et al.*, 2022). Mānuka occupies a wide range of habitats and can tolerate a range of conditions, including low-nutrient soils, harsh environmental conditions, and high or low soil moisture (Stephens *et al.*, 2005; Derraik, 2008).

The mechanism by which DHA accumulates in *Leptospermum* nectar, and any associated physiological function, is unknown. In many plant species, nectaries do not have photosynthetic capacity, so nectar carbohydrates from photosynthesis elsewhere in the plant are translocated through the phloem (Pacini & Nepi, 2007). These sugars can be directly secreted into nectar (Ren *et al.*, 2007) but are often stored in nectaries as starch that is later degraded (Lin *et al.*, 2014). However, there are also some plant species, including mānuka, that have green, chloroplast-containing nectaries (Lüttge, 2013). Clearwater *et al.* (2021) recently demonstrated that the nectaries of mānuka photosynthesise and are a significant – although not the only – source of the fructose, glucose and trace amounts of sucrose that are characteristic of mānuka nectar. Furthermore, the

study also found that nectary photosynthesis was important for the accumulation of DHA in mānuka nectar.

Phosphorylated DHA (DHAP) is an intermediate of several primary metabolic pathways, including the Calvin cycle, glycolysis, and gluconeogenesis. Until recently *Leptospermum* was the only genus with species reported to accumulate DHA in their nectar, but it has recently been detected in two other species within the Myrtaceae (Obeng-Darko *et al.*, 2023). While at least 40 species across the genus *Leptospermum* have nectar DHA in various concentrations, only two, *L. scoparium* and *L. polygalifolium*, are currently commercially relevant to the *Leptospermum* honey industry (Williams *et al.*, 2018; Binks *et al.*, 2021). The presence of DHA in the nectar of mānuka has made it the basis of Aotearoa New Zealand's (Aotearoa-NZ) mānuka-dominated honey exports, valued at almost NZ$450M in 2022 (MPI, 2023). The nectar DHA converts to the antimicrobial compound methylglyoxal (MGO) in the derived honey, through a process that has been well studied (see Owens *et al*. 2019 and references therein). The nectar DHA content and subsequent MGO concentration underpins the antibacterial activity of the honey (Adams *et al.*, 2008; Mavric *et al.*, 2008), and determines the premium price the honey attracts.

Mānuka is highly polymorphic, owing to both genotypic and environmental effects (Cockayne, 1919; Yin *et al.*, 1984; Dawson, 2009). Variation has been observed in morphology of leaves, growth form, root anatomy, stem and foliage colour, leaf density and flower colour and size (Stephens *et al.*, 2005; Dawson, 2009). The DHA nectar trait and nectar characteristics in mānuka are also variable, affected by environmental factors and floral developmental stage (Clearwater *et al.*, 2018; Smallfield *et al.*, 2018; Noe *et al.*, 2019). Noe *et al.* (2019) found that mānuka nectar characteristics varied more among plants than among sites in a study across five sites in the North Island of Aotearoa-NZ. This study suggests that genetic factors influence the amount of nectar DHA. The importance of the genetics of the plant is further supported by the fact that not all mānuka plants accumulate DHA amounts useful for honey production. Individuals sourced from different regions in Aotearoa-NZ that are designated as 'high' nectar DHA accumulators retain their relative status when grown in another region (Hamilton *et al.*, 2013). Clarifying the role of genetics in the nectar DHA trait and understanding the key genetic factors will not only help answer questions on how and why *Leptospermum* has this unique nectar trait, but also will deliver knowledge that could assist the honey industry. There is a projected global shortage of high MGO honey, with existing mānuka stands unlikely to be sufficient to meet projected demand (McPherson, 2016).

We report here the use of two complementary approaches to find key gene(s) underpinning the nectar DHA trait in *Leptospermum*. The first was to identify genes with nectary-associated expression

that were differentially expressed between relatively high or low nectar DHA genotypes (H-DHA and L-DHA, respectively), and that encoded proteins that could have a predicted role in regulating DHA amounts from a primary metabolism or biochemical perspective. Of particular interest were transcripts encoding phosphatases, as it has been postulated that the first step in the process for DHA to accumulate in nectar would be dephosphorylation of DHAP (Clearwater *et al.*, 2021).

The second and unbiased approach used an improved version of the *L. scoparium* whole genome sequence reported by Thrimawithana *et al.* (2019) along with a mānuka high-density linkage map and quantitative loci (QTL) mapping population (Chagné *et al.*, 2019) to reveal genetic regions associated with the nectar DHA trait. Multiple QTLs were identified, each on separate chromosomes. One of these QTLs was found to contain the top candidate gene from the expression analysis, which is predicted to encode a phosphatase. The expression of this phosphatase was further characterised in relation to nectar DHA accumulation. This is the first report of a gene associated to this unique and commercially valuable nectar trait, which can now be studied further to establish its physiological role.

2.3 Materials and methods

2.3.1 Plant material

The lines used in each experiment are detailed in Supplementary Table **S2.1.**

Regional samples

To ensure a range of genotypes were represented, mānuka originating from six regions across Aotearoa-NZ were grown in a common garden design (all within approx. 500m^2) at Plant & Food Research (PFR) Palmerston North (Aotearoa-NZ). 'East Cape' plants were grown from seed collected from the East Cape region of the Gisborne District (described in Chagne *et al*., 2019). 'Reinga' and 'Kaipara' plants were sourced as young trees from Kauri Park Nursery (SH1 Kaiwaka, North Auckland) and originate from the corresponding districts of Northland. 'Waikato', 'Manawatū' and 'Otago' plants were grown from seed collected directly from the corresponding Aotearoa-NZ districts. Material from non-commercial sources was collected from private land with the permission of the owner. At the first year of sampling the 'East Cape' trees were approx. 10 years old. The 'Kaipara', 'Reinga', 'Waikato', 'Otago' and 'Manawatū' trees were sourced in 2014, prior to PFR establishing guidelines for sourcing samples from native species (Morgan *et al.,* 2019) and no prior informed consent was obtained from Māori guardians for this material. We recognise that consultation with the appropriate iwi about the use of material sourced by a commercial nursery in their ancestral boundaries would have been appropriate prior to the start of the project. We recommend for future

projects that such consultation should occur before the onset of the project. Initial nectar DHA sampling on at least 10 individual genotypes per region was followed with repeated sampling for up to six years on a smaller number of genotypes. Trees chosen for multiyear repeat sampling and gene expression analysis were those with consistently higher or lower nectar DHA.

Ornamental varieties

L. scoparium (J.R. Forst et G. Forst) cv. 'Martinii' was sourced from Annton Nursery (1/249 Peake Road, Hautapu, Tamahere), cv. 'Crimson Glory' as described in Thrimawithana *et al.* (2019) and cv. 'Electric Red' from Leafland (1 Roberts Line, Palmerston North).

Segregating population

The segregating population (EC201×EC103) was developed by crossing two genotypes from the 'East Cape' plants described above. Victor Goldsmith who worked for Ngāti Porou Miere Ltd and represented the Ngāti Porou iwi at the onset of this research approved the use of plant material from the EC201xEC103 segregating population that was developed from crossing parental trees collected within their ancestral boundaries. The population, genotyping by sequencing and linkage map construction are described in Chagné *et al.* (2019) and Thrimawithana *et al.* (2019). Nectars were sampled from these plants over three consecutive flowering seasons from 2019 until 2021, and Pearson correlation coefficients were calculated to assess correlation between sampling years.

2.3.2 Phenotyping

Nectar sampling

At least two branches per plant with at least 10 flowers per branch were covered with fine mesh drawstring bags approx. 24 hrs before collection. Branches with flowers were harvested each morning between 9 am and 10 am and stored (for up to 5 hrs) in plastic bags at 4°C until flowers were sampled for nectar. The floral developmental stage sampled was FDS6-7 according to Smallfield *et al.* (2018), which is defined as: stamens unfolding and anthers beginning to dehisce, nectar secretion occurring. Nectar sampling used a protocol modified from Morrant *et al.* (2009). For each flower, the hypanthium (where secreted nectar accumulates) was washed with a single 10µL aliquot of water by pipetting it on and off the hypanthium at least five times. Each sample was the combined nectar washes from 10 individual flowers. As a measure of total sugars, the °Brix value of each nectar wash sample was determined using a hand-held refractometer (ADE Advanced Optics, Oregon, USA).

Quantitative analyses for dihydroxyacetone (DHA)

This method, which measures DHA and MGO, was modified from that reported previously by Windsor *et al.* (2012). A subsample (20 µL) of nectar wash was combined with internal standard solution (ISS; 200 µL, 0.1 M citrate buffer pH 4, containing 0.1 µL mL^{-1} hydroxyacetone [HA]), pentafluorobenzylhydroxylamine (PFBHA; 30 µL, 10 mg mL^{-1}) in an Eppendorf tube and centrifuged (10,600 rcf, 5 min). The solution was incubated at room temperature for at least 60 min, MeCN (350 µL) added and centrifuged again prior to subsampling for HPLC analysis. Analyses were performed on an Agilent 1100 HPLC (Agilent Technologies) fitted with a diode array detector, using a Phenomenex Kinetex (100 × 3.0 mm RP-18, 2.6 µm) column and Phenomenex guard column (SecurityGuard 4 × 2 mm RP-18) at 30°C. Peaks were monitored at 263 nm. The mobile phase was CH3CN in H_2O, both containing formic acid (0.1%), with a linear programme: 30% MeCN at 0 min, 100% at 4 min, 30% at 5 min and 30% at 10 min. The flow rate was 0.5 mL min^{-1} with injection volumes of 5 µL. DHA was expressed as mg g^{-1} nectar sugar as estimated from the °Brix measurements.

2.3.3 'Crimson Glory' v2.0 genome and annotation

Genome assembly and scaffolding

A 'Crimson Glory' v2.0 genome was generated through a hybrid genome assembly approach, using a combination of Illumina short read data and Oxford Nanopore (ONT) long reads. High molecular weight DNA was extracted from the same plant as in Thrimawithana *et al.* (2019) and one flow cell (FLO-MIN106) was run using a minION (Mk1B) Oxford Nanopore sequencing device. The ONT data were base called using Guppy (version 6.4.6). In addition to the Illumina paired end (PE) short read data generated in Thrimawithana *et al.* (2019), two libraries of Illumina PE data were generated from the same plant with an insert size of 500bp. The PE short read data and ONT read data were then assembled using MaSuRCA assembler with 'FLYE_ASSEMBLY=0 (version 4.1.0). Hi-C reads described in Thrimawithana *et al.*, (2019) were mapped to the new assembly using BWA (-5SP option) (Li, 2013) and reads were flagged as PCR duplicates and removed using samblaster (Faust & Hall, 2014) and samtools (Li *et al.*, 2009). The contigs from this assembly were scaffolded using the information from the Hi-C data using YAHS (Version v1.2a.2) (Zhou *et al.*, 2023). Non-mānuka contigs from the scaffolded assembly were identified using Blobtools (Challis *et al.*, 2020). Contigs that did not blast to Streptophyta, had GC contents outside of 38%-42% or were less than 10,000 bp long were removed from the assembly. The larger scaffolds were named by synteny mapping to the *Eucalyptus grandis* (W. Hill) reference assembly (NCBI RefSeq assembly GCF_016545825.1), using a Ls (for *L. scoparium*) prefix. Synteny mapping was done using Minimap2 (Li, 2018) and visualised with Dgenies (Cabanettes & Klopp, 2018).

Gene prediction and annotation

Gene prediction was performed in two ways using BRAKER3 (version 3.0.3) (Gabriel *et al.*, 2023) and RNAseq data from multiple tissues described in Thrimawithana et al., (2019). Firstly, we produced training proteins using PASA (Haas *et al.*, 2008) from a de novo transcriptome assembly using Trinity (version 2.14.0) and secondly, we mapped the RNAseq data to the reference genome using STAR (version 2.5.3a) (Dobin *et al.*, 2013) to provide evidence for splice sites and coding regions. Functional annotation was assigned to predicted proteins using eggNOG mapper (Cantalapiedra *et al,* 2021) and by blastp against the TAIR 10 release of *Arabidopsis* (Reiser *et al.*, 2017), and the Swiss-Prot dataset from UniProt (Uniprot Consortium, 2015).

Genome data sovereignty

As a native species to Aotearoa-NZ, the ownership and rights to intellectual property and scientific data derived from mānuka raise concerns. Genomic data are a representation of peoples' biological material and are also considered taonga and a highly valuable strategic asset for Māori. Māori rights and full authority over natural resources are defined by the Treaty of Waitangi, a treaty signed in 1840 between Māori and the British Crown. Internationally, Māori ownership and control of natural resources is also supported by the UN Declaration on the Rights of Indigenous Peoples 2007 (UNDRIP, 2007) which Aotearoa-NZ endorsed in 2010. Considering the above, the mānuka genome assembly and predicted genes were deposited in a protected Aotearoa-NZ based repository. This seeks to ensure there is controlled access to genome data for taonga species from Aotearoa-NZ. The guardians (kaitiaki) of the genome data obtained from 'Crimson Glory' grown in the Manawatū region of Aotearoa-NZ are the local iwi (Rangitāne o Manawatū).

2.3.4 Determination of transcript abundance in high and low nectar DHA genotypes, and across a developmental series

Nectaries in mānuka flowers are indistinct and not readily delineated from the surrounding hypanthium, so the whole hypanthium was sampled. Hypanthium tissue was collected by a transverse cut between the ovary and hypanthium surface. The petals, sepals, stamens and stigma were removed. Top portions of the ovary locules were included, and any ovules within them were removed (Fig. **S2.1**).

Three mānuka genotypes producing higher amounts of DHA (H-DHA: 1N41V, 1N424 and 1N414) and three genotypes producing lower amounts of DHA (L-DHA: 1N40L, 1N40A and 1N3ZR) were chosen for RNAseq analysis. Two floral developmental stages (according to Smallfield *et al*., 2018) were targeted – FDS4, just prior to flower opening (pre-nectar secretion), and FDS6-7, nectar secretion

occurring (Fig. **S2.1**). Each genotype was sampled on three different days, with 20 hypanthia harvested per day. All 60 hypanthia for each genotype were pooled into one tissue sample for extraction and library preparation. Each mature hypanthium sample had 10 of the 20 hypanthia from which nectar was sampled.

The cv. 'Martinii' was used for a developmental series including seven stages. Hypanthia were collected the day before anthesis and on days zero, one, three, six, nine, twelve and fourteen post anthesis. Ten hypanthia from two to three plants were pooled for each sample, with each stage being sampled in triplicate as biological replication. Each replicate was taken across three to five days.

RNA was extracted from the hypanthium tissue based on the CTAB method from Gambino *et al.* (2008). RNAseq library preparation and sequencing was done by AGRF (Melbourne, Australia). Twelve Illumina stranded libraries were prepared (two developmental stages for each H-DHA genotype and L-DHA genotype) and run in one HiSeq lane of 100 bp paired-end reads, resulting in an average of 20 million reads per sample. The data was then quality checked with FastQC (version 0.11.2) and adapter trimmed using bbduck package of BBMap toolkit (version 38.33). Quality trimmed reads were then aligned to the newly improved mānuka genome ('Crimson Glory' v2.0) using STAR (version 2.7.10a) aligner with '--quantMode GeneCounts '. The resulting count files were modified to reflect HTSeq format using awk and sed one-liners and then used in differential expression analysis using DESeq2 (version 1.28.0) using R (version 4.0.0), where low counts reads were pre-filtered (at least three samples with a count of 10 or higher) and a surrogate variable was introduced to account for genotype differences between the samples. The DeSeq design formula contained a condition variable which accounted for both development stage and the DHA level. Furthermore, the nectar secretion occurring stage was selected for exploring the differentially expressed genes.

To amplify from transcripts in a range of species, RT-qPCR primers were designed from conserved regions to match all *Sgpp* (named in line with *AtSgpp* described in Caparrós-Martín *et al*., 2013) transcripts identified through RNAseq. For RT-qPCR, 1µg of RNA was used for cDNA synthesis. The cDNA synthesis, RT-qPCR reactions and conditions were as described in Lafferty *et al.* (2022). For each sample, 1µl of DNased RNA was diluted as a gDNA contamination control. The developmental series used *Sgpp2* specific primers and was run with the same reaction conditions on a BioRad CFX Maestro (5.2.008.0222). Relative transcript abundance for all experiments was calculated using the geometric mean of three reference genes – Actin, SAND and RPL3. Primer efficiencies were calculated using serial dilution (Primer pairs listed in Table **S2.2**).

2.3.5 Determination of transcript abundance in six tissue types

Six tissue types were harvested from each of three 'Crimson Glory' plants. Samples taken were stem and leaf from the previous year's growth, stem and leaf from the current year's new growth, roots and whole flowers. RNA was extracted from all tissues using the CTAB method as above. Stranded RNA libraries were prepared by Novogene using Ribo-zero ™ magnetic kit (Illumina, San Diego, California) and sequenced. The sequencing data were then mapped to 'Crimson Glory' v2.0, using STAR aligner (version 2.7.10a) with the '--quantMode GeneCounts 'and soft-clipping turned on '--clip5pNbases 15'.

2.3.6 QTL mapping of DHA content in nectar using a segregating mānuka population

The EC201×EC103 segregating population described in Chagné *et al.* (2019) was used for phenotypic assessment for DHA concentration over three flowering seasons (2019, 2020 and 2021). To account for environmental effects, weather information was taken from a weather station (Agent number 21963, network number E0536D, CliFlo, retrieved March 2021) approx. 800 m from the field location. Exploratory data analysis (plots, tables) and exploratory models (forward/ backward automatic variable selection using basic linear models) were used to select a subset of environmental variables to include in the final model. Brix values of <0.8 were excluded due to the inability to get an accurate measure of sugar content or samples with little or no nectar. Out of 181 genotypes, a total of 135 had nectar sampled, and 106 genotypes were included in the QTL analysis (with genotypes excluded that did not have genotyping data or had only one replicate for nectar). A linear mixed effects model was used to calculate Best Linear Unbiased Prediction (BLUP) values for each genotype (line) via the lmer function (lme4 package, Bates *et al.*, 2015). The genotype random effects (BLUPs) were iid (independent identically distributed) as all lines were equally related (full siblings). The model incorporated fixed effects for row, average relative humidity of day before, hours on sun of day of sampling and average night temperature of previous night. Inclusion of these covariates increased the conditional R2 from 0.43 to 0.49 (performance package, Lüdecke *et al*., 2021) indicating that the covariates explained some additional variation in log(DHA) values. Adjusting for variation explained by environmental factors enables the model to more accurately estimate additive genetic effects (BLUPs), which improves QTL mapping. The DHA data were log transformed to restrict estimated means and predicted values to be positive, and to stabilise the residual variance. Model checks were performed to assess for violations of the model assumptions, and no problematic departures were found.

In total, 185 individuals from the EC201×EC103 segregating population were screened using an Affymetrix Axiom™ (Thermo-Fisher Scientific) Single Nucleotide Polymorphism (SNP) array. The array

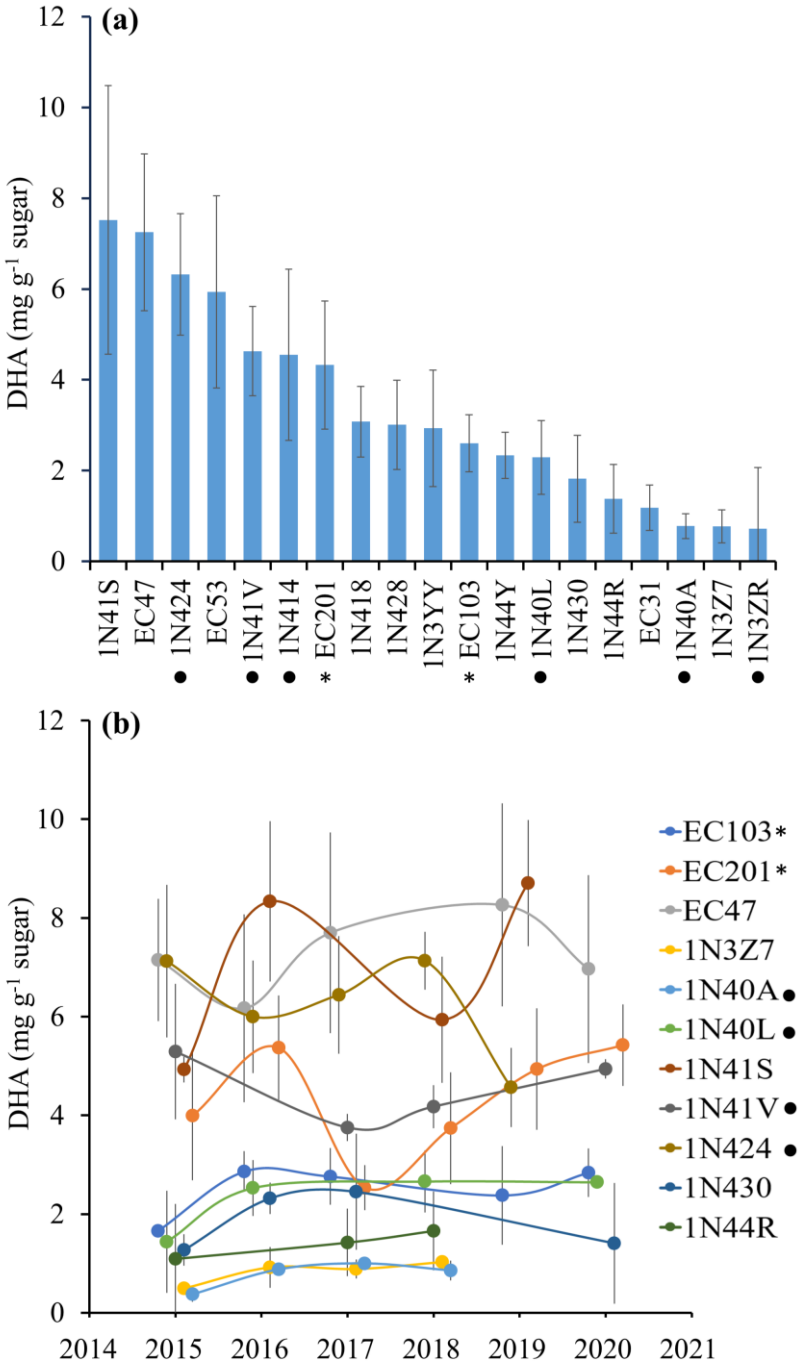


Figure 2.1: Dihydroxyacetone (DHA) nectar trait in a common garden design, for a range of mānuka (*Leptospermum scoparium)* genotypes across years. (a) DHA mg g^{-1} sugar in selected lines (n = 19), with at least 8 replicates across at least 3 seasons in total for each line. Parents of the segregating population are indicated with an asterisk, lines used for RNAseq indicated with a filled circle (b) DHA mg g^{-1} sugar in lines (n = 11) within and across seasons, for lines with at least two replicates per season and at least three seasons. Lines considered High-DHA were EC47, 1N424, 1N41V, 1N41S, and EC201. Lines considered Low-DHA were EC103, 1N40L, 1N430, 1N44R, 1N3Z7, and 1N40A. Values are the mean DHA mg g^{-1} sugar and error bars show the standard deviation. Data points have been offset and lines added linking them to assist readability. Each nectar sample consists of combined nectar washes from ten flowers.

containing 9,002 SNPs for mānuka was described by Montanari *et al.* (2023) and the genotyping was performed by Labogena (Jouy-en-Josas, France) using standard protocols. After genotyping, the SNP data were analysed using the Axiom Analysis Suite v5.1.1 software applying a dish QC threshold of 0.82 and a QC Call Rate > 95%. Visual inspection of SNP cluster plots was performed for SNPs classified as 'PolyHighResolution' (PHR) and 'CallRateBelowThreshold' (CRBT) by the Axiom Analysis Suite software, and SNP calls were manually curated. Genetic maps were constructed using the SNP array data using JoinMap v5.0 (www.kyazma.nl), based on the two-way pseudo-testcross strategy of Grattapaglia and Sederoff (1994), and SNP markers polymorphic for each parent. Linkage groups were determined with a logarithm of the odds (LOD) score of 6 or higher for grouping, and the Kosambi function was used to calculate map distances. The linkage groups were aligned to and numbered based on the new genome assembly, using a LG (linkage group) prefix. QTLs were detected using the Interval Mapping (IM) module in MapQTL v6.0 software (www.kyazma.nl), using BLUP values. A permutation test (1000 permutations) was performed to calculate the threshold to declare a QTL significant (genome-wide 90% LOD 3.3).

2.4 Results

2.4.1 Identification of mānuka genotypes with genetic variation for nectar DHA content

To determine that genetics contributes to DHA trait diversity, we measured nectar DHA content within and across flowering seasons for genotypes collected from across Aotearoa-NZ but grown in a single location (common garden). Up to a ten-fold difference in average nectar DHA content was observed between genotypes (7.5 versus 0.7 mg DHA g^{-1} sugar; Fig. **2.1a,** Fig. **S2.2**). Genotypes identified as H-DHA or L-DHA were chosen for repeat sampling for up to six years (Fig. **2.1b**). For most H-DHA and L-DHA genotypes, a difference in DHA was maintained across flowering seasons (Fig. **2.1b**). This indicated that nectar DHA content was strongly determined by genetic variation between individual plants.

Although most genotypes had consistently relatively high or relatively low average DHA content across seasons, significant variation of individual measurements was observed both within a season and between seasons for genotypes with relatively high DHA content (Fig. **2.1b**). This suggested environmental factors also influenced the higher nectar DHA amounts observed in some mānuka populations. As nectar samples were taken across the season, daily environmental factors that would influence nectar DHA would have varied.

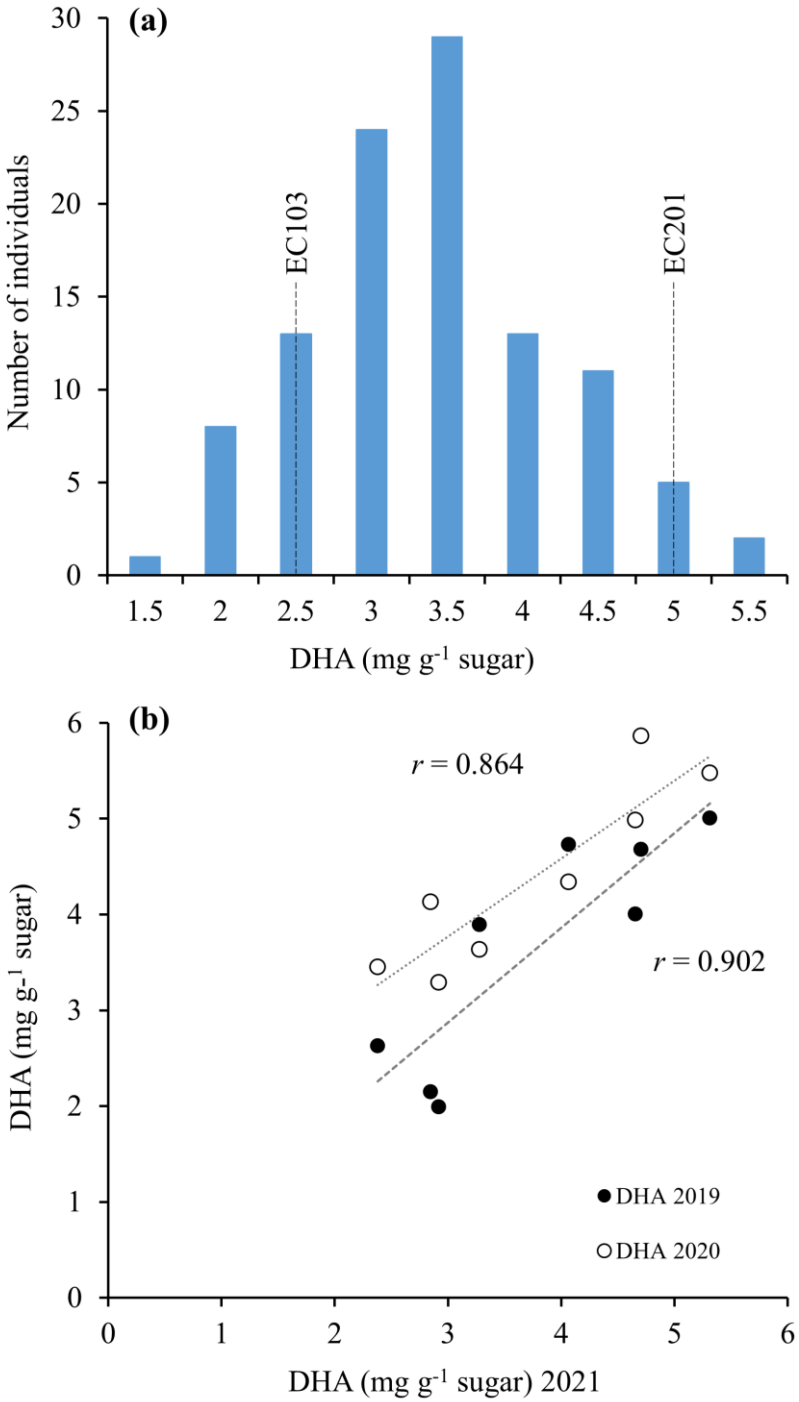


Figure 2.2: Nectar dihydroxyacetone (DHA) variation in a mānuka (*Leptospermum scoparium)* segregating population. (a) Phenotypic distribution of DHA in the EC201×EC103 segregating population in 2019 (average DHA concentration per individual). The dotted lines indicate the phenotype of the EC201 (4.9 mg g^{-1} sugar) and EC103 (2.4 mg g^{-1} sugar) parents. (b) Correlation of DHA concentration for the eight individuals (1N4HD, 1N4DW, 1N4BY, 1N4F0, 1N4CH, 1N4EE, 1N4H3, 1N4EB) from the EC201×EC103 segregating population that were phenotyped across three flowering seasons (2019 and 2020 on y axis against 2021 DHA values on x axis). Trendline fitted by linear regression, *r* values are the Pearson correlation coefficient.

We phenotyped a population, EC201×EC103 (Chagné *et al.*, 2019), to determine segregation of nectar DHA amounts. The data showed a normal distribution (Fig. **2.2a**), with values ranging from 1.7 to 5.4 mg DHA g^{-1} sugar. DHA phenotypes of the parents of the segregating population fell at either end of the range (EC103 = 2.4 mg g^{-1} and EC201 = 4.9 mg g^{-1}). A subset of the population was also sampled in the two subsequent years, and the DHA values correlated across years (Fig. **2.2b**; r = 0.86 and 0.90).

The range of DHA content and the stability of the phenotypes was considered sufficient to select members of the collection for candidate gene identification and to use the segregating population for QTL analysis.

2.4.2 Identification of a phosphatase gene differentially expressed between nectaries of high- and low-DHA genotypes

Approximately 35Gb of RNAseq data were generated from the 12 mānuka nectary samples, with a map back rate of 85% to the 'Crimson Glory' v2.0. DE-seq analysis was conducted on transcriptome data from nectary-enriched tissue of H-DHA and L-DHA genotypes at two developmental stages: pre-nectar secretion and active secretion (for volcano, PCA and MA plots see Fig. **S2.3**). As the RNA was from mixed tissue due to the nectary not having clear delineation, the DE-seq data was first checked for orthologues of nectary-associated genes to give confidence that analysis would be sensitive enough to capture nectary genes, even those with low transcript abundance. Three known nectary-associated genes, *SWEET9* (g16350.t1; Table **S2.3**; Lin *et al.*, 2014), *CRABS CLAW* (gene model g7327.1; Bowman & Smyth, 1999) and *Pin6* (gene model g2930.1; Bender *et al.*, 2013) had relatively high (transcript base mean 147,282), medium (3,190) and low (14) transcript abundance, respectively, indicating sufficient nectary tissue had been captured. SWEET9 is a sugar transporter required for nectar secretion in multiple species (Lin *et al.*, 2014; Roy *et al.*, 2017), and was included because of the presumed eccrine model of nectar secretion (Roy *et al.*, 2017) in mānuka (Clearwater *et al*., 2021).

Regarding differentially expressed genes between H-DHA and L-DHA genotypes (adjusted p-value cutoff of < 0.01 and log2Fold change > 1.0), 121 and 167 genes were identified in nectar secreting and pre-secreting stages, respectively (Supporting Dataset **S1**; Fig. **S2.3**). The standout gene in the dataset encodes a Haloacid dehalogenase-like hydrolase domain-containing protein. The HAD protein super family contains members with phosphatase activity. Transcript amount for the gene, subsequently named *Sgpp2* (described further below), was approximately doubled in H-DHA genotypes at both developmental stages (with adjusted p-values of $6.87E^{-10}$ for nectar secretion and $2.01E^{-08}$ for pre-secretion stages) (Table **S2.3;** Supporting Dataset **S1**). Reciprocal Blastp identified the

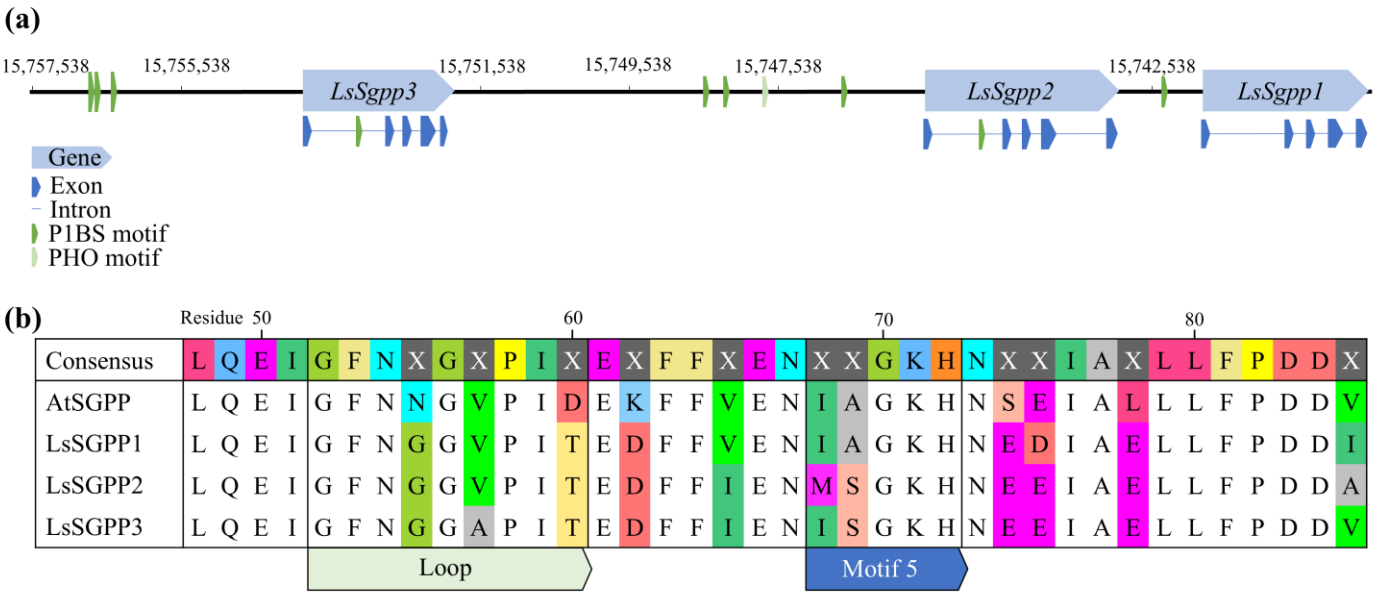


Figure 2.3: Three *Sgpp* genes identified in a complex locus. (a) The *Leptospermum scoparium Sgpp* locus from 15,757,538 to 15,739,538 on Chromosome 4, showing the three *Sgpp* genes, their introns and exons, and P1BS and PHO motifs. (b) Alignments of the residues from 48-85 of the three *L. scoparium* SGPPs, and the *Arabidopsis* SGPP (AT2G38740). Motif 5 (substrate specificity motif) and its preceding loop are indicated.

Arabidopsis HAD phosphatase gene *AtSgpp* (AT2G38740) as a putative homologue. AtSGPP has dephosphorylating activity on a range of pentose and hexose phosphor-sugars as well as on the triose glycerol 3-phosphate (although DHAP was not tested) (Caparrós-Martín *et al.*, 2013). As a phosphatase has been predicted to be required for releasing DHA from DHAP, thus allowing DHA to move into the nectar (Clearwater *et al*., 2021), if mānuka SGPP2 could use DHAP as a substrate it would directly affect DHA concentrations in the cell. Thus, we considered this gene to be the lead candidate from this dataset for the nectar trait.

The nectary-enriched transcriptome contained transcripts for two additional *Sgpp* genes closely related to *LsSgpp2*, termed *LsSgpp1* and *LsSgpp3.* All three mānuka genes reside in a complex locus located within a 15 kb region from 15,753,980 (*LsSgpp3* start codon) to 15,739,704 (*LsSgpp1* stop codon) on Chromosome 4 (Fig. **2.3a**). The *Sgpp* with the most similar sequence to *AtSgpp* was termed *Sgpp1*, and *Sgpp2* and *Sgpp3* were named for the sequence order in the locus. Of the three encoded proteins, SGPP1 is the closest match to AtSGPP, with 68% amino acid identity. SGPP1 is 84.5% identical to SGPP2 and 86.5% identical to SGPP3, and SGPP2 and SGPP3 are 94.3% identical to each other. Phylogenetic analysis using a subset of the *Arabidopsis* HAD family members showed the mānuka proteins cladded with AtSGPP, with strong bootstrap support (Fig. **S2.4**). All the predicted mānuka proteins have a similar loop sequence to AtSGPP that is involved in the protein recognising multiple substrates (Caparrós-Martín *et al.*, 2019) (Fig. **2.3b**). Both the mānuka proteins and AtSGPP were predicted to localise to the cytoplasm (PredictProtein web tool; Bernhofer *et al.,* 2021).

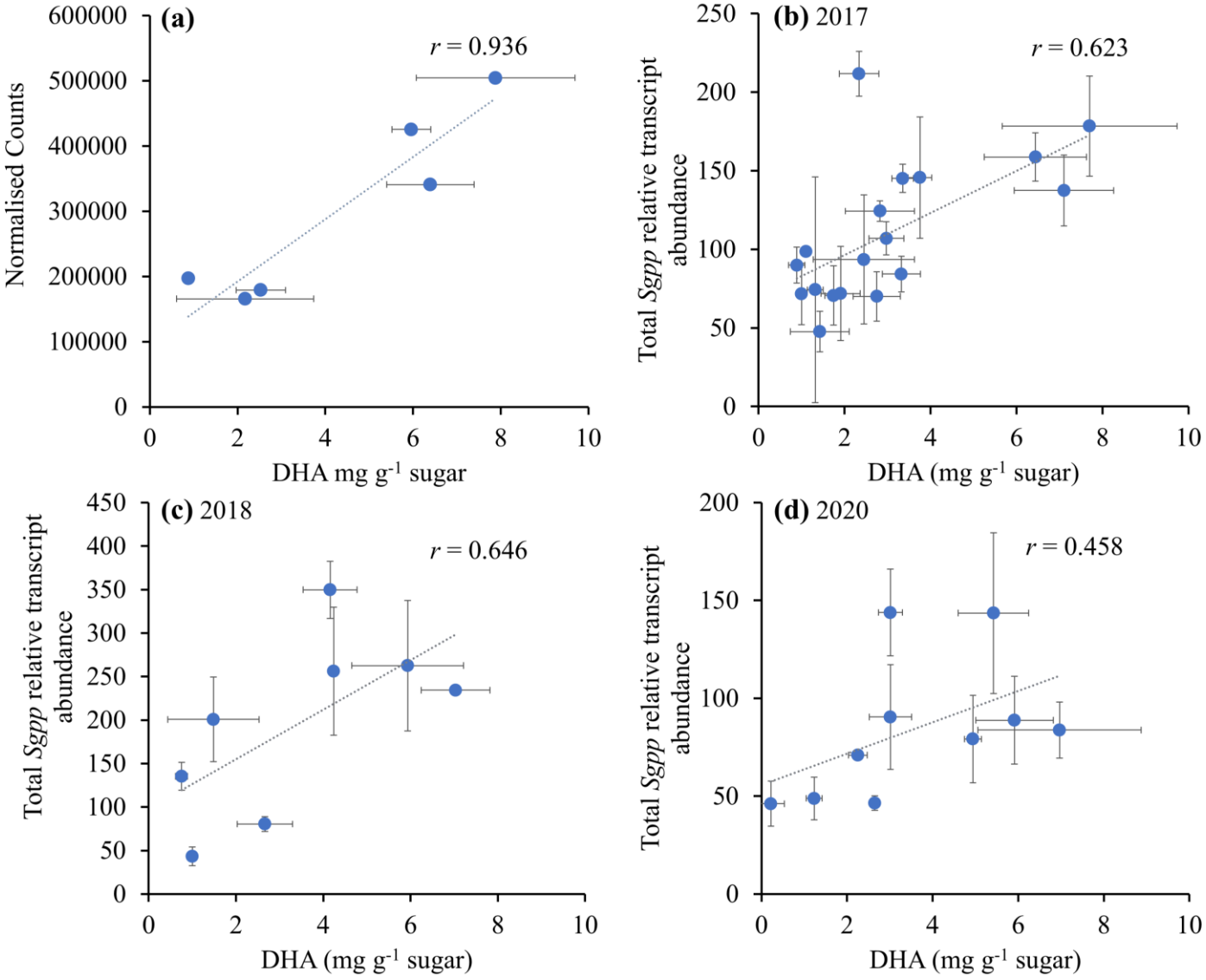


Figure 2.4: *Sgpp* expression in a range of *Leptospermum scoparium* genotypes. (a) *Sgpp2* normalised RNAseq counts from Low-DHA (N40A, N3ZR, and N40L) and High-DHA (N41V, N424 and N414) lines against DHA mg g^{-1} sugar in the corresponding samples at FDS 6-7. High-DHA average normalised counts were 382,724, and Low-DHA 209,122. (b) Total *Sgpp* expression (RT-qPCR) in 2017 with 18 genotypes (EC0103, EC031, EC038, EC047, EC201, EC49, EC53, N3Y7, N3YY, N3Z7, N40A, N418, N41V, N424, N430, N44R, N44Y, N457) (c) Total *Sgpp* expression (RT-qPCR) in 2018 with 8 genotypes (N3Z7, N40A, N40L, N414, N41S, N41V, N424, N44R). (d) Total *Sgpp* expression (RT-qPCR) in 2020 with 10 genotypes (EC103, EC201, EC31, EC37, EC47, EC53, Electric red, N40L, N41V, N428). Each genotype's nectar sample consists of combined nectar washes from ten flowers, each RNA sample consists of the hypanthium tissue of those same ten flowers. Error bars represent one standard deviation. Trendline fitted by linear regression, *r* values are the Pearson correlation coefficient.

2.4.3 Transcript abundance for *Sgpp2* correlates with nectar DHA content

Expression patterns of the three *Sgpp* genes were further examined regarding a possible function in DHA production. RNAseq normalised transcript abundance of *Sgpp2* correlated with the average DHA amounts in the nectar samples (r = 0.94, Fig. **2.4a**). In contrast neither *Sgpp1* nor *Sgpp3* were strongly expressed and no correlation with nectar DHA was found (LsSgpp1.t1 - log_2Fold change in high-DHA genotypes of 0.119, adjusted p-value of 0.933; LsSgpp3.t1, log_2Fold change in high-DHA genotypes of 0.02, adjusted p-value of 0.962, Fig. **2.5a**, Fig. **S2.5**). Of total *Sgpp* transcript abundance in the mature nectary-enriched tissue samples, an average of 99.46% was *Sgpp2* (based on normalised RNAseq read counts), with negligible contribution from *Sgpp1* or *Sgpp3* (0.25% and 0.28%, respectively). A similar ratio, with *Sgpp2* dominant, was also found in RNA extracted from whole 'Crimson Glory' flowers (Fig. **2.5a**).

Using RT-qPCR, *Sgpp* gene expression was measured in nectary-enriched tissue of a wider range of genotypes across three seasons, and *Sgpp2* gene expression again correlated with DHA accumulation (Fig. **2.4b, c, d**). DNA primers that effectively distinguished between the transcripts from the three genes were not able to be designed at that stage. However, given *Sgpp1* and *Sgpp3* have negligible expression in the flower, the data represented mostly *Sgpp2* (Fig. **2.5a**). Further evidence of this is the low transcript abundance of these genes using *Sgpp* gene specific primers for RT-qPCR in 2020 (Fig. **S2.5**). When *Sgpp2*-specific transcript abundance was measured across seven developmental stages of the flower, a peak was found at 3 days after flower opening (Fig. **2.5b**), which is just prior to the period that shows the highest DHA accumulation (Clearwater *et al*., 2021).

2.4.4 An improved complete chromosomal scale assembly of *L. scoparium* 'Crimson Glory'

An unbiased QTL approach was used to complement the RNAseq differential transcript analysis. A whole genome sequence for the *L. scoparium* cv. 'Crimson Glory' was previously published (v1.0, Thrimawithana *et al*., 2019), but to facilitate the QTL analysis, we developed an improved version ('Crimson Glory' v2.0). A hybrid *de novo* genome sequence was assembled using 6Gb of Oxford Nanopore sequencing and 70 Gb of Illumina PE short read data. A genome assembly of 61 scaffolds larger than 100kb was produced and spanned 294,018,319 bp in total (Table **S2.4**). The percentage of complete BUSCOs and the LAI score were 97.2% (Fig. **S2.6**) and 10.75, respectively, indicating higher completeness compared to the 'Crimson Glory' v1.0 genome. The contig ordering was also more accurate, based on improved collinearity with both the *E. grandis* genome and the EC201xEC103 linkage map, and one contig was a complete telomere to telomere assembled chromosome. Hi-C and

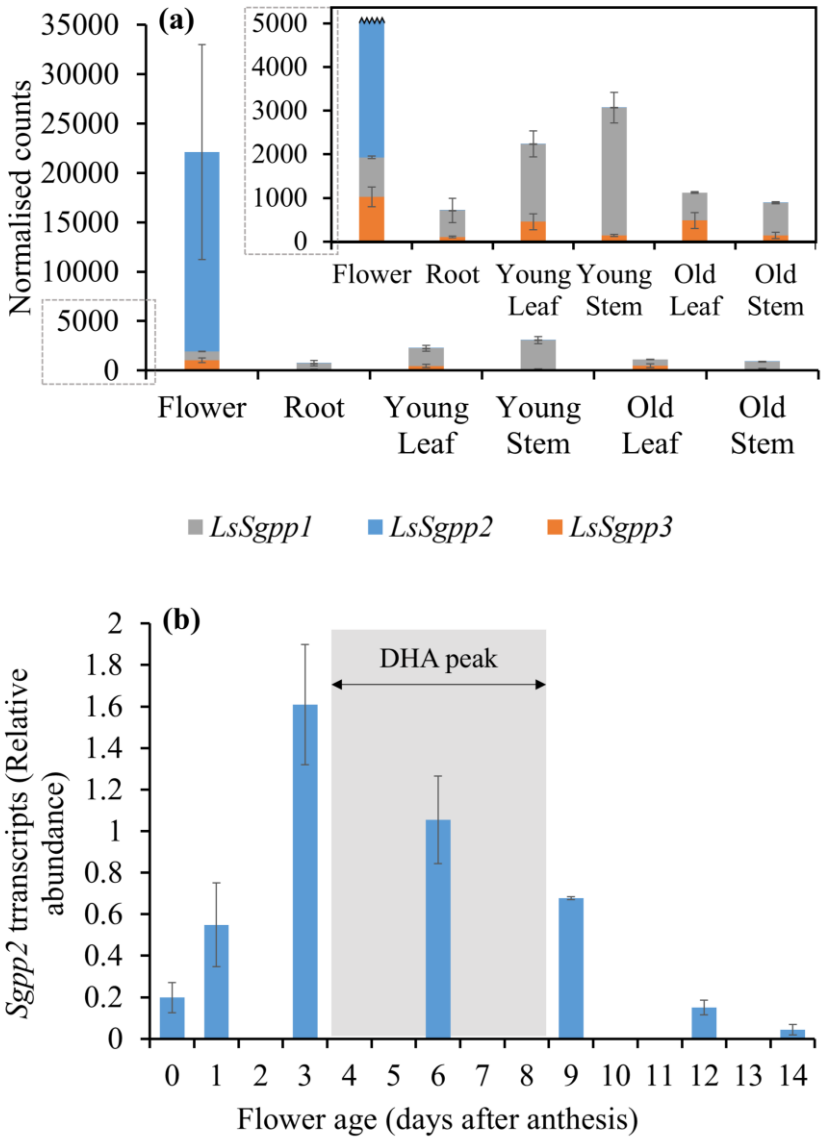


Figure 2.5: *Sgpp* expression in different tissues and developmental stages of *Leptospermum scoparium*. (a) *Sgpp1, Sgpp2* and *Sgpp3* RNAseq transcript counts in 'Crimson Glory' tissues – flowers, mature and young leaves, mature and young stems, and roots. The grey box indicates the area of the axis that is represented in the inset. (b) Relative expression (RT-qPCR) of *Sgpp2* in a developmental series of 'Martinii' hypanthium. The shaded area indicates the days after the flower opens that DHA is peaking (days 4-8 based on Figure 2b in Clearwater *et al.* 2021; FDS6 and 7 in Smallfield *et al.* 2018). Error bars represent one standard deviation.

synteny with *E. grandis* (NCBI RefSeq assembly GCF_016545825.1) was used to scaffold and orientate the new assembly into 11 pseudo-chromosomes ('Crimson Glory' v2.0) (Fig. **S2.7** and **S2.8**). In total 30,256 gene models were predicted in the 'Crimson Glory' v2.0 assembly. The *Sgpp* locus in 'Crimson Glory' v1.0 genome was not well resolved, and only contained *Sgpp1* and part of *Sgpp3*, but the 'Crimson Glory' v2.0 genome enabled full resolution of all three *Sgpp* genes.

2.4.5 Identification of QTLs for DHA content in nectar

QTL analysis for the DHA trait was performed using a segregating population of mānuka phenotyped over multiple years and genotyped with SNPs anchored to the new near-complete genome assembly (Supporting Dataset **S2**). In total, 766 SNP markers were polymorphic in the EC201×EC103 segregating population and were used for linkage map construction. Two parental maps of 12 linkage groups (LG) syntenic to the 'Crimson Glory' genome were obtained, with 254 and 303 markers totalling 996 cM and 767 cM for EC201 and EC103, respectively. LG8 and LG5 were split into two groups for EC201 and EC103, respectively. QTLs were identified for nectar DHA concentration on LG9 and LG10 of the EC201 parent, and LG4, LG9 and LG10 of the EC103 parent (Fig. **2.6a** and Table **2.1**). The chromosome 9 and 10 QTLs segregate from both parents of the population, while the chromosome 4 QTL is inherited from only one of the parents, indicating that the high DHA allele may be fixed/homozygous in the other parent. The confidence intervals of the three QTLs in the 'Crimson Glory' genome assembly ranged from approximately 0 to 7.5 Mb, 0 to 8 Mb and 11 to 17 Mb on chromosomes Ls4, Ls9 and Ls10, respectively.

2.4.6 *Sgpp2* is in a QTL for nectar DHA content

After the QTL intervals were defined, the genomic location of *Sgpp2* was determined – between positions 15,745,584 and 15,742,961 bp on the new 'Crimson Glory' v2.0 chromosome Ls4, co-locating with the LG4 QTL. Of the 586 genes in the interval that had >10 read counts in at least three RNA samples (out of 834 total genes), at a stringent adjusted *p*-value cut off at < 0.001, *Sgpp2* was the only gene significantly differentially expressed at the nectar secretion stage, and one of only three genes differentially expressed at the pre-nectar secretion stage. The other two genes at the pre-nectar secretion stage encode a protein of unknown function (gene model g18842.t1) and LEUCOANTHOCYANIDIN DIOXYGENASE/ANTHOCYANIDIN SYNTHASE (ANS, gene model g18659.t1). ANS is part of the biosynthetic pathway of anthocyanins, the pigments putatively responsible for the red colour of the mānuka hypanthium, and unlikely to be associated with the DHA trait. At a relaxed adjusted *p*-value cutoff at < 0.01, there were an additional four differentially expressed genes, none of which have an annotation indicating clear potential in influencing the DHA trait (Supporting Dataset **S1)**.

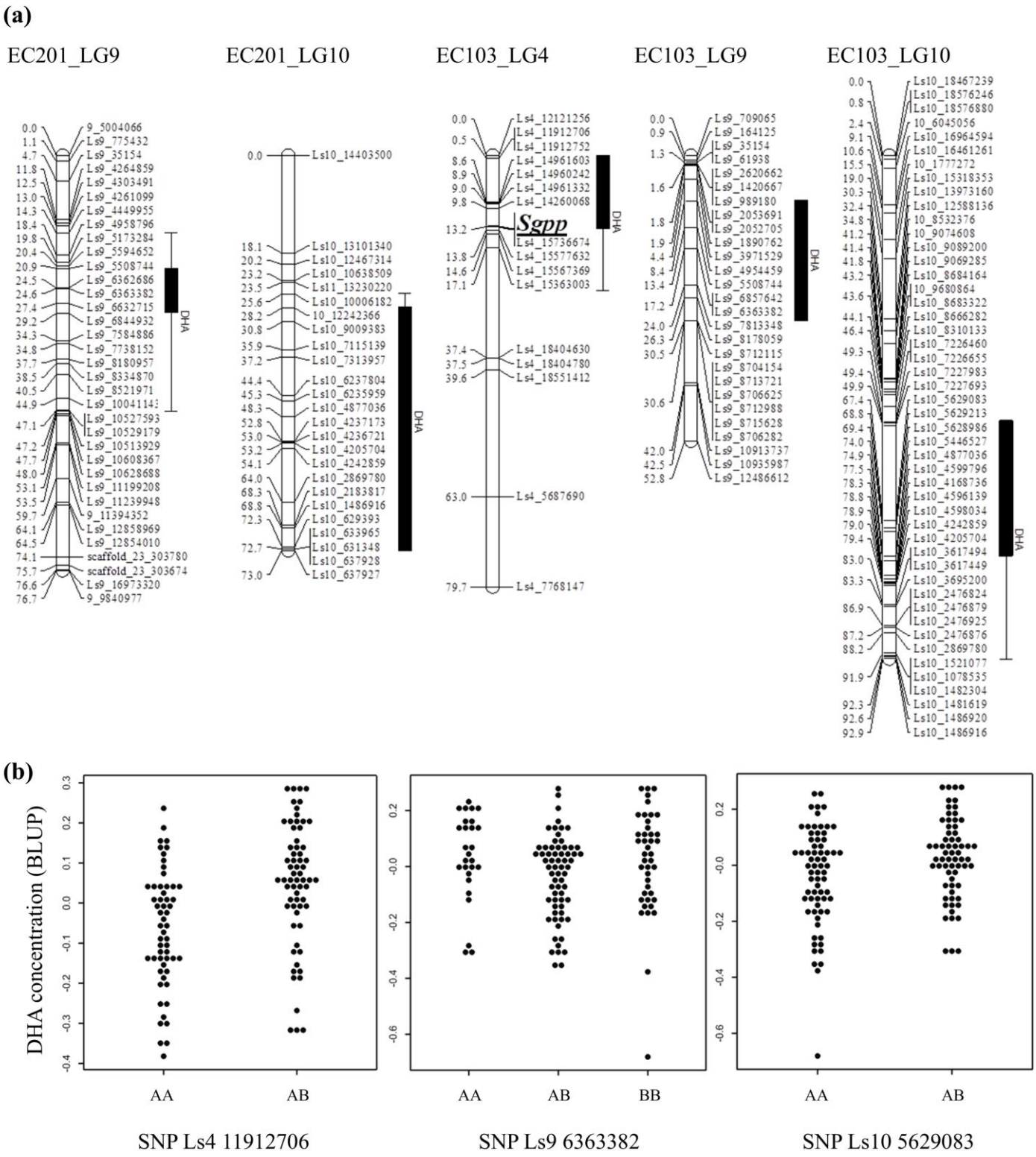


Figure 2.6: Quantitative Trait Locus (QTL) for nectar dihydroxyacetone (DHA) concentration in the *Leptospermum scoparium* EC201×EC103 segregating population. (a) Linkage maps of EC201 and EC103 showing the locations of QTLs for DHA concentration calculated using Best Linear Unbiased Prediction (BLUP) values. A linear mixed effects model was used to calculate the BLUP values for each genotype to more accurately estimate the genetic effects for improving QTL mapping. The markers' locations on the linkage map are indicated in centiMorgans (left) and based on their location in the 'Crimson Glory' v2.0 genome assembly (right). (b) Dot histogram of the phenotypes (BLUP) of segregating individuals against the SNP markers with the highest LOD score on linkage group 4, 9 and 10 located at positions 11,912,706 bp (EC103 = AB, EC201 = AA), 6,363,382 bp (both EC103 and EC201 = AB) and 5,629,083 bp (EC103 = AB, EC201 = AA) of 'Crimson Glory' v2.0 chromosome 4, 9 and 10 (Ls4, Ls9 and Ls10), respectively. Ls4 is linked to the *Sgpp* locus.

Table 2.1: Quantitative Trait Loci (QTL) for DHA concentration in the *Leptospermum scoparium* EC201×EC103 segregating population. QTL locations are calculated based on BLUP values and their locations are presented according to the new 'Crimson Glory' v2.0 genome assembly. The marker with the highest LOD score are heterozygous for the parent indicated in the column 'Parental map'.

L. scoparium chromosome	Marker with highest LOD position (bp)	LOD score	% variance explained	QTL peak location and interval (cM)	Parental map
Ls4	11912706	3.53	12.7	0.5 (0– 25)	EC103
Ls9	6363382	3.88	13.9	17.2 (8– 30.6)	EC103
Ls10	5629083	4.09	16.3	67.4 (49– 93)	EC103
Ls10	7115139	3.86	21.5	35.9 (25.5– 73)	EC201
Ls9	6363382	3.58	12.8	24.6 (14.3– 47)	EC201

Candidate genes that could influence DHA content in nectar have not yet been identified for the QTLs on LG9 and LG10, which contain 1,781 and 1,112 predicted gene models, respectively. Of these, a total of nine and five genes in LG9 and LG10, respectively were significantly differentially expressed (adjusted $p < 0.01$ and log2Fold change > 1.0, Supporting Dataset **S1)**.

2.5 Discussion

2.5.1 Plant genetics is a key influence on nectar DHA accumulation

Understanding how nectar is formed and why it varies in composition between and within a species is a question receiving increasing research attention (Roy *et al.*, 2017; Nepi *et al.*, 2018; Prasifka *et al.*, 2018; Slavković *et al.*, 2021). However, although progress has been made on identifying the genetic basis for variation of nectar volume (Galliot *et al.*, 2006; Wessinger *et al.*, 2014; Barstow *et al.*, 2022) or sucrose as a percentage of total sugars (Prasifka *et al.*, 2018), such knowledge is limited for the other nectar components. Enzymes and/or candidate genes (Roy *et al.*, 2022; Magner *et al.*, 2023) and a QTL (Kostyun *et al.*, 2019) have been identified responsible for nectar colour in a few species, but data on the control of production of other metabolites is lacking, perhaps because of the potential difficulties in obtaining sufficient phenotypic and genomic data. In this study, by growing mānuka plants from across Aotearoa-NZ in a common environment and combining nectar screening with genetic sequence analysis of a population segregating for nectar DHA content, facilitated by a new genome sequence assembly, we were able to identify QTLs associated with this notable nectar trait. The presence of QTLs demonstrated, for the first time, that the concentration of DHA in the nectar of *L. scoparium* (mānuka) is under genetic control, in addition to defining genomic regions for assisting with identification of associated causal genes.

The accumulation of DHA in nectar of mānuka and some other species of *Leptospermum* results in honey which is highly valued both as a food and medicinal product. The economic importance of mānuka honey has resulted in it being not only one of the only woody perennial species for which nectary biology and nectar traits have been studied in detail, but also one of the few species with green, photosynthetic nectaries to have been characterised. Such studies have found that mānuka nectar traits, including DHA content, can be highly variable (Smallfield *et al.*, 2018; Noe *et al.*, 2019; Clearwater *et al.*, 2021), making it challenging in previous work to separate the relative influence of plant genetics from environment on nectar composition (Noe *et al.*, 2019).

2.5.2 A phosphatase gene candidate for control of nectar DHA production

Two complementary approaches, unbiased QTL discovery and differential gene expression analyses on nectaries, independently identified a candidate gene associated with high-DHA production in mānuka nectar: *Sgpp2*. *LsSgpp2* putatively encodes a phosphatase, based on sequence similarity to the characterised *Arabidopsis* gene *AtSgpp*, which dephosphorylates a range of phosphor-sugar substrates, including a triose (Caparrós-Martín *et al.*, 2013). *LsSgpp2* expression was essentially restricted to the flower (Fig. **2.5a**), and in nectary-enriched tissue it had the highest transcript count in the transcriptome, higher than the candidate orthologue for *SWEET9* – a gene known to be highly expressed in nectaries (Lin *et al*., 2014, Table **S3**). Also in the top 10 most highly expressed genes in the dataset were candidate cell wall invertase and sucrose synthase sequences, which could reflect the highly active carbohydrate metabolism pathways typical of nectaries (Table **S3**).

The complex locus that contains *LsSgpp1*, *LsSgpp2* and *LsSgpp3* lies within one of the three QTLs found to contribute to the nectar DHA phenotype. *LsSgpp2* transcript abundance correlated with DHA accumulation, although we did not determine if this was reflected in a corresponding change in protein abundance or enzyme activity.

Based on the RNAseq analysis, *LsSgpp2* was the only gene in the Ls4 QTL interval with significant differential transcript abundance between high- and low-DHA genotypes (adjusted $p < 0.001$). The expression characteristics of *LsSppg2* and the putative function of the encoded protein, which would directly affect the amount of free DHA in the nectary, make it the strongest candidate for the causal gene underlying the Ls4 QTL for the DHA trait. However, it has not been ruled out that other genes in the interval are responsible for, or co-contribute, to the phenotype. Almost 600 genes expressed in the nectary-enriched tissue were present in the Ls4 QTL interval, with six of these differentially expressed (Supporting Dataset **S1**). Additionally, fourteen differentially expressed genes were identified in the LG9 and LG10 QTL intervals. Further research would be required to draw conclusions

on any possible roles these differentially expressed genes have in contributing to this complex multifactorial trait.

2.5.3 A proposed role for SGPP2

AtSgpp expression is responsive to abiotic stresses, with inorganic phosphate (Pi) starvation causing one of the largest inductions of expression (Caparrós-Martín *et al.*, 2013). In wider studies of HAD genes, they have been found to be upregulated in response to Pi stress (Sega *et al.*, 2021) and contain a high frequency of Pi responsive *cis* elements in their promoters (Du *et al.*, 2021). The physiological function of AtSGPP is thought to be housekeeping detoxification, modulation of sugar-phosphate balance and Pi homeostasis (Caparrós-Martín *et al.*, 2013; He *et al.*, 2019). Based on analyses of putative regulatory motifs, the mānuka *Sgpp* locus, including *Sgpp2*, is predicted to also be responsive to changes in Pi status. The *LsSgpp* locus contains multiple P1BS motifs (GNATATNC), which are the target element for the transcription factor PHOSPHATE STARVATION RESPONSE1 (PHR1; Rubio *et al.*, 2001). In *Arabidopsis,* PHR1 has an important role in Pi starvation responses including adaptation to high light stress and preserving photosynthesis (Müller *et al.*, 2007; Bustos *et al.*, 2010; Nilsson *et al.*, 2012). In particular, *LsSgpp2* contains high affinity variants of the P1BS sequence (GAATATTC, Ruan *et al*., 2015) in both its promoter and first intron, which could enhance transcription (e.g., Aleksza *et al*., 2017).

In a proteome study, AtSGPP has been found to be associated with the plasma membrane along with hundreds of other proteins, including other primary metabolism proteins such as the glycolysis- and gluconeogenesis-related fructose-bisphosphate aldolase and cytosolic triose-phosphate isomerase (Marmagne *et al.*, 2007), both of which directly affect pools of DHAP in the cytoplasm owing to their roles in using and making DHAP. If LsSGPP2, which has a similar *in silico* predicted localisation as AtSGPP, is also in the plasma membrane then it would be appropriately located for facilitating movement of DHA into nectar.

Based on our results and the data on AtSGPP function, we propose a model in which LsSGPP2 dephosphorylates DHAP, allowing DHA to move freely out of the nectaries into the nectar (Fig. **2.7**). The proposed activity of LsSGPP2 may be related to a role in Pi homeostasis during sugar production and secretion by the nectary. Nectary cells have rapid sugar phosphate turnover and undergo periods of general oxidative stress as well as osmotic stress during nectar secretion (Nicholson & Thornburg 2007). AtSGPP is associated with stress tolerance (Caparrós-Martín *et al.*, 2013), and it is possible that LsSGPP2 assists in maintaining appropriate Pi concentrations in different cellular compartments during metabolic flux in a photosynthesising nectary (Tegeder & Weber, 2006; Nilsson *et al.*, 2012).

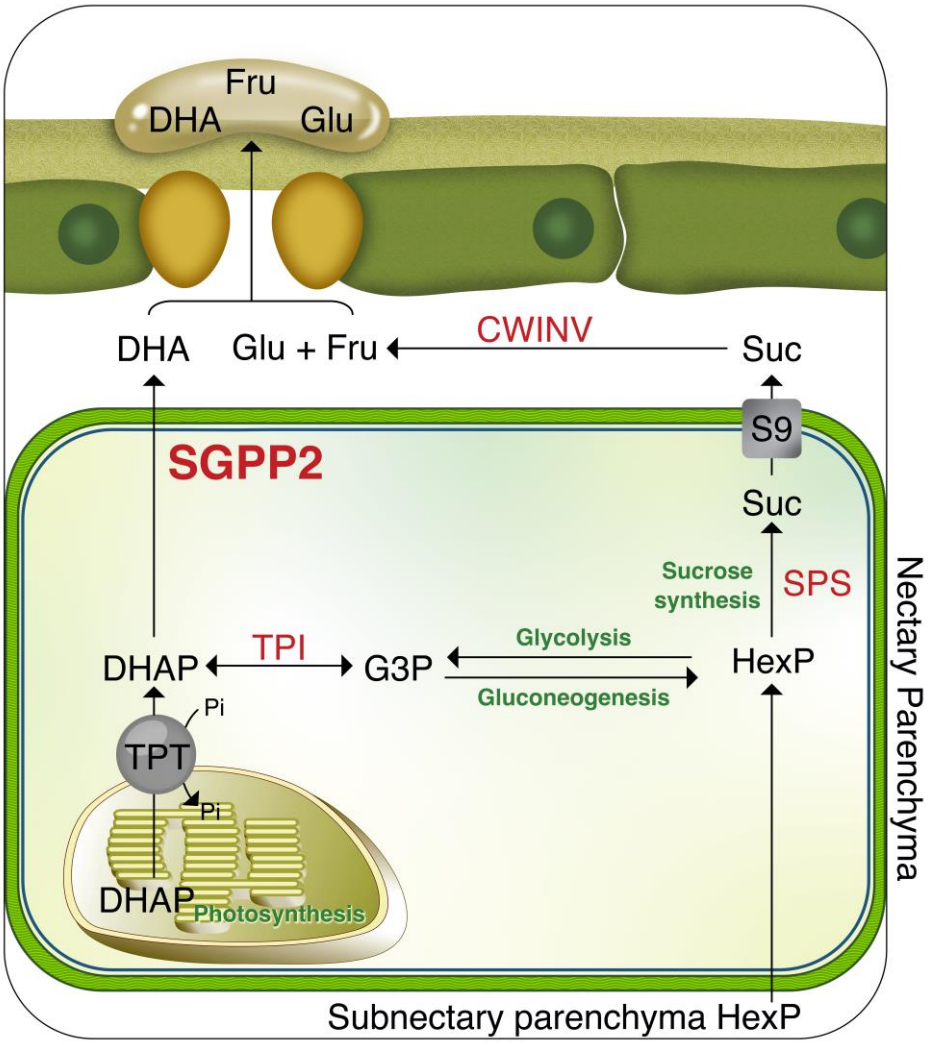


Figure 2.7: Integration of SGPP2 phosphatase into proposed mānuka (*Leptospermum scoparium)* nectar production model. (Developed from the proposed nectar production model in Clearwater *et al.* 2021) In this model, the nectary Hexose phosphates (HexP) enter through subnectary parenchyma as well as from photosynthesis from nectary chloroplasts (*via* dihydroxyacetone-phosphate (DHAP) and gluconeogenesis). DHAP is exported from the nectary chloroplasts via the chloroplastic Triose Phosphate Translocator (TPT). Some of this exported DHAP is dephosphorylated by SGPP2 to form dihydroxyacetone (DHA), which then diffuses into the apoplast and into the nectar. As LsSGPP2 has a similar *in silico* predicted location to AtSGPP, it could be associated with the plasma membrane.

A possible explanation for the evolution of the unique trait of DHA accumulation could relate to *Leptospermum*'s tolerance to nutrient-poor soils. *Leptospermum* is the main genus reported to have high levels of DHA in the nectar of some species (Williams *et al.*, 2018), and it is centred in Australia which has a high proportion of nutrient poor soils (Orians & Milewski, 2007). The Australian flora also has many unique features thought to have evolved in response to nutrient poverty, including some related to nectar traits (Orians & Milewski, 2007). Mānuka often grows on infertile sites that are low in phosphorous (Stephens *et al.*, 2005), and exhibits highly plastic root architecture in response to variable soil nutrient concentrations (Gutiérrez-Ginés *et al.*, 2019; Burge *et al.*, 2021). Mānuka also maintains rapid growth with comparatively low requirements for key nutrients (Gutiérrez-Ginés *et al.*, 2019), behaviour that suggests efficient strategies to maintain sufficient Pi in its tissues. Perhaps the accumulation of DHA in nectar is a consequence of evolutionary changes for tolerance of nutrient poor soils, rather than a response to pollination-related evolutionary drivers acting directly on nectar composition.

2.5.4 Concluding comments

The data confirm genetic control of DHA content in mānuka nectar and identify a candidate gene for the biochemical mechanism of DHA accumulation. Further analysis can now be used to confirm whether the LsSGPP2 enzyme does indeed release DHA from DHAP, as proposed. Understanding the genetic basis of nectar chemistry opens opportunities for breeding for nectar traits, which could contribute to increasing production of high-value mānuka honey through plantations of improved plant material. As a treasured species to Māori and with a large proportion of mānuka growing on indigenous-owned ancestral land, such breeding efforts would require co-innovation and benefit sharing with Māori. The QTLs identified may be used for more efficient breeding using marker-assisted selection, while DNA variants unique to plants growing on ancestral lands will preserve the authenticity and uniqueness of indigenous food products, while maintaining the resilience of local ecosystems. The 'Crimson Glory' v2.0 assembly will facilitate ongoing analysis of the genetic basis of other traits in mānuka, including those valuable to both the honey industry, such as flowering time and nectar volume, and plant conservation, such as genes for resistance to *Austropuccinia psidii* (myrtle rust) and adaptation to climate change. Improving our mānuka genomic resources and furthering our understanding of mānuka nectar production can contribute to wider nectar research, specifically for species with green photosynthesising nectaries. Further work can establish whether the same locus structure and gene expression patterns are found in other *Leptospermum* and Myrtaceae species, and whether the proposed model of DHA accumulation is retained across the Myrtaceae.

Acknowledgements

We thank Ian King and Julie Ryan for propagation and care of plants; Steve Arathoon, Chethi Waite, Sarah Moss, Elena Hilario and Chris Kirk for lab assistance; Sylvester Obeng-Darko for tissue harvest for the developmental series; Bruce Smallfield and Ed Morgan for seed collection; Tony Corbett for figure illustration; and Catrin Guenther and Simona Nardozza for useful comments on the manuscript. The EC103, EC201 and other EC lines were obtained from a collaboration with Horouta Mānuka Company Ltd (Te Tairāwhiti). This research was funded by a Plant & Food Research 'Discovery Science' project [New Zealand Ministry for Business Innovation and Employment (MBIE), Science Strategic Investment Funding (SSIF)], and MBIE Smart Idea funding (C11x2102) and MBIE SSIF Platform Genomics Aotearoa.

Author Contributions

KES and DC conceived the project. KES, DC, KD and ERPG planned and designed the research. ERPG, AHT, JWvK and DHL performed the experiments. ERPG, AHT, SCD, DC, PM, KES and MC analysed and/or interpreted the data. AHT, DC, IC and JS produced v2.0 of the genome. ERPG, KES, KMD and DC wrote the manuscript with contributions from the other authors. All authors read and approved the manuscript.

Competing interests

None to declare.

Data accessibility statement

The genome and transcriptome for mānuka are considered taonga (i.e. a treasured resource to Māori) and therefore it is hosted in an Aotearoa-New Zealand-based data repository (https://data.agdr.org.nz/). The 'Crimson Glory' v2.0 genome and transcript data are accessible to researchers, provided that their use benefits Māori and respects their intellectual properties, traditional knowledge and cultural values. Researchers can access that data by filling an access request form located on the AGDR which is then evaluated by the guardians of the 'Crimson Glory' v2.0 genome.

Supporting datasets

Can be accessed online at: https://doi.org/10.1111/nph.19714

2.6 References

Adams CJ, Boult CH, Deadman BJ, Farr JM, Grainger MN, Manley-Harris M, Snow MJ. 2008. Isolation by HPLC and characterisation of the bioactive fraction of New Zealand manuka (*Leptospermum scoparium*) honey. *Carbohydrate Research* **343**: 651–659.

Aleksza D, Horváth GV, Sándor G, Szabados L. 2017. Proline accumulation is regulated by transcription factors associated with phosphate starvation. *Plant Physiology* **175**: 555–567.

Barstow AC, Prasifka JR, Attia Z, Kane NC, Hulke BS. 2022. Genetic mapping of a pollinator preference trait: Nectar volume in sunflower (*Helianthus annuus* L.). *Frontiers in Plant Science* **13**: 1056278.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**: 1–48.

Bender RL, Fekete ML, Klinkenberg PM, Hampton M, Bauer B, Malecha M, Lindgren K, J AM, Perera MA, Nikolau BJ*, et al.* 2013. PIN6 is required for nectary auxin response and short stamen development. *Plant Journal* **74**: 893–904.

Bernardello G. 2007. A systematic survey of floral nectaries. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*, Springer: Dordrecht, Netherlands, pp. 19–128.

Bernhofer M, Dallago C, Karl T, Satagopam V, Heinzinger M, Littmann M, Olenyi T, Qiu J, Schütze K, Yachdav G. 2021. PredictProtein – predicting protein structure and function for 29 years. *Nucleic Acids Research* **49**: W535–W540.

Binks RM, Heslewood M, Wilson PG, Byrne M. 2021. Phylogenomic analysis confirms polyphyly of *Leptospermum* and delineates five major clades that warrant generic recognition *Taxon* **71**: 348–359.

Bowman JL, Smyth DR. 1999. *CRABS CLAW*, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **126**: 2387–2396.

Burge OR, R. CB, Eger A, Fitzgerald N. 2021. Wetland plant foliage nutrients as indicators of soil nutrients. *Manaaki Whenua – Landcare Research Contract Report LC3955*.

Bustos R, Castrillo G, Linhares F, Puga MI, Rubio V, Pérez-Pérez J, Solano R, Leyva A, Paz-Ares J. 2010. A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in *Arabidopsis*. *PLoS Genetics* **6**: e1001102.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**: e4958.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**: 5825 –5829.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. 2013. HAD hydrolase function unveiled by substrate screening: enzymatic characterization of *Arabidopsis thaliana* subclass I phosphosugar phosphatase AtSgpp. *Planta* **237**: 943–954.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. 2019. Sequence determinants of substrate ambiguity in a HAD phosphosugar phosphatase of *Arabidopsis thaliana*. *Biology* **8**: 77.

Chagné D, Montanari S, Kirk C, Mitchell C, Heenan P, Koot E. 2023. Single nucleotide polymorphism analysis in *Leptospermum scoparium* (Myrtaceae) supports two highly differentiated endemic species in Aotearoa New Zealand and Australia. *Tree Genetics & Genomes* **19**: 31.

Chagné D, Ryan J, Saeed M, Van Stijn T, Brauning R, Clarke S, Jacobs J, Wilcox P, Boursault E, Jaksons P*, et al.* 2019. A high density linkage map and quantitative trait loci for tree growth for New Zealand mānuka (*Leptospermum scoparium*). *New Zealand Journal of Crop and Horticultural Science* **47**: 261–272.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit - Interactive Quality Assessment of Genome Assemblies. *G3 (Bethesda)* **10**: 1361–1374.

Clearwater MJ, Noe ST, Manley-Harris M, Truman GL, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. 2021. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

Clearwater MJ, Revell M, Noe S, Manley-Harris M. 2018. Influence of genotype, floral stage, and water stress on floral nectar yield and composition of manuka (*Leptospermum scoparium*). *Annals of Botany*. **121:** 501–512.

CliFlo: NIWA's (National Institute of Water and Atmospheric Research) National Climate Database on the Web. http://cliflo.niwa.co.nz/. Retrieved March 2021.

Cockayne L. 1919. New Zealand plants and their story: Government Printer, South Africa.

Dawson M. 2009. A history of *Leptospermum scoparium* in cultivation: Discoveries from the wild. *New Zealand Garden Journal* **12**: 21–25.

Derraik J. 2008. New Zealand manuka (*Leptospermum scoparium*; Myrtaceae): a brief account of its natural history and human perceptions. *New Zealand Garden Journal*. **11:** 4–8.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Du Z, Deng S, Wu Z, Wang C. 2021. Genome-wide analysis of haloacid dehalogenase genes reveals their function in phosphate starvation responses in rice. *Plos One* **16**: e0245600.

Faust GG, Hall IM. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**: 2503–2505.

Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2023. BRAKER3: Fully Automated Genome Annotation Using RNA-Seq and Protein Evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *BioRxiv*. [Preprint]. 2023 Nov 27:2023.06.10.544449.

Galliot C, Hoballah ME, Kuhlemeier C, Stuurman J. 2006. Genetics of flower size and nectar volume in *Petunia* pollination syndromes. *Planta* **225**: 203–212.

Gambino G, Perrone I, Gribaudo I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* **19**: 520–525.

Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**: 1121–1137.

Gutiérrez-Ginés MJ, Madejón E, Lehto NJ, McLenaghen RD, Horswell J, Dickinson N, Robinson BH. 2019. Response of a pioneering species (*Leptospermum scoparium* J.R.Forst. & G.Forst.) to Heterogeneity in a Low-Fertility Soil. *Frontiers in Plant Science* **10**: 93.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**: R7.

Hamilton G, Millner J, Robertson A, Stephens J. 2013. Assessment of manuka provenances for production of high 'unique manuka factor' honey. *Agronomy New Zealand* **43**: 139–144.

He JZ, Dorion S, Lacroix M, Rivoal J. 2019. Sustained substrate cycles between hexose phosphates and free sugars in phosphate-deficient potato (*Solanum tuberosum*) cell cultures. *Planta* **249**: 1319–1336.

Koot E, Arnst E, Taane M, Goldsmith K, Thrimawithana A, Reihana K, González-Martínez SC, Goldsmith V, Houliston G, Chagné D. 2022. Genome-wide patterns of genetic diversity, population structure and demographic history in mānuka (*Leptospermum scoparium*) growing on indigenous Māori land. *Horticulture Research* **9**: uhab012.

Kostyun JL, Gibson MJS, King CM, Moyle LC. 2019. A simple genetic architecture and low constraint allow rapid floral evolution in a diverse and recently radiating plant genus. *New Phytologist* **223:** 1009–1022.

Kram BW, Carter CJ. 2009. *Arabidopsis thaliana* as a model for functional nectary analysis. *Sexual Plant Reproduction* **22**: 235–246.

Lafferty DJ, Espley RV, Deng CH, Günther CS, Plunkett B, Turner JL, Jaakola L, Karppinen K, Allan AC, Albert NW. 2022. Hierarchical regulation of MYBPA1 by anthocyanin- and proanthocyanidin-related MYB proteins is conserved in *Vaccinium* species. *Journal of Experimental Botany* **73**: 1344–1356.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lin IW, Sosso D, Chen LQ, Gase K, Kim SG, Kessler D, Klinkenberg PM, Gorder MK, Hou BH, Qu XQ, *et al.* 2014. Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**: 546–549.

Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* **6**: 3139.

Lüttge U. 2013. Green nectaries: the role of photosynthesis in secretion. *Botanical Journal of the Linnean Society* **173**: 1–11.

Magner ET, Roy R, Freund Saxhaug K, Zambre A, Bruns K, Snell-Rood EC, Hampton M, Hegeman AD, Carter CJ. 2023. Post-secretory synthesis of a natural analog of iron-gall ink in the black nectar of *Melianthus* spp. *New Phytologist* **239**: 2026–2040.

Marmagne A, Ferro M, Meinnel T, Bruley C, Kuhn L, Garin J, Barbier-Brygoo H, Ephritikhine G. 2007. A high content in lipid-modified peripheral proteins and integral receptor kinases features in the *Arabidopsis* plasma membrane proteome. *Molecular & Cellular Proteomics* **6**: 1980–1996.

Mavric E, Wittmann S, Barth G, Henle T. 2008. Identification and quantification of methylglyoxal as the dominant antibacterial constituent of Manuka (*Leptospermum scoparium*) honeys from New Zealand. *Molecular Nutrition & Food Research* **52**: 483–489.

McPherson AJ. 2016. Manuka—A viable alternative land use for New Zealand's hill country. *New Zealand Journal of Forestry* **61**: 11–19.

Montanari S, Deng C, Koot E, Bassil NV, Zurn JD, Morrison-Whittle P, Worthington ML, Aryal R, Ashrafi H, Pradelles J*, et al.* 2023. A multiplexed plant–animal SNP array for selective breeding and species conservation applications. *G3 Genes|Genomes|Genetics* **10:** jkad170.

Morgan ER, Perry NB, Chagné D. 2019. Science at the intersection of cultures – Māori, Pākehā and mānuka. *New Zealand Journal of Crop and Horticultural Science* **47**: 225–232.

Morrant DS, Schumann R, Petit S. 2009. Field methods for sampling and storing nectar from flowers with low nectar volumes. *Annals of Botany* **103**: 533–542.

MPI (Ministry for Primary Industries). 2023 *New Zealand Honey Exports* URL https://www.mpi.govt.nz/dmsdocument/42360-New-Zealand-honey-exports. [Accessed 1 October 2023]

Müller R, Morant M, Jarmer H, Nilsson L, Nielsen TH. 2007. Genome-Wide Analysis of the *Arabidopsis* leaf transcriptome reveals interaction of phosphate and sugar metabolism. *Plant Physiology* **143**: 156–171.

Nicholson S, Thornburg R. 2007. Nectar Chemistry. In: Nicholson SW, Nepi, M Pacini E, eds, *Nectaries and nectar.* Dordrecht: Springer Netherlands 215–264.

Nilsson L, Lundmark M, Jensen PE, Nielsen TH. 2012. The *Arabidopsis* transcription factor PHR1 is essential for adaptation to high light and retaining functional photosynthesis during phosphate starvation. *Physiologia Plantarum* **144**: 35–47.

Nepi M, Grasso DA and Mancuso S. 2018. Nectar in plant–insect mutualistic relationships: From food reward to partner manipulation. *Frontiers in Plant Science* **9**: 1063.

Noe S, Manley-Harris M, Clearwater MJ. 2019. Floral nectar of wild mānuka (*Leptospermum scoparium*) varies more among plants than among sites. *New Zealand Journal of Crop and Horticultural Science* **47**: 282–296.

Obeng-Darko SA, Sloan J, Binks RM, Brooks PR, Veneklaas EJ, Finnegan PM. 2023. Dihydroxyacetone in the floral nectar of *Ericomyrtus serpyllifolia* (Turcz.) Rye (Myrtaceae) and *Verticordia chrysantha* Endl. (Myrtaceae) demonstrates that this precursor to bioactive honey is not **r**estricted to the **g**enus *Leptospermum* (Myrtaceae). *Journal of Agricultural and Food Chemistry* **71:** 7703–7709.

Orians GH, Milewski AV. 2007. Ecology of Australia: the effects of nutrient-poor soils and intense fires. *Biological Reviews* **82**: 393–423.

Owens A, Lane JR, Manley-Harris M, Jensen AM, Jørgensen S. 2019. Kinetics of conversion of dihydroxyacetone to methylglyoxal in New Zealand mānuka honey: Part V – The rate determining step. *Food Chemistry* **276**: 636–642.

Pacini E, Nepi M. 2007. Nectar production and presentation. In: Nicholson SW, Nepi, M Pacini E, eds, *Nectaries and nectar.* Dordrecht: Springer Netherlands 167–214.

Prasifka JR, Mallinger RE, Portlas ZM, Hulke BS, Fugate KK, Paradis T, Hampton ME, Carter CJ. 2018. Using nectar-related traits to enhance crop-pollinator interactions. *Frontiers in Plant Science* **9**: 812.

Reiser L, Subramaniam S, Li D, Huala E. 2017. Using the *Arabidopsis* Information Resource (TAIR) to Find Information About *Arabidopsis* Genes. *Current Protocols in Bioinformatics* **60**: 1.11.11-11.11.45.

Ren G, Healy RA, Klyne AM, Horner HT, James MG, Thornburg RW. 2007. Transient starch metabolism in ornamental tobacco floral nectaries regulates nectar composition and release. *Plant Science* **173**: 277–290.

Roy R, Schmitt AJ, Thomas JB, Carter CJ. 2017. Review: Nectar biology: From molecules to ecosystems. *Plant Science* **262**: 148–164.

Roy R, Moreno N, Brockman SA, Kostanecki A, Zambre A, Holl C, Solhaug EM, Minami A, Snell-Rood EC, Hampton M, Bee MA, *et al.* 2022. Convergent evolution of a blood-red nectar pigment in vertebrate-pollinated flowers. *Proceedings of the National Acadamy of Sciences USA* 1**19:** e2114420119**.**

Ruan W, Guo M, Cai L, Hu H, Li C, Liu Y, Wu Z, Mao C, Yi K, Wu P*, et al.* 2015. Genetic manipulation of a high-affinity PHR1 target *cis*-element to improve phosphorous uptake in *Oryza sativa* L. *Plant Molecular Biology* **87**: 429–440.

Rubio V, Linhares F, Solano R, Martín AC, Iglesias J, Leyva A, Paz-Ares J. 2001. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes & Development* **15**: 2122–2133.

Sega P, Kruszka K, Bielewicz D, Karlowski W, Nuc P, Szweykowska-Kulinska Z, Pacak A. 2021. Pi-starvation induced transcriptional changes in barley revealed by a comprehensive RNA-Seq and degradome analyses. *BMC Genomics* **22**: 165.

Silva FA, Guirgis A, Thornburg R. 2018. Nectar analysis throughout the genus *Nicotiana* Suggests conserved mechanisms of nectar production and biochemical action. *Frontiers in Plant Science* **9**: 1100.

Smallfield BM, Joyce NI, van Klink JW. 2018. Developmental and compositional changes in *Leptospermum scoparium* nectar and their relevance to mānuka honey bioactives and markers. *New Zealand Journal of Botany* **56**: 183–197.

Solhaug EM, Roy R, Chatt EC, Klinkenberg PM, Mohd-Fadzil N-A, Hampton M, Nikolau BJ, Carter CJ. 2019. An integrated transcriptomics and metabolomics analysis of the *Cucurbita pepo* nectary implicates key modules of primary metabolism involved in nectar synthesis and secretion. *Plant Direct* **3**: e00120.

Stephens J, Molan P, Clarkson B. 2005. A review of *Leptospermum scoparium* (Myrtaceae) in New Zealand. *New Zealand Journal of Botany* **43**: 431–449.

Slavković F, Dogimont C, Morin H, Boualem A, Bendahmane A. 2021. The genetic control of nectary development. *Trends in Plant Science* **26**: 260–271.

Tegeder M, Weber APM. 2006. Metabolite transporters in the control of plant primary metabolism. *Annual Plant Reviews **22**:* 85–120.

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

Thrimawithana AH, Jones D, Hilario E, Grierson E, Ngo HM, Liachko I, Sullivan S, Bilton TP, Jacobs JME, Bicknell R*, et al.* 2019. A whole genome assembly of *Leptospermum scoparium* (Myrtaceae) for mānuka research. *New Zealand Journal of Crop and Horticultural Science* **47**: 233–260.

Uniprot Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Research* **43**: D204–D212.

UNDRIP – United Nations Declaration on the Rights of Indigenous Peoples. 2007. https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html. Accessed October 2023

Wessinger CA, Hileman LC, Rausher MD. 2014. Identification of major quantitative trait loci underlying floral pollination syndrome divergence in *Penstemon*. *Philosophical Transactions of the Royal Society B* **369**: 20130349.

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.

Windsor S, Pappalardo M, Brooks P, Williams S, Manley-Harris M. 2012. A convenient new analysis of dihydroxyacetone and methylglyoxal applied to Australian *Leptospermum* honeys. *Journal of Pharmacognosy and Phytotherapy* **4**: 6–11.

Yin R, Mark AF, Wilson JB. 1984. Aspects of the ecology of the indigenous shrub *Leptospermum scoparium* (Myrtaceae) in New Zealand. *New Zealand Journal of Botany* **22**: 483–507.

Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**: btac808

Chapter 3: Genome and transcriptome assemblies of Australian tea trees *Leptospermum morrisonii* and *Gaudium laevigatum*.

David Chagné[1*], Cecilia Deng[2], **Ella Grierson[1]**

[1]The New Zealand Institute for Plant and Food Research Ltd (Plant & Food Research), Food Innovation Research Centre, Batchelar Road, Palmerston North 4472, New Zealand (Aotearoa)

[2]Plant & Food Research, Mt Albert Research Centre, Private Bag 92169, Auckland Mail Centre, Auckland 1142, New Zealand (Aotearoa)

[*]Author for correspondence: David.Chagne@plantandfood.co.nz

3.1 Abstract

Leptospermeae species (Myrtaceae) are shrubs that are often termed as tea trees in Australia and that form the basis of honey and essential oil industries. We produced high-quality haplotyped genome assemblies of *Gaudium laevigatum* and *Leptospermum morrisonii* using high-fidelity long-read sequencing and synteny-based scaffolding against the *Leptospermum scoparium* genome. The 11 pseudo-chromosomes of *G. laevigatum* haplotypes 1 and 2, and *L. morrisonii* haplotypes 1 and 2 covered 265 Mb, 255 Mb, 341 Mb and 322 Mb, respectively. These chromosome-scale assemblies were near complete, with complete BUSCOs of 98.3% and 97.9% for *G. laevigatum* and 97.1% and 95.6% for *L. morrisonii,* and eight to seven contigs achieving telomere to telomere for *G. laevigatum* and *L. morrisonii*, respectively. In total, 33,663 and 31,035 genes were predicted in *L. morrisonii* hap1 and hap2, respectively, whereas *G. laevigatum*'s haplotypes had 30,563 and 30,422 genes. Of these genes, 18,121 orthologue groups were shared between both species and *L. scoparium*. These new genome assemblies provide an important resource for studying traits related to nectar and foliage chemistry, and for understanding the genetic diversity of the species.

Keywords: mānuka, Myrtaceae, synteny; genome sequencing; Leptospermeae

3.2 Introduction

Leptospermeae is a tribe of the Myrtaceae family consisting of shrubby trees that are often termed as tea trees, due to European settlers to Australia using their leaves to make herbal tea. Taxa in the

Leptospermeae tribe include species belonging to the *Leptospermum*, *Gaudium* and *Kunzea* genera. They bear flowers with five petals that are white, pink or red, and the fruit is a woody capsule. Tea trees are hardy and can grow in moist soils, making Leptospermeae successful colonising species. There are approximately 80 species of Leptospermeae, however the taxonomy in the genus is often debated (Wilson and Heslewood, 2023). Most species of Leptospermeae are endemic to South-East Australia (mainland and Tasmania), however one species can be found in Aotearoa-New Zealand (*Leptospermum scoparium*) and one in Malaysia and Indonesia (*L. recurvum*). In addition to their ecological importance as colonising species, some *Leptospermum* species have economic value. While the honey produced from *L. scoparium* in Aotearoa-New Zealand is termed "mānuka honey" due to its traditional and cultural connection to Māori, other species of *Leptospermum* form the basis for Australian honeys. For example, the honey produced from *Leptospermum* species is often known as jelly bush or lemon-scented tea tree honey. In Australia, *Leptospermum* species are also the basis of a significant essential oil industry.

The *Leptospermum* genome is diploid with 11 base chromosomes (2n = 2x = 22) (Thrimawithana *et al., 2019)*, although some polyploid forms have been described (Bicknell *et al.,* 2019). The genome is relatively compact (~300 million base pairs; Mb) compared to other plant species. Recent sequencing of the *L. scoparium* genome (Thrimawithana *et al.,* 2019, Grierson *et al.,* 2024, Chapter 2) provided a resource for *Leptospermum* biology, in particular for understanding the genetic control of traits relevant to the honey and oil industries (Grierson *et al.,* 2024, Chapter 2), and for understanding the genetic diversity and population structure within the species (Chagné *et al.,* 2023; Koot *et al.,* 2022).

Here we used high-fidelity long-read PacBio sequencing to produce high-quality, near complete, haplotyped genome assemblies of Australian species *Gaudium laevigatum* and *Leptospermum morrisonii*. Synteny analysis between both species and the genome of Aotearoa-New Zealand mānuka (*L. scoparium*) was performed test for building chromosome-scale scaffolds. Transcriptome sequencing data was used for predicting gene models and transposable elements. Orthologous protein analysis was performed for identifying protein-coding genes unique to each species and haplotypes.

3.3 Material and methods

3.3.1 Plant material and DNA extraction

Gaudium laevigatum is known as the Australian coastal tea tree. The species' natural range is from Southwest Victoria to Southeast New South Wales and Northern Tasmania. It is naturalised in Aotearoa-New Zealand and other parts of Australia, where it is considered a weed. Specimens of *G.*

laevigatum were sourced from plants found in the coastal dunes near to Whanganui, Aotearoa-New Zealand (39°56'24.2"S 174°58'44.0"E).

Leptospermum morrisonii is endemic to New South Wales. The species is used by honeybees to produce honey. The red foliage ornamental variety 'Copper Sheen' was used for this study, using an individual tree growing at the Plant & Food Research Palmerston North site (40°22'35.6"S 175°36'52.8"E).

Young expanding leaves were collected for both species. High molecular weight nuclear DNA was purified as described by Hilario (2018).

Hypanthium tissue was collected from *G. laevigatum* and *L. morrisonii* flowers at approximately floral developmental stage 6-7 (Smallfield *et al.,* 2018) as described in Grierson *et al.* (2024, Chapter 2), however the ovules within the top portion of locules were not removed. Three *G. laevigatum* hypanthium samples and four *L. morrisonii* hypanthium samples were taken, and each sample consisted of 10 hypanthia. RNA was extracted from the hypanthium tissue based on the CTAB method from Gambino *et al.* (2008).

3.3.2 Transcriptome sequencing

Seven stranded Illumina mRNA libraries were prepared by Novogene (Hong Kong) from the *G. laevigatum* and *L. morrisonii* hypanthium RNA and run using one Novaseq lane of 150bp paired ends, with an average of approximately 24 million reads per sample.

3.3.3 Genome sequencing

High-fidelity (HiFi) Pacific Biosciences (PacBio) long read sequencing was produced using the Sequel® II SMRT® Cell 8M system for both DNA samples at the McDonnell Genome Institute, Washington University in Saint Louis, MI, USA. In total, 706,506 and 783,109 circular consensus sequencing (CCS) reads were produced for *L. morrisonii* and *G. laevigatum*, totalling 15,456,070,399 bp and 15,526,766,806 bp (approximately 51x coverage for both), respectively.

3.3.4 Genome assembly and quality checking

Contig assembly was performed for both sequencing datasets using hifiasm (Cheng *et al.,* 2021) using default parameters. Assembly quality checking was performed using the assemblyQC pipeline (Rashid *et al.,* 2024). Adaptor sequences detected by the assemblyQC (Rashid *et al.,* 2024) were chopped off. Contigs classified as contaminants (such as fung:ascomycetes, fung:basidiomycetes, *Pseudomicrostroma glucosiphilum*) were discarded. Contigs anchoring, orientation, and chromosome

numbering were done by comparative genome mapping against the *L. scoparium* 'Crimson Glory' genome assembly versions CG_v2e (Grierson *et al.*, 2024, Chapter 2).

3.3.5 Synteny analysis for pseudo-chromosome scaffolding

The synteny analysis was carried out using minimap2 v2.22 (Li 2018) and visualised with NGenomeSyn[11]. Firstly, the final curated assemblies were aligned to the 11 chromosomes in CG_v2e using 'minimap2 -x asm5'. The alignments were filtered with the minimum aligned length of 5KB (*G. laevigatum*) or 10KB (*L. morrisonii*) and converted to link files using 'GetTwoGenomeSyn.pl Paf2Link' in the NGenomeSyn package. The results were plotted with 'NGenomeSyn -InConf config'. Evaluation of the synteny plots was used to join and orient the *L. morrisonii* and *G. laevigatum* assembly contigs into pseudo-chromosome size scaffolds that are syntenic to the mānuka genome.

3.3.6 Repeat and gene predictions

De novo transposable elements (TE) detection was performed for each haplotyped assemblies in parallel using edta v2.2.2 (Ou *et al.,* 2019). Each haplotype was soft-masked with the corresponding TE library using RepeatMasker v4.1.5 (https://www.repeatmasker.org/). Gene model prediction was performed on soft-masked genomes using genePal (https://github.com/Plant-Food-Research-Open/genepal/), using supports of RNA-Seq data from hypanthium of both *L. morrisonii* and *G. laevigatum*, as well as liftoff the gene models from mānuka (Thrimawithana *et al.,* 2019).

3.3.7 Orthologous proteins analysis

Orthologous analysis and species tree generation were based on the predicted peptide sequences in each haplome, using orthofinder v2.5.4 (Emms and Kelly 2015, 2019).

3.4 Results and Discussion

3.4.1 Genome assembly of *Gaudium laevigatum*

Two haplotyped assemblies were produced for *G. laevigatum*. The initial assemblies for *G. laevigatum*'s haplotypes (hap) 1 and 2 had 749 and 206 scaffolds (Supplementary Fig. S**3.1**), respectively, for a total of 302,808,981 bp and 270,392,341 bp, with N50 for the scaffolds of 21.5 Mb and 12.5 Mb. This initial assembly was fed into the assemblyQC pipeline to check the presence of adapters, contaminants, assembly metrics and synteny with the L. scoparium 'Crimson Glory' v2.0 genome assembly. One contig in *G. laevigatum* hap1 was trimmed. The assembly was then scaffolded to achieve pseudo-chromosome length using synteny with *L. scoparium*. The final assemblies for *G. laevigatum*'s hap1 and hap2 after removing contaminants and using synteny for anchoring had 708 and 167 scaffolds (Table **3.1**), respectively, for a total of 302,812,000 bp and

270,396,241 bp. The N50 for the scaffolds were 23.9 Mb and 22.8 Mb, with half of the assembled sequenced contained into six scaffolds for both (L50) and 11 scaffolds larger than 15 Mb achieving chromosome-scale. In total, the 11 pseudo-chromosomes of *G. laevigatum* hap1 and hap2 covered 265 Mb and 255 Mb (87.7% and 94.4% of the total assembly for both haplotypes, respectively). BUSCO analysis using the embryophyta_odb10 reference indicated that the chromosome-scale assemblies were near complete, with complete BUSCOs of 98.3% and 97.9%. Analysis of telomeric repeats using TIDK confirmed that the assemblies are near-complete and spanning entire chromosomes. For *G. laevigatum*, eight out of the 11 scaffolds were telomere-to-telomere and gapless for both hap1 and hap2, and therefore can be considered as pseudo-chromosomes (Table **3.2**). The remaining three scaffolds had one telomere at one end. Seven scaffolds had telomeric repeats that were internal, i.e. not located at the chromosome ends.

Table 3.1: Genome assembly metrics for *Gaudium laevigatum* and *Leptospermum morrisonii*.

	***G. laevigatum* haplotype 1**		***G. laevigatum* haplotype 2**		***L. morrisonii* haplotype 1**		***L. morrisonii* haplotype 2**	
Number of scaffolds	708		167		722		238	
Total assembly	302,812,000		270,396,241		379,818,864		343,205,770	
N50	23,953,855		22,860,685		31,493,504		29,049,683	
L50	6		6		6		6	
Complete BUSCOs	1586	98.3%	1581	97.9%	1566	97.1%	1542	95.6%
Complete and single-copy BUSCOs	1552	96.2%	1550	96.0%	1510	93.6%	1510	93.6%
Complete and duplicated BUSCOs	34	2.1%	31	1.9%	56	3.5%	32	2.0%
Fragmented BUSCOs	15	0.9%	14	0.9%	18	1.1%	18	1.1%
Missing BUSCOs	13	0.8%	19	1.2%	30	1.8%	54	3.3%
Total BUSCO groups searched	1614		1614		1614		1614	

3.4.2 Genome assembly of *Leptospermum morrisonii*

Two haplotyped assemblies were produced for *L. morrisonii*. The initial assemblies for *L. morrisonii*'s hap1 and hap2 had 880 and 395 scaffolds (Supplementary Fig. **3.1**), respectively, for a total of 386,546,005 bp and 349,845,363 bp, with N50 for the scaffolds of 6 Mb and 4 Mb, respectively. Eighteen and five contigs were removed from *L. morrisonii* hap1 and hap2 respectively, due to contamination according to NCBI FCS-GX and Kraken2/Krona results. The cleaned-up *L. morrisonii*'s

hap1 and hap2 had 722 and 238 scaffolds, respectively, for a total of 379,818,864 bp and 343,205,770 bp. The N50 for the scaffolds were 31.4 Mb and 29 Mb, with half of the assembled sequences contained into six scaffolds for both (L50) and 11 scaffolds achieving chromosome-scale. The 11 pseudo-chromosomes of *L. morrisonii* hap1 and hap2 covered 341 Mb and 322 Mb (89.9% and 93.8% of the total assembly for both haplotypes, respectively). BUSCO analysis had complete BUSCOs of 97.1% and 95.6% for *L. morrisonii*.

Table 3.2 Pseudo-chromosome assembly of *Gaudium laevigatum* and *Leptospermum morrisonii*.

Chr: chromosome; T2T: telomere-to-telomere; 1T: one telomere; *: internal telomeric sequence

	***G. laevigatum* haplotype 1**	**Telomeric ends**	***G. laevigatum* haplotype 2**	**Telomeric ends**	***L. morrisonii* haplotype 1**	**Telomeric ends**	***L. morrisonii* haplotype 2**	**Telomeric ends**
Chr1	21,003,270	T2T	20,678,918	1T	25,889,885	T2T	29,008,821	T2T
Chr2	25,397,309	T2T	23,904,347	T2T	32,231,048	T2T*	29,736,634	1T
Chr3	29,902,649	1T	26,715,370	T2T	31,830,518	T2T	29,503,176	T2T
Chr4	18,678,803	1T*	20,026,561	1T*	29,946,940	1T*	25,558,253	1T*
Chr5	24,565,314	T2T	22,860,685	T2T	41,213,220	n	24,312,826	T2T
Chr6	29,855,063	T2T	29,083,629	1T	31,493,504	1T	33,585,923	1T
Chr7	23,953,855	T2T*	23,987,127	T2T	29,110,600	T2T	31,430,387	T2T
Chr8	29,553,281	T2T	27,240,377	T2T*	36,811,872	T2T	38,097,710	T2T*
Chr9	19,296,033	1T	18,870,950	T2T	18,605,778	1T	25,276,758	T2T*
Chr10	21,791,746	T2T	21,259,159	T2T	31,105,979	T2T	27,140,084	T2T
Chr11	21,889,327	T2T	20,565,536	T2T	33,052,471	T2T	29,049,683	T2T
Total chromosome-scale assembly	265,886,650		255,192,659		341,291,815		322,700,255	

For *L. morrisonii*, seven and eight out of the 11 scaffolds were telomere-to-telomere and gapless for hap1 and hap2, and therefore can be considered as pseudo-chromosomes (Table **3.2**). Three scaffolds had one telomere at one end for both hap1 and hap2, and only chromosome 5 missed both telomeres in hap1 (while still being the longest scaffold assembled with >41 Mb). Five scaffolds had telomeric repeats that were internal. Internal telomeres were found for all four assemblies (both haplotypes of both species) of chromosome 4.

3.4.3 Repeats and gene model predictions

Repeats predictions were completed in both haplotypes of the *L. morrisonii* and *G. laevigatum* assemblies. In total, 166Mb (43.9%) and 139Mb (40.7%) were repeat sequences in both *L. morrisonii*

assembled haplotypes, respectively (Table **3.3**). This total number of repeats is larger than in the *G. laevigatum* genome, which has 106Mb (35%) and 81Mb (30%), and in mānuka (87Mb, 29.6% of the genome). This indicates that the larger genome of *L. morrisonii* compared to *G. laevigatum* and *L. scoparium* can be attributed to a larger quantity of repeated elements.

Gene model prediction was performed on soft-masked genomes. In total, 33,663 and 31,035 genes were predicted in *L. morrisonii* hap1 and hap2, respectively, whereas *G. laevigatum*'s haplotypes had 30,563 and 30,422 genes.

Table 3.3: Repeat and gene predictions in the *Gaudium laevigatum* and *Leptospermum morrisonii* genomes.

		***L. morrisonii* hap 1**	***L. morrisonii* hap 2**	***G. laevigatum* hap 1**	***G. laevigatum* hap 2**	***L. scoparium***
Total assembled size		379,818,789	343,205,770	302,694,896	270,341,035	294,018,319
GC content		40.74%	40.57%	40.90%	40.93%	39.65%
De novo repeats detection and masking	Base pairs (bp) masked	166,750,051 (43.90%)	139,897,441 (40.76%)	106,016,252 (35.02%)	81,220,902 (30.04%)	87,194,611 (29.69%)
Gene model predictions	Number of genes	33,663	31,035	30,563	30,422	26,419
	Number of mRNAs	37,349	34,484	33,877	33,677	30,231
	Number of CDS	159,927	152,601	153,584	151,618	147,978
	Average gene length (bp)	2,456	2,534	2,538	2,502	2,998
	Gene length range (bp)	96 - 118,484	120 - 61,399	114 - 88,847	96 - 73,836	116 - 133,809
	Number of single exon genes	15,188	13,254	12,543	12,694	9,629

3.4.4 Comparative analysis of predicted proteins in the Leptospermeae

A comparison of the predicted proteins in *L. morrisonii* and *G. laevigatum* was performed against themselves and mānuka (*L. scoparium*) (Fig. **3.1**). Among the total of 169,618 protein-coding genes in the assemblies, 161,235 (95.1%) were assigned to 30,258 orthogroups and 4.9% genes as un-assigned singletons. A total of 18,121 groups were shared between all three species. There are 1,580 groups present in *L. morrisonii* and *G. laevigatum*, but absent from mānuka. In total, 2,512 groups (3,367 and 2,876 genes in hap1 and hap2, respectively) were detected in both haplotypes and specific to *G. laevigatum*; while 1,857 groups (with 2,506 and 2,701 genes in hap1 and hap2) in both haplotypes and specific to *L. morrisonii*. There were more groups (1,035) with orthologous genes in *L. morrisonii* and mānuka but absent from *G. laevigatum*, comparing to 283 groups that are common in mānuka and *G. laevigatum*, but absent from *L. morrisonii*. The predicted proteins that were different between the haplotypes of the same species are of interest as these may be variants that segregate in the species and may be linked to trait variability.

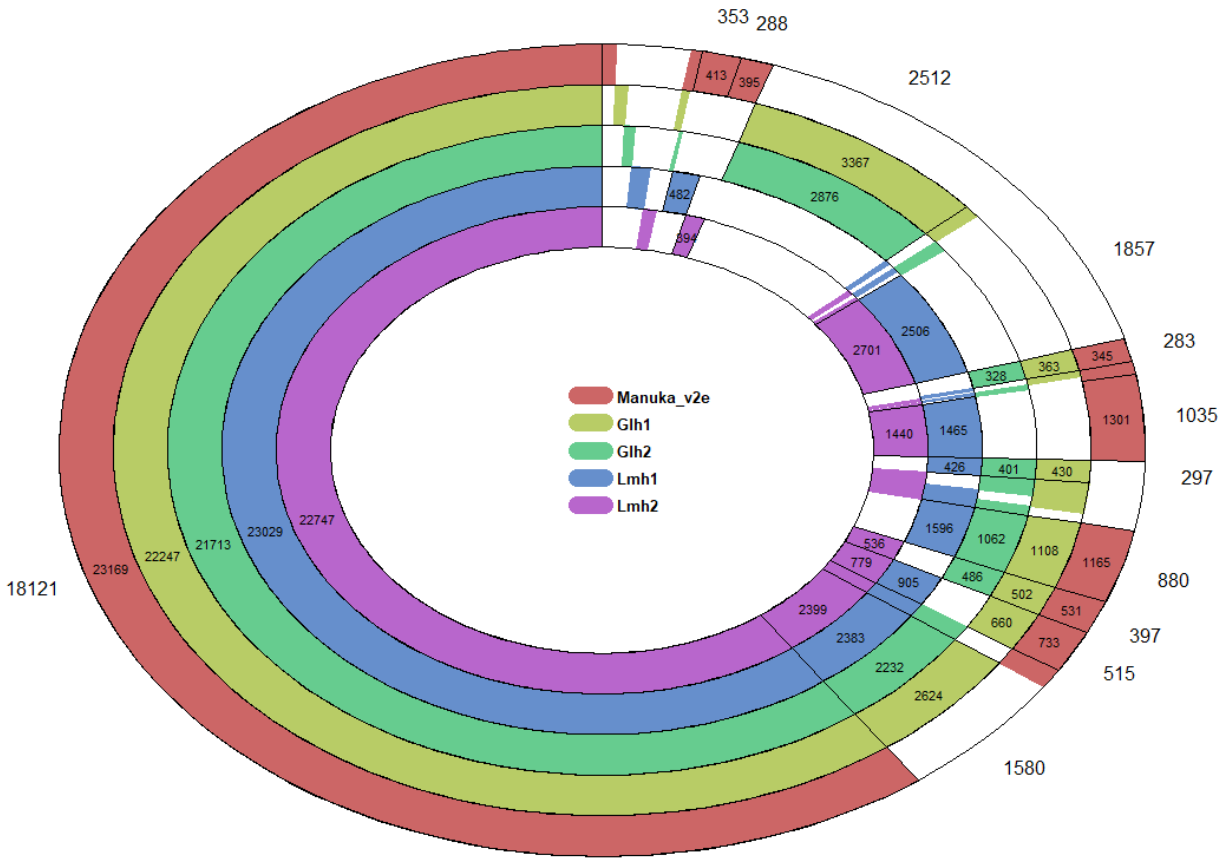


Figure 3.1: Orthologue analysis of protein coding genes across in three Leptospermeae genome assemblies. Each ring represents a haplotype (Gh1, Gh2, Lh1, Lh2 for *Gaudium laevigatum* and *Leptospermum morrisonii,* respectively) or the primary assembly (*Leptospermum scoparium*). The external numbers denote the count of orthologue groups (OGs) shared by the coloured assemblies, while the number on each ring represents the total of orthologous genes in the OGs. White sections indicate the absence of orthologous genes within that specific assembly.

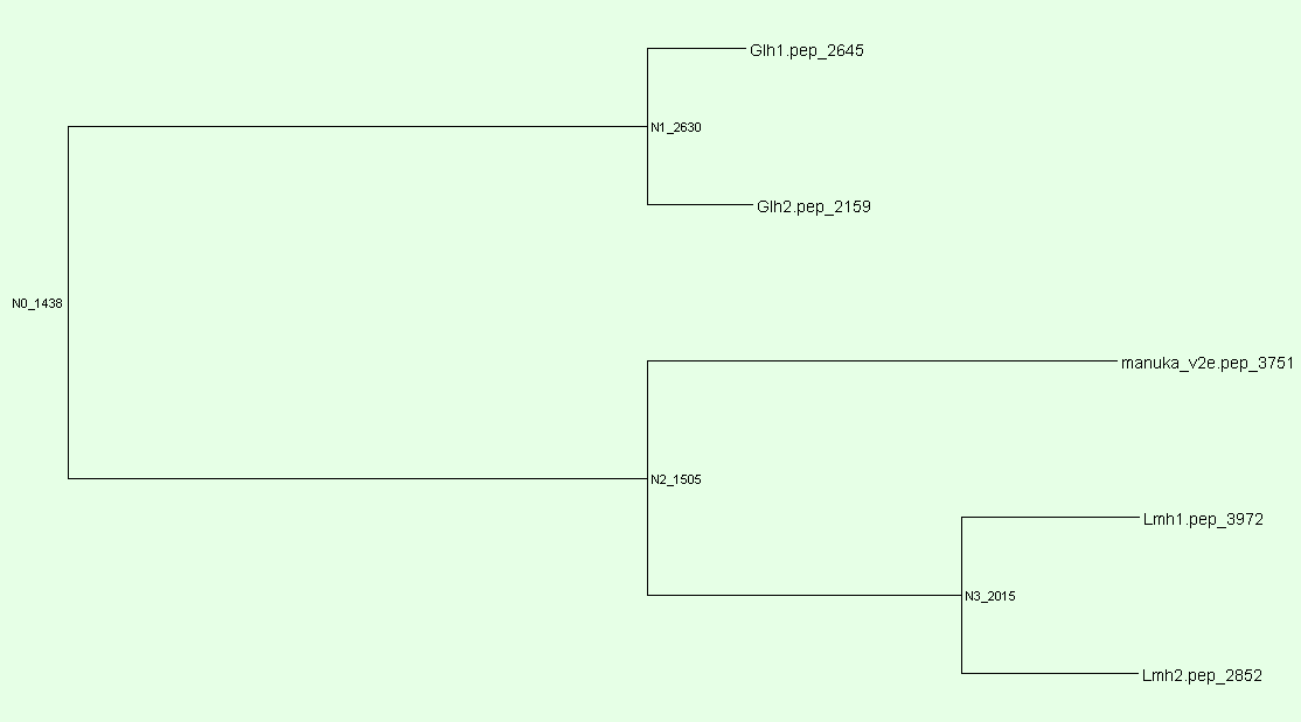


Figure 3.2: Species tree based on genome-wide orthologous gene analysis. Nodes were labelled with their hierarchical orthologue groups (Nx) and the number of duplication events (≥50% support) along the branch leading to that node/species. Species *Gaudium laevigatum* (Gh1 and Gh2) is separated from the clade containing species *Leptospermum morrisonii* and *Leptospermum scoparium* (New Zealand mānuka).

The phylogenetic tree constructed from the amino acid sequences illustrated that *G. laevigatum* branched away from the other two organisms (Fig. **3.2**), while *L. morrisonii* and mānuka shared a common ancestor N2 with 1,505 supporting duplication events. Gene trees analyses revealed 32,341 hierarchical orthogroups (HOGs) at the root node of the species tree (N0). There are 29,413, 29,974 and 30,116 HOGs for nodes N1, N2 and N3, respectively. This new phylogenetic tree constructed using orthologous genes confirms the expected phylogenetic relationship between the three species in the Leptospermeae.

In conclusion, we have assembled and annotated reference quality genomes for *G. laevigatum* and *L. morrisonii*, two Myrtaceae species endemic to Australia. Conserved synteny across Leptospermeae genomes suggests high levels of genome stability among members of the genus. These genome assemblies are useful resources for future research on the species biology, including studying traits related to nectar and foliage chemistry, and understanding the genetic diversity in their natural ranges.

Data availability

Genome assemblies were submitted to NCBI under the following BioProjects: *Gaudium laevigatum* (PRJNA1238035 and PRJNA1238034); *Leptospermum morrisonii* (PRJNA1238037 and PRJNA1238036).

Acknowledgements

The authors thank Pieter van As (Neogen Australasia) for his help with PacBio sequencing sequencing and Graham Pearson (Castlecliff Coastcare) for assistance with locating *Gaudium laevigatum*.

Funding

DC and EG received funding from the New Zealand Ministry for Business Innovation and Employment (MBIE) Endeavour Smart Idea funding (C11X2102). DC and CD received funding from the MBIE Strategic Science Investment Fund's platform Genomics Aotearoa (https://www.genomics-aotearoa.org.nz/).

3.5 References

Bicknell R, Boase M, and Morgan E, 2019. Developing polyploid genotypes of *Leptospermum scoparium*. *New Zealand Journal of Crop and Horticultural Science.* **47**:273-281.

Chagné D, Montanari S, Kirk C, Mitchell C, Heenan P, Koot E. **2023**. Single nucleotide polymorphism analysis in *Leptospermum scoparium* (Myrtaceae) supports two highly differentiated endemic species in Aotearoa New Zealand and Australia. *Tree Genetics & Genomes* **19**: 31.

Cheng H, Concepcion GT, Feng X, Zhang H, and Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**: 170-175.

Emms DM, and Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**: 157.

Emms DM, and Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Grierson ERP, Thrimawithana AH, van Klink JW, Lewis DH, Carvajal I, Shiller J, Miller P, Deroles SC, Clearwater MJ, Davies KM, *et al.* 2024. A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*). *New Phytologist* **242**: 2270–2284.

Hilario E. 2018. *Plant nuclear genomic DNA preps*. [WWW document] URL dx.doi.org/10.17504/protocols.io.rncd5aw [accessed 15 March 2022]

Koot E, Arnst E, Taane M, Goldsmith K, Thrimawithana A, Reihana K, González-Martínez SC, Goldsmith V, Houliston G, Chagné D. **2022**. Genome-wide patterns of genetic diversity, population structure and demographic history in mānuka (*Leptospermum scoparium*) growing on indigenous Māori land. *Horticulture Research* **9**: uhab012.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34:** 3094-3100.

Ou S, Su W, Liao Y, Chougule K, Agda JRA et al., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20:** 275.

Rashid U, Wu C, Shiller J, Smith K, Crowhurst R et al., 2024. AssemblyQC: a Nextflow pipeline for reproducible reporting of assembly quality. *Bioinformatics* **40**: btae477.

Thrimawithana AH, Jones D, Hilario E, Grierson E, Ngo HM, Liachko I, Sullivan S, Bilton TP, Jacobs JME, Bicknell R, *et al.* 2019. A whole genome assembly of *Leptospermum scoparium* (Myrtaceae) for mānuka research. *New Zealand Journal of Crop and Horticultural Science* **47**: 233–260.

Wilson PG, Heslewood MM. **2023**. Revised taxonomy of the tribe Leptospermeae (Myrtaceae) based on morphological and DNA data. *TAXON* **72**: 550–571.

Chapter 4*:* Regulation of *Sgpp2* and the structure of the *Sgpp* locus in the Leptospermeae

4.1 Introduction

Mānuka (*Leptospermum scoparium*, J.R. Forst et G. Forst) nectar contains variable amounts of a triose sugar – dihydroxyacetone (DHA). DHA is the precursor to Methylglyoxal (MGO), the antimicrobial compound underpinning the high value of mānuka honey and forming the basis of the Aotearoa-New Zealand honey industry. The presence of DHA in nectar is a rare trait and has only been reported in some species within the Myrtaceae. Understanding the basis of DHA production has been a recent research focus. A model has been proposed for mānuka nectar production that could explain the origin of DHA (Clearwater *et al.*, 2021; Grierson *et al.*, 2024, Chapter 2), and a phosphatase gene (*Sgpp2*) has recently been linked to nectar DHA accumulation in *L. scoparium* (Grierson *et al.*, 2024, Chapter 2).

The *Leptospermum* genus been shown to be polyphyletic (Binks *et al.*, 2022). Based on an analysis of the entire tribe Leptospermeae (Myrtaceae), the genus has recently been revised and split into five genera – *Aggreflorum*, *Acetospermum*, *Gaudium*, *Leptospermopsis* and *Leptospermum* (Wilson & Heslewood, 2023) – four of which contain species with DHA in their nectar. This trait until recently was thought to be confined to the previously defined *Leptospermum* genus. However two more nectar-DHA producing species outside of the Leptospermeae have recently been identified (Obeng-Darko *et al.*, 2023). These species, *Ericomyrtus serpyllifolia* and *Verticordia chrysantha*, belong to genera within the Chamelaucieae, which is the sister tribe to the Leptospermeae (Wilson *et al.*, 2001). The presence of DHA in nectar is a trait that has been found so far among a total of six genera within the Myrtaceae family and is often observed in some but not all members of the genus – such as presence in *V. chrysantha*, absent in *Verticordia picta* (Obeng-Darko *et al.*, 2023), and presence in *Aggreflorum whitei*, and absence in *Aggreflorum parviflorium* (Williams *et al.*, 2018). The recent discovery of DHA producing species outside of the Leptospermeae highlights that there are many species within these tribes with nectar that is so far untested for presence of DHA (Obeng-Darko *et al.*, 2023).

The *Sgpp2* gene has been recently linked to nectar DHA amounts in *L. scoparium*, and it resides within a complex locus containing three duplicated *Sgpp* gene copies of which *Sgpp2* is the highly and differentially expressed copy linked to nectar DHA (Grierson *et al.*, 2024, Chapter 2). So far, we

know that the *Arabidopsis* genome contains one copy of the *Sgpp* gene and it has not yet been established how many *Sgpp* genes there are, or how they expressed in other DHA producing taxa and the wider Myrtaceae.

Expression of the *Sgpp2* gene is very high and flower specific in *L. scoparium* (Grierson *et al.*, 2024, Chapter 2). The mechanisms controlling gene expression and tissue specificity in general are complex and often involve multiple *cis* (e.g DNA sequence motifs) and *trans* (e.g transcription factors) acting regulatory elements (See reviews Schmitz *et al.*, 2022; Marand *et al.*, 2023; Beernink *et al.*, 2025). The promoter of a gene is generally considered to be the 1-2kbp of the genome upstream of its transcription start site (TSS) and contains the core promoter and the proximal promoter (Schmitz *et al.*, 2022). The core promoter is the minimum sequence required to initiate transcription, and it often contains key *cis* elements (Smale & Kadonaga, 2003). The proximal promoter contains other *cis*-regulatory motifs that can enhance or repress gene expression through the binding of transcription factors. Distal *cis* elements include enhancers and repressors further from the target gene that can be multiple megabases away and are brought into close proximity through chromatin looping (Beernink *et al.*, 2025). Many other factors can affect gene expression and transcript abundance, such as chromatin accessibility and microRNA-triggered degradation (Bartel, 2009; Schmitz *et al.*, 2022).

The genomes of two *Leptospermum* species – *Leptospermum morrisonii (*Joy Thomps.), which accumulates DHA in its nectar, and *Gaudium laevigatum* (Gaertn. Peter G. Wilson), which does not accumulate DHA in its nectar (Williams *et al.*, 2018), have recently been assembled (Chagné, Deng & Grierson, Chapter 3). We aim to further understand the regulation of the *Sgpp2* gene and the structure of the *Sgpp* locus within the Leptospermeae and wider Myrtaceae family. We report here an investigation of the *Sgpp* locus, with specific focus on the promoter of *Sgpp2* genes that are highly expressed. The *Sgpp* locus, gene sequences, and transcript abundances were compared among three Leptospermeae species – *L. scoparium*, *L. morrisonii*, and *G. laevigatum*. Conserved and unique regions were identified between and within the *Sgpps*, and regions and binding motifs unique to the highly expressed *Sgpp* genes were identified. Within these were two C-box motifs, and a bZIP transcription factor predicted to bind to them was identified as significantly differentially expressed in the RNAseq dataset presented in Grierson *et al.* (2024, Chapter 2), and was subsequently found to co-locate with another of the QTLs identified in Grierson *et al.* (2024, Chapter 2). The promoter region within 13 *Leptospermum scoparium* genotypes with varying DHA phenotypes was found to be highly similar indicating further complex transcriptional regulation fine tuning *Sgpp2* expression beyond this promoter region. The *L. morrisonii Sgpp2* promoter driving *Venus* lines of *Petunia* and

Nicotiana were made to assess the *LmSgpp2* promoter *in vitro*, and it was found to have strong nectary specific expression.

4.2 Methods

4.2.1 Plant material

Plant material for the genomes used have been described previously (*L. scoparium* – Grierson *et al.* 2024, Chapter 2, *L. morrisonii* and *G. laevigatum* – Chagné, Deng & Grierson, Chapter 3). Fourteen genotypes of *L. scoparium* and *Leptospermum* crosses were used for further analysis by PCR. The genotypes and species selected and which aspects of this study they were used for are listed in Supplementary Table **S6.1**.

4.2.2 DNA extraction

DNA was extracted from leaf tissue using a rapid CTAB-based extraction protocol based on Stewart & Via (1993).

4.2.3 Nectar and nectary sampling, RNA extraction

As myrtaceous nectaries are generally not well delineated, the whole hypanthium was sampled to provide nectary-enriched tissue. Hypanthium tissue was collected from *L. morrisonii* and *G. laevigatum* flowers at floral developmental stage 6-7 (Smallfield *et al.*, 2018) as described in Figure **S2.1** within Chapter 2, however the ovules within the top portion of locules were not removed. Four *L. morrisonii* hypanthium samples were taken across two days and three *G. laevigatum* hypanthium samples were taken in one day. RNA was extracted from nectary enriched tissue based on the CTAB method from Gambino *et al.* (2008). Each sample consisted of 10 hypanthia and had a paired nectar sample taken from the same hypanthia. Nectar was sampled as described in Grierson *et al.* (2024, Chapter 2).

4.2.4 RNAseq and RT-qPCR

RNA sequencing (RNAseq) library preparation and sequencing were completed by Novogene (Hong Kong). Seven Illumina stranded mRNA libraries were prepared and run in one Novaseq lane of 150 bp paired-end reads, resulting in an average of approximately 24 million reads per sample. The reads were then quality checked with FASTQC (v.0.12.1) and adapter trimmed using fastp (0.23.02). The *L. morrissonii* and *G. laevigatum Sgpp* loci genome sequences were obtained from their respective genomes (Chagné, Deng & Grierson, Chapter 3) and were manually annotated. The quality trimmed reads were mapped to their respective loci using STAR (v.2.7.10a, Dobin *et al.*, 2013) aligner with '--quantMode GeneCounts'.

RNAseq was validated with RT-qPCR. To amplify from transcripts in a range of species, primers were designed from conserved regions to match all *Sgpp* transcripts. The primers for *Sgpp* and the reference genes *Actin* and *rpl3* are listed in Table **S2.2** in Grierson *et al*. (2024, Chapter 2).

4.2.5 Nectar DHA quantification

Nectar DHA was analysed as described in Grierson *et al*. (2024, Chapter 2).

4.2.6 Genome sequences for comparison

High quality genomes from throughout the Myrtaceae, as well as one from within the Myrtales, were blasted for the *Sgpp1* gene (genomes and accession details are listed in Table **S4.3**). If the genomes were annotated those annotations were used, and if not the *Sgpp* locus was then manually annotated based on *Sgpp* transcripts for the species if present on NCBI or using the *Leptospermum* gene sequences. Predicted whole genes (all exons and introns) were used to construct phylogenies using MrBayes (v3.2.6, Ronquist *et al.*, 2012) in Geneious Prime (v2022.0.1) using the default parameters.

4.2.7 Promoter element analysis

The promoter sequences of *Sgpp*s from the three Leptospermeae genomes were compared (*L. scoparium* cv. 'Crimson Glory' – Chapter 2, *L. morrisonii* and *G. laevigatum* – Chapter 3). Approx 1kb of the proximal promoter from each *LsSgpp2*, *LsSgpp3*, *LmSgpp2*, *LmSgpp3*, and *GlSgpp2* were annotated with promoter elements using the plantCARE database (PlantCARE webtool: Lescot *et al.*, 2002). The unique conserved regions were also annotated with the newPLACE database (newPLACE webtool: Higo *et al.*, 1999). The promoter sequences were compared by MUSCLE (Edgar, 2004) alignment followed by manual alignment of the sequences. Structural differences and conserved regions were identified by pairwise alignments and annotations, followed by alignment of all five promoter sequences to identify the regions unique to the highly expressed *Sgpp2* promoters (*L. morrisonii* and *L. scoparium*, as discussed further below), and conserved across all five promoters (defined here as >85% identical). Motifs unique to the highly expressed *Sgpp2* promoters were identified through comparison of conserved unique regions, as well as motifs present only in the highly expressed *Sgpp2s* due to single nucleotide polymorphisms (SNPs).

The RNAseq dataset published in Grierson *et al*. (2024) was searched for significantly differentially expressed transcription factors of types that might bind to motifs identified in the unique conserved promoter regions. These were identified by literature search of the annotations and motif sequences from the plantCARE and newPLACE databases.

4.2.8 PCR amplification of locus

A series of PCRs spanning the *Sgpp* locus were conducted to sequence the locus in a range of *Leptospermum* genotypes (see Supplementary Tables **S4.1**, **S4.4** and **S4.5**). Product lengths ranged from 1.7 to 6.5kb. Multiple PCR primers were designed in conserved regions (based on 'Crimson Glory' v2.0 genome – Chapter 2, *L. morrisonii* and *G. laevigatum* genomes – Chapter 3), and different combinations tried for each genotype. The PCR reactions resulting in the cleanest, ideally single band product were chosen for minION sequencing. PCRs were done using iProof (BioRad), with further information in Supplementary Tables **S4.4** and **S4.5**.

4.2.9 MinION sequencing and assembly

A total of 64 amplicons were sequenced across three runs – using one FLO-MIN06 flowcell and one FLO-MIN112 flowcell – on a Mk1B minion device (Oxford Nanopore Technologies). The protocol followed and kit used was ligation sequencing of amplicons using native barcoding (NBD112.24, Oxford Nanopore Technologies). In brief, the protocol was as follows: Amplicons were end-prepped using the NEBNext Ultra II end repair/dA tailing module (NEB, E7546) according to manufacturer's instructions. DNA cleanup was performed using an equal volume of Ampure XP (Beckman Coulter) beads as per manufacturers' instructions. The barcodes were then ligated using Blunt/TA ligase Master Mix (New England Biolabs, M0367) and pooled into a 1.5mL DNA LoBind tube (Eppendorf). The pooled barcoded libraries were then cleaned using 0.4x volume of Ampure beads. Adapters were then ligated using the NEBNext Quick ligation Reaction Module (E6056) and then cleaned with 0.4 x volume of Ampure beads. The library was quantified on a Qubit fluorometer (Invitrogen) using a dsDNA HS kit (Invitrogen, Q32851). Approximately 11-15fmol (depending on average length of amplicons within each run) was then loaded onto the flowcell. The flowcell was loaded according to manufacturer's instructions, and ran for at least 2 hours, yielding at least 200,000 reads per flowcell. The FLO-MIN112 flowcell was washed using the Flowcell Wash Kit (EXP-WSH004) and reused.

Sequenced amplicons were basecalled and sorted into barcodes using Guppy (version 6.4.6-gpu). Nanoplot (nanopack v2.2.2) was used to visualise reads, and Chopper (v.0.6.0) was used to trim the reads and filter based on quality to result in at least 100 high quality reads (Q scores used were between 20 and 25, with an average of 23.5). Consensus sequences were assembled in Geneious Prime (v2022.0.1), by mapping reads to the 'Crimson Glory' v2.0 genome (Grierson *et al*., 2024, Chapter 2) and manually extracting each haplotype. Reads for each haplotype were then aligned, and a consensus sequence extracted.

4.2.10 Construct design

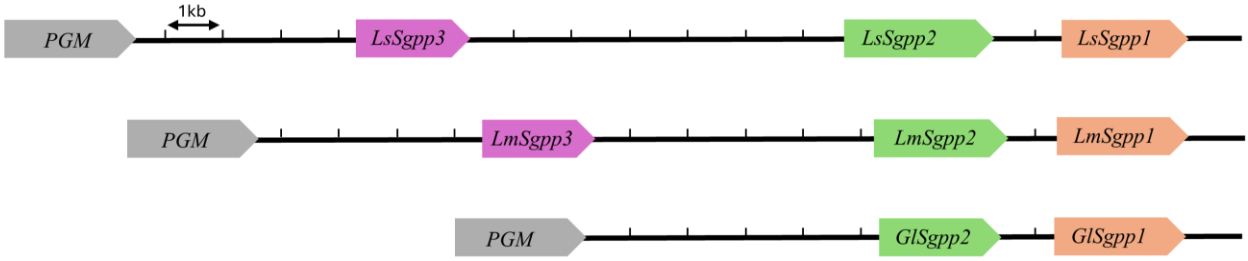
Promoter sequences were identified through blasting the *Leptospermum morrisonii* genome (Chagné, Deng & Grierson, Chapter 3) with sequences for the *Petunia hybrida Nec1/Sweet9* (from Ge *et al.* (2000), AF313914.1) and *LsSgpp2* genes. The *LmNec1/Sweet9* promoter was based on the length of *Nec1* promoter as described in Ge *et al.* (2000), and the *LsSgpp2* promoter encompassed the conserved regions between *LmSgpp2* and *LsSgpp2* promoters. The *LmSgpp2*promoter-Venus (3.36kb) and *LmNec1promoter*-Venus (2.53kb) expression cassettes in pUC57 were synthesised by Genscript (USA). Both expression cassettes were then cloned into pART27 for *Agrobacterium tumefaciens* (*Agrobacterium*)-mediated transformation The resulting plasmids were termed pEG53 (*LmSgpp2promoter-Venus-OCSterminator*) and pEG54 (*LmSweet9promoter-Venus-OCSterminator*) (Supplementary Fig. **S4.1**).

4.2.11 Transformation protocol

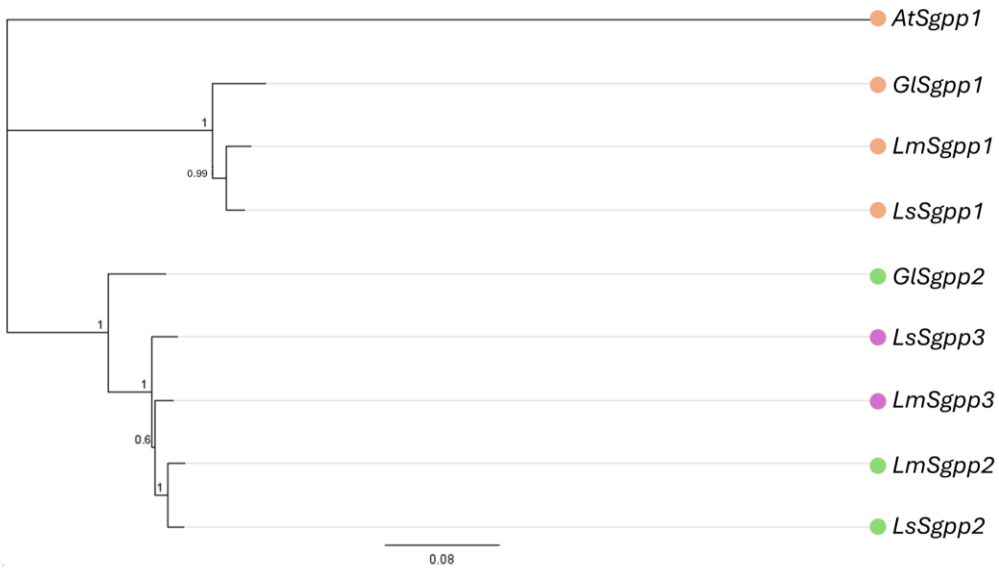
Petunia cv 'Mitchell' (*Petunia axillaris* x (*P. axillaris* x *P. hybrida*)) and *Nicotiana bethamiana* were grown in greenhouse conditions and young leaves were taken for *Agrobacterium*-mediated transformation. The leaves were sterilised by submersion in 10% bleach with Tween20 for 10 mins, swirling occasionally. The leaves were rinsed three times in Milli-Q water and then blotted dry. Leaf discs were made using a 5mm Rapid-Core biopsy punch (ProSciTech) and placed on cocultivation media for 4 days (media contents listed in Supplementary Table **S4.2**). When the *Agrobacterium* cultures reached an A_{600} of 0.6-08, they were pelleted and resuspended in 50ml minimal media + 200µM acetosyringone to an A_{600} of 0.2-03 and incubated with shaking for 2-4 hours at 28. Leaf discs were incubated in 25ml *Agrobacterium* cultures for 5 mins, then blotted with sterile tissue paper and plated back onto cocultivation media. Uninoculated regeneration and kill controls were also plated on cocultivation media. Leaf discs were then cocultivated at 22°C for 3 days. Leaf discs were then transferred to selective media (with kanamycin for pEG53 and pEG54 treatments and kill controls) and regeneration media (regeneration controls). Each plate was transferred to new media every 3 weeks, and shoots were harvested as they arose into tubs of the same media to continue to grow. When shoots were at least 2-3cm, they were transferred onto selective rooting media (treatments) or regeneration rooting media (regeneration controls). After 3 to 6 weeks, rooted shoots were exflasked into the greenhouse.

At least 20 independent lines for each treatment were exflasked and an approximately 1kb portion of the T-DNA insert amplified by PCR to check they were transgenic.

(a)



(b)



(c)

	Residue								60										70						
Consensus	G	F	N	G	G	X	P	I	T	E	D	F	F	X	E	N	X	X	G	K	H	N	E	X	I	A	X
LmSGPP1	G	F	N	G	G	V	P	I	T	E	D	F	F	I	E	N	I	A	G	K	H	N	E	D	I	A	E
LmSGPP2	G	F	N	G	G	V	P	I	T	E	D	F	F	I	E	N	M	S	G	K	H	N	E	E	I	A	E
LmSGPP3	G	F	N	G	G	V	P	I	T	E	D	F	F	I	E	N	I	S	G	K	H	N	E	E	I	A	E
LsSGPP1	G	F	N	G	G	V	P	I	T	E	D	F	F	V	E	N	I	A	G	K	H	N	E	D	I	A	E
LsSGPP2	G	F	N	G	G	V	P	I	T	E	D	F	F	I	E	N	M	S	G	K	H	N	E	E	I	A	E
LsSGPP3	G	F	N	G	G	A	P	I	T	E	D	F	F	I	E	N	I	S	G	K	H	N	E	E	I	A	E
GlSGPP1	G	F	N	G	G	V	P	I	T	E	D	F	F	V	E	N	I	A	G	K	H	N	E	D	I	A	E
GlSGPP2	G	F	N	G	G	V	P	I	T	E	D	F	F	I	E	N	V	S	G	K	H	N	E	D	I	A	Q
	Loop																Motif 5									

Figure 4.1: *Sgpps* in Leptospermeae. (a) Locus structure in Leptospermeae genomes showing *Sgpp1*, *Sgpp2*, and *Sgpp3* where present. *Pgm* (phosphoacetylglucosamine mutase) is the gene flanking the *Sgpp* locus. (b) Whole gene sequence phylogeny of Leptospermeae *Sgpps*. *Sgpp1, Sgpp2* and *Sgpp3* sequences from *Leptospermum morrisonii*, *Leptospermum scoparium* and *Gaudium laevigatum* with *Arabidopsis Sgpp1* as an outgroup. Phylogeny created using in Geneious (v2022.0.1) with the MrBayes plugin (v3.2.6, Ronquist *et al.*, 2012). (c) Alignment of the residues from 52-78 of the SGPPs if present from *L. morrisonii*, *L. scoparium* and *G. laevigatum*. Motif 5 and its preceding loop are indicated.

4.2.12 Visual screening and imaging of transgenic lines

All lines were visually screened for the presence of fluorescence in the nectaries and floral parts. For *LmSgpp2promoter-Venus* and *LmSweet9promoter-Venus Petunia*, developmental series (stage 1 2 and 3) images were taken of the nectaries from three lines. Floral developmental stages were based on Ge *et al.* (2000) and are defined as – Stage 1- short (2-4cm) closed corolla, white nectary; stage 2: elongated (5-6cm) closed corolla, yellow nectary; and stage 3: elongated (5-6cm) open corolla, yellow to orange nectary (Ge *et al*. (2000) stage 3 and 4 combined). For *LmSgpp2promoter-Venus* and *LmSweet9promoter-Venus* transformed *Petunia* and *Nicotiana*, three lines for each treatment had stage 3 flowers imaged 3. YFP fluorescence was viewed with a YFP filter and blue light. Paired white light and blue light (with YFP filter). images were taken on a Leica camera (DF550) on a Leica dissection microscope (M205FA).

4.3 Results

4.3.1 Differences in the *Sgpp* locus between *L. scoparium*, *L. morrisonii* and *G. laevigatum*

The *L. morrisonii* genome contains all three *Sgpp* genes as described in Grierson *et al*. (2024, Chapter 2), and *G. laevigatum* only has two *Sgpp* copies, *Sgpp1* and *Sgpp2* (Fig. **4.1a**). The three *Sgpp1*s clade together, and the *Sgpp2s* and *3*s clade together (Fig. **4.1b**), and in both clades the *GlSgpp*s clade separately reflecting that the two *Leptospermum* species share a more recent common ancestor compared with *G. laevigatum*. Within the *Sgpp2-3* clade, all the *Leptospermum Sgpps* clade together indicating that the duplication event that gave rise to *Sgpp3* may have occurred after their separation from *Gaudium*. The substrate specificity motif is identical within the *Leptospermum* SGPP2s (MSGKH) and SGPP3s (ISGKH), but GlSGPP2 has a change to Valine at position 68 (VSGKH, Fig. **4.1c**).

4.3.2 Two *Sgpp* gene copies probably the ancestral trait within the Myrtaceae

Myrtaceae *Sgpp* gene phylogeny is shown in Figure **4.2**, with nodes annotated with Bayesian posterior probability (PP). Posterior probability above 0.95 was considered a strongly supported clade. All *Sgpp1*s within Myrtaceae clade together with strong support (PP=0.95), and within that, the Leptospermeae tribe *Sgpp1*s clade together (PP=1, Fig. **4.3**). All *Sgpp2*s within Myrtaceae clade together, along with *Sgpp3*s. *Leptospermum Sgpp2*s and *3*s clade together (PP=1, Fig. **4.2**). The *Eucalyptus Sgpp2* and *3* clade together, indicating a probable separate duplication from the *Leptospermum Sgpp3* duplication (Fig. **4.3**). Gene phylogenies across both the *Sgpp1* and *Sgpp2-3* clades reflect their overall phylogenetic relationships within Myrtaceae (Fig. **4.2** and Fig. **4.3**).

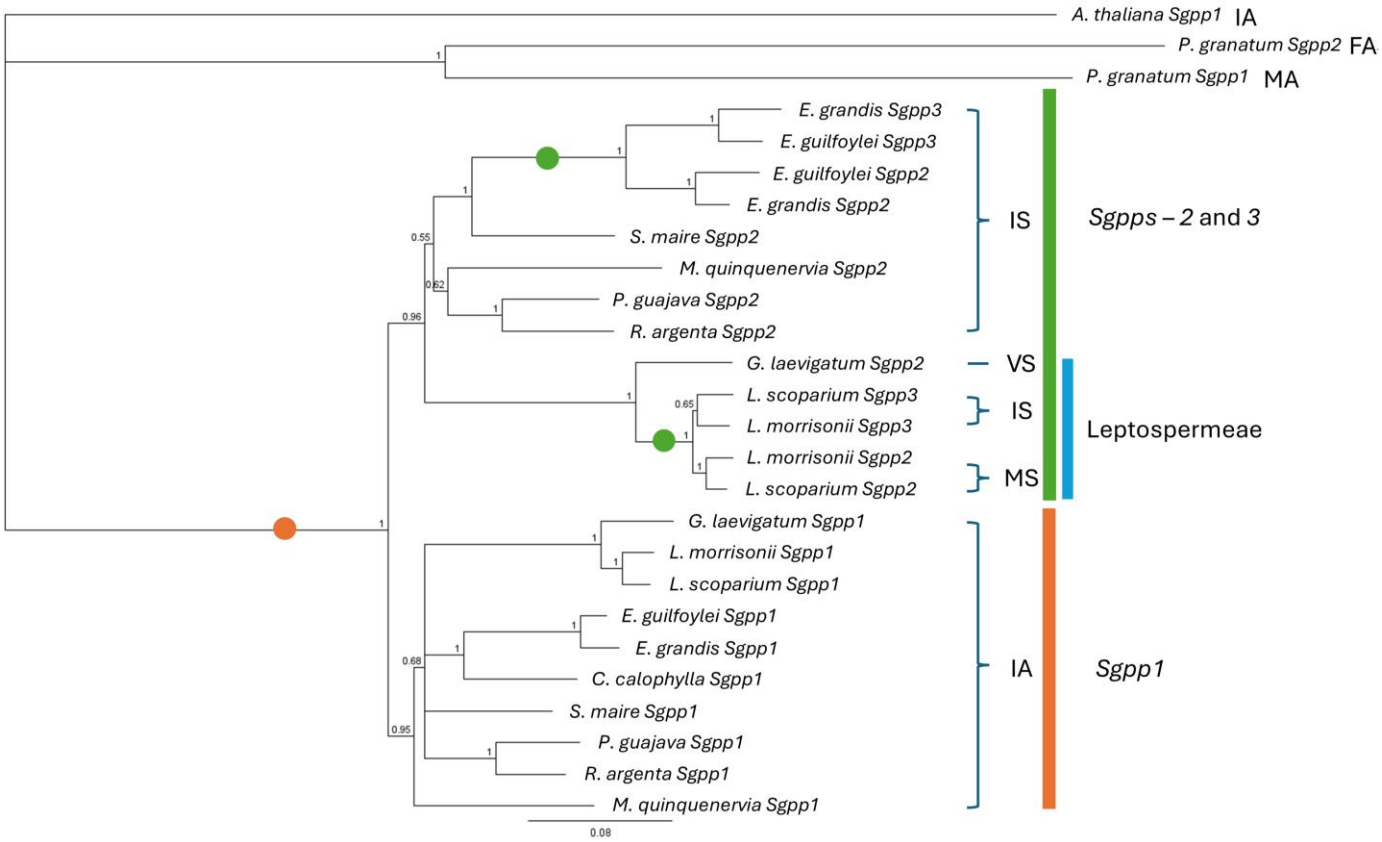


Figure 4.2: Whole gene sequence phylogeny focussing on Myrtaceae *Sgpps*. *Sgpp1, Sgpp2* and *Sgpp3* sequences from *Punica granatum*, *Melaleuca quinquenervia*, *Syzigium maire*, *Psidium guajava*, *Rhodamnia argenta*, *Leptospermum morrisonii*, *Leptospermum scoparium*, *Gaudium laevigatum*, *Corymbia calophylla*, *Eucalyptus grandis*, and *Eucalyptus guilfoylei* with *Punica granatum* representing a Myrtales species, and *Arabidopsis Sgpp1* as an outgroup. Phylogeny created using in Geneious (v2022.0.1) using the MrBayes plugin (v3.2.6, Ronquist *et al.*, 2012). Nodes are annotated with bayesian posterior probabilities, highly supported nodes (PP 0.95-1). The variable residues from the substrate specificity motif (positions 69 and 69, see Figure 1c) are indicated next to each gene. The orange circle indicates duplication of *Sgpp1* to *Sgpp2*, and the green circles indicate the two independent duplications of *Sgpp2* to *Sgpp3.*

The presence of both *Sgpp1* and *Sgpp2* across most of the Myrtaceae included here indicates that the presence of a second *Sgpp* copy is probably the ancestral state of this clade (Fig. **4.3**). The presence of DHA in the nectar of many species within the sister tribes Leptospermeae and Chamelaucieae suggests that nectar DHA is the ancestral trait for that clade. The lack of DHA in the nectar of many species within these tribes also suggest that the ancestral trait has often been lost.

All *Sgpp1*s retain the motif 5 sequence of IAGKH (Fig. **4.2**) and all *Sgpp2*s have the alanine to serine shift at position 69 supporting a shared evolutionary origin. The isoleucine to methionine shift at position 68 appears to have occurred within the *Leptospermum Sgpp2* clade, after the duplication event that produced *Sgpp3*.

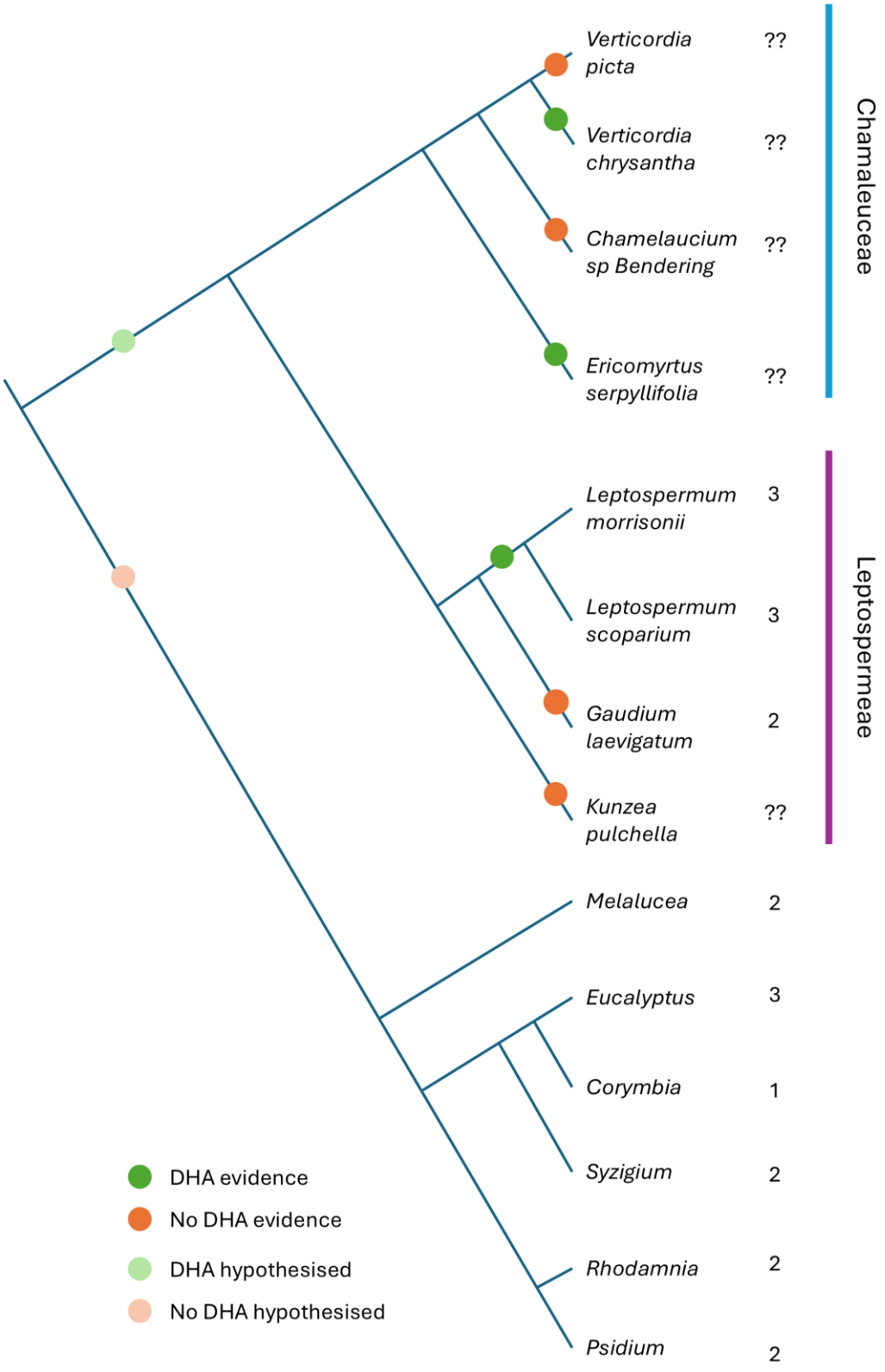


Figure 4.3: Relationships within the Myrtaceae. Genera or species are included based on their inclusion in the *Sgpp* tree (Figure 4.2) or known presence or absence of DHA. The Leptospermeae clade has many more species with reported DHA that are not included here. The coloured circle indicates the demonstrated presence (green) or absence (orange) of DHA within a branch, as well as hypothesised presence of DHA (light green) or absence (light orange). The number of *Sgpp*s within the locus is indicated above the taxa label. If currently unknown, is marked with ??.

4.3.3 Expression patterns of *Sgpp2*: high in *L. morrisonii* and low in *G. laevigatum*

RNAseq transcript counts indicated high levels of *Sgpp2* expression (read counts averaging 232,189) in the *L. morrisonii* hypanthium, which was validated with RT-qPCR (Fig. **4.4**). The transcript levels of *LmSgpp1* (average of 2,122) and *LmSgpp3* (average of 2,376) were low in comparison. The *G. laevigatum Sgpp2* transcript levels were also low in comparison to the *LmSgpp2* (average 1,166), which was also validated with RT-qPCR (Fig. **4.4**). The transcript levels of *GlSgpp1* were similar to *GlSgpp2* (average of 949). The results of the transcript abundance correlated with the DHA amounts found in the nectar of each sample (*L. morrisonii* average nectar DHA = 4.21mg/g sugar, *G. laevigatum* nectar DHA was below the quantitation limit of 0.02mg/g sugar, Fig. **4.4**)

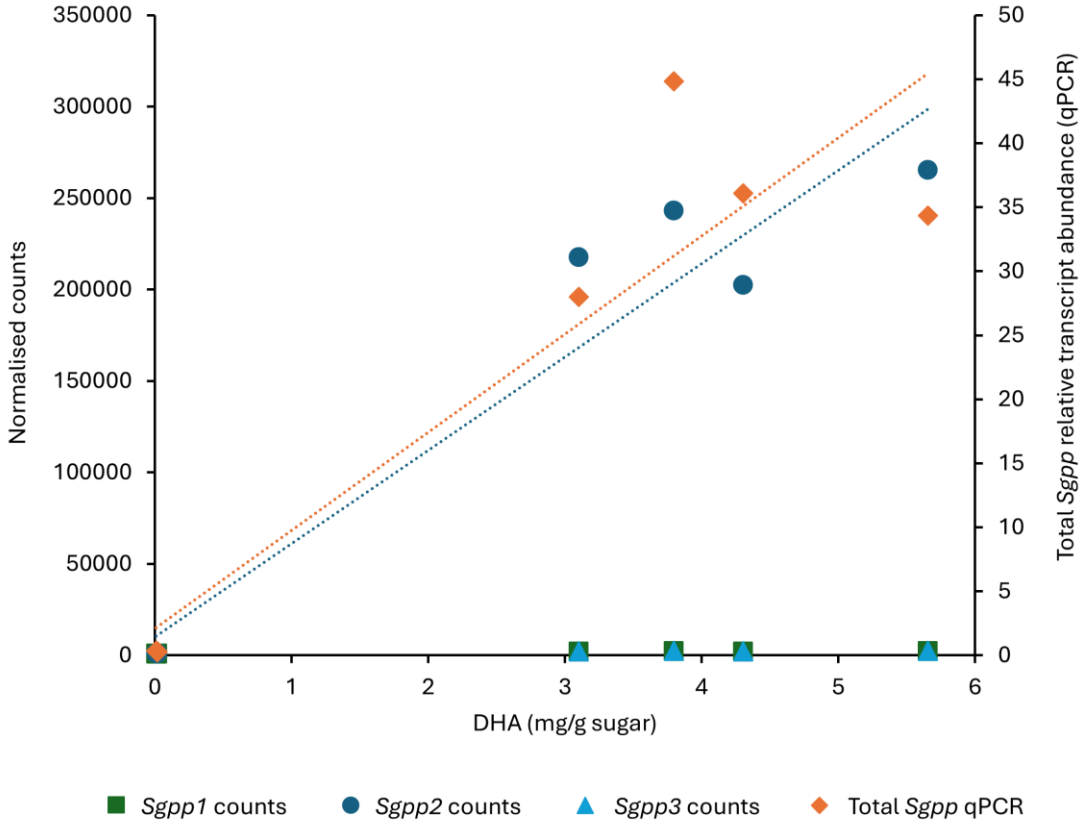


Figure 4.4: Gene expression data for *Leptospermum morissonii* and *Gaudium laevigatum* – mapped to their own genomes, and validated by qPCR – using primers designed to amplify all *Sgpps*. RNAseq normalised counts for the three *Sgpps* are indicated by the squares, circles and triangles, and use the left x axis. Total *Sgpps* as assayed by RT-qPCR are indicated by the diamonds, and use the right x axis.

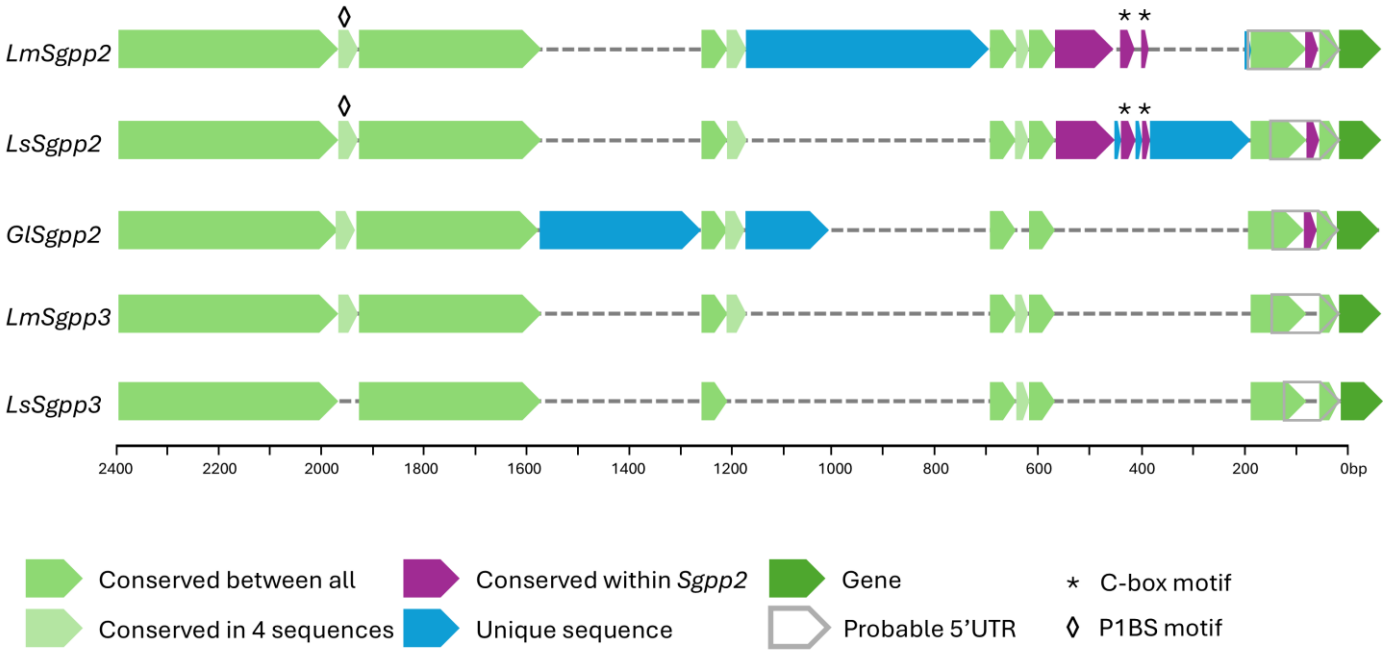


Figure 4.5: Approximately 1kb of the promoter sequence of *Leptospermum morrisonii, Leptospermum scoparium* and *Gaudium laevigatum Sgpp2* and *Sgpp3* sequences. Highlighted are the regions conserved within and between the promoters of each gene and structural differences (generally defined here as insertions/deletions over 15bp, except for 7bp shared between the two *Leptospermum Sgpp2s*).

4.3.4 *Sgpp* promoter sequences in *L. scoparium*, *L. morrisonii* and *G. laevigatum* have structural differences and unique motifs

There were similarities within *Sgpp2* promoters and between *Sgpp2* and *3* promoters (Fig. **4.5**). Three main regions are unique to the highly expressed *Sgpp2* promoters – meaning they are in both *Leptospermum Sgpp2* sequences and not in the *G. laevigatum Sgpp2* or either *Sgpp3* sequence. The unique regions are 169bp in total and start 348bp from first exon in *L. scoparium* and 188bp from first exon in *L. morrisonii*. They consist of 7bp, 40bp and 122bp conserved stretches, with two 16bp inserts between them (Fig. **4.5**). These regions contain multiple conserved motifs – notably two C-boxes (GACGTCA) and a GT1 (GRWAAW). The C-box was the only one of the two types of motifs identified that wasn't present in the promoter of *Sgpp3*s. There are multiple GT-1 binding sites within the first kb of promoter for each – 6, 9, and 11 in *LsSgpp2 GlSgpp2* and *LmSgpp2* respectively, and 8 and 11 in *LsSgpp3* and *LmSgpp3* respectively.

One region is also shared among all three *Sgpp2*s, located within the 5'UTR at 25bp from the first exon – it is 46bp long, which alters the length and content of the 5'UTR. This change could affect gene transcription and/or mRNA stability and translation efficiency (Srivastava *et al.*, 2018). The P1BS high affinity motif (GAATATTC) discussed previously in Grierson *et al*. (2024, Chapter 2) is also present

only in the proximal 1kb of the promoter of the highly expressed *Sgpp2*s (Fig. **4.5**) due to multiple SNPs. However, there are also multiple additional P1BS motifs (GNATATNC) further upstream in all five *Sgpp* sequences - between flanking genes and *Sgpp*s there are a total of three P1BS motifs in *GlSgpp2*, *LmSgpp2*, *LsSgpp3* and *LmSgpp3*, and four in *LsSgpp2* (not shown in Figure **4.5**, as they are further upstream).

4.3.5 bZIP transcription factor in QTL for nectar DHA

As we had identified C-box motifs that are bound by bZIP transcription factors, the RNAseq dataset described in Grierson *et al*. (2024) was searched for significantly differentially expressed bZIP type transcription factors. A bZIP gene (gene model g19850.t1) was identified as being significantly differentially expressed between high-DHA and low-DHA lines (adjusted $p = 0.0065$), with average normalised RNAseq read count of 3,120 in high-DHA lines and 2,266 in low-DHA lines. The gene model g19850.t1 Swiss-Prot annotation was OCS1_MAIZE – an OCS element binding factor, and the TAIR 10 annotation was AT4G34590.1 – a bZIP11 type transcription factor. The maize OCS1 has been demonstrated to bind to the CREB binding site – which is identical to the palindromic C-box/C-box (Singh *et al.*, 1990) that is present only in the highly expressed *Sgpp2* promoters.

The location of g19850.t1 was between 2,914,936bp and 2,915,396bp on chromosome 9 – which is within the Ls9 QTL for nectar DHA identified in Grierson *et al*. (2024 - Chapter 2). The co-location of this differentially expressed transcription factor with a QTL for nectar DHA suggests it has a role in DHA production, probably by enhancing or inducing the expression of *Sgpp2* through binding to the C-box motifs present in its proximal promoter.

4.3.6 Sequence of *Sgpp2* gene in *Leptospermum* genotypes

In the 14 genotypes sequenced for the *Sgpp2* gene (11 *L. scoparium* genotypes, five with two haplotypes and three *L. scoparium* crosses, Table **S2.1**), there was a high amount of sequence similarity between the lines, with the most variation found within the final intron. Within the *L. scoparium* genotypes, there were 98.9% identical sites in exons and 99.8% pairwise identity. In the whole gene, there were 83.6% identical sites and 96.2% pairwise identity. Most variability was observed in the final intron – with 58% identical sites and 88.8% pairwise identity. In the whole gene there were a total of 111 SNPs and indels. When the three *L. scoparium* crosses were included in the alignment there were 95.9% identical sites in the exons with 99.4% pairwise identity. In the whole gene there were 55.8% identical sites and 92.7% pairwise identity. Again most variation was observed in the final intron – with 8% identical sites. There were a total of 199 SNPs and indels. Introns can enhance gene expression and the changes in sequence of the final intron do alter the

motifs present. However, there is no obvious cosegregation between intron sequence and nectar DHA in the present dataset.

4.3.7 Promoter region in *L. scoparium* genotypes highly similar

In the 13 genotypes sequenced for the promoter (12 *L. scoparium* genotypes - six with two haplotypes, one *L. scoparium* cross, Table **S2.1**) there was a high similarity between the lines – with 96% identical sites and 99.4% pairwise identity in 1kb of the promoter and 93.2% identical sites and 98.9% pairwise identity in 3kb of the promoter. There were a total of 101 SNPs and indels in the 3kb of promoter. In the 169bp of conserved regions in the highly expressed Ls- and Lm- *Sgpp2*s conserved regions identified in the previous promoter sequence analysis (Fig. **4.5**), there were 96% identical sites and 99.2% pairwise identity. Only one sequence – one of two haplotypes in 1N41V – contained a SNP in one of the C-box motifs. There was no obvious cosegregation between nectar DHA and *Sgpp2* promoter sequence differences in this dataset.

4.3.8 *LmSgpp2* promoter induces nectary specific YFP in *Petunia* and *Nicotiana* transgenic lines

In both *Petunia* (Fig. **4.6**) and *Nicotiana* (Supplementary Fig. **S4.2**) transgenic lines, the *LmSgpp2* promoter directed strong YFP expression in the nectaries, indicating that elements within the 3.3kb of the promoter used are key to the strong nectary specific expression of *Sgpp2*. The developmental series in *Petunia* transgenic lines showed YFP fluorescence in the nectaries of all three stages, with a peak in fluorescence in stage 2 nectaries, just prior to nectar secretion and corolla opening (Fig. **4.6 a**, **b** and **c**). The *LmSweet9promoter* transformed lines had variable fluorescence – with the majority being faintly fluorescent in the vasculature rather than in the nectary (Fig. **4.6d** and **e**, Supplementary Fig. **S4.3**)

In the *LmSgpp2promoter-Venus* transformed *Petunia* lines the transgene was detected in 15 out of 20 lines, and in the *LmSweet9promoter-Venus* transformed lines, 8 out of 20 had the transgene detected. In the *LmSgpp2promoter-Venus* transformed *Nicotiana* lines the transgene was detected in 5 out of 5 lines assessed, and in the *LmSweet9promoter-Venus* transformed lines assessed, 3 out of 3 had the transgene detected. No transgenes were detected in the regeneration controls of either species.

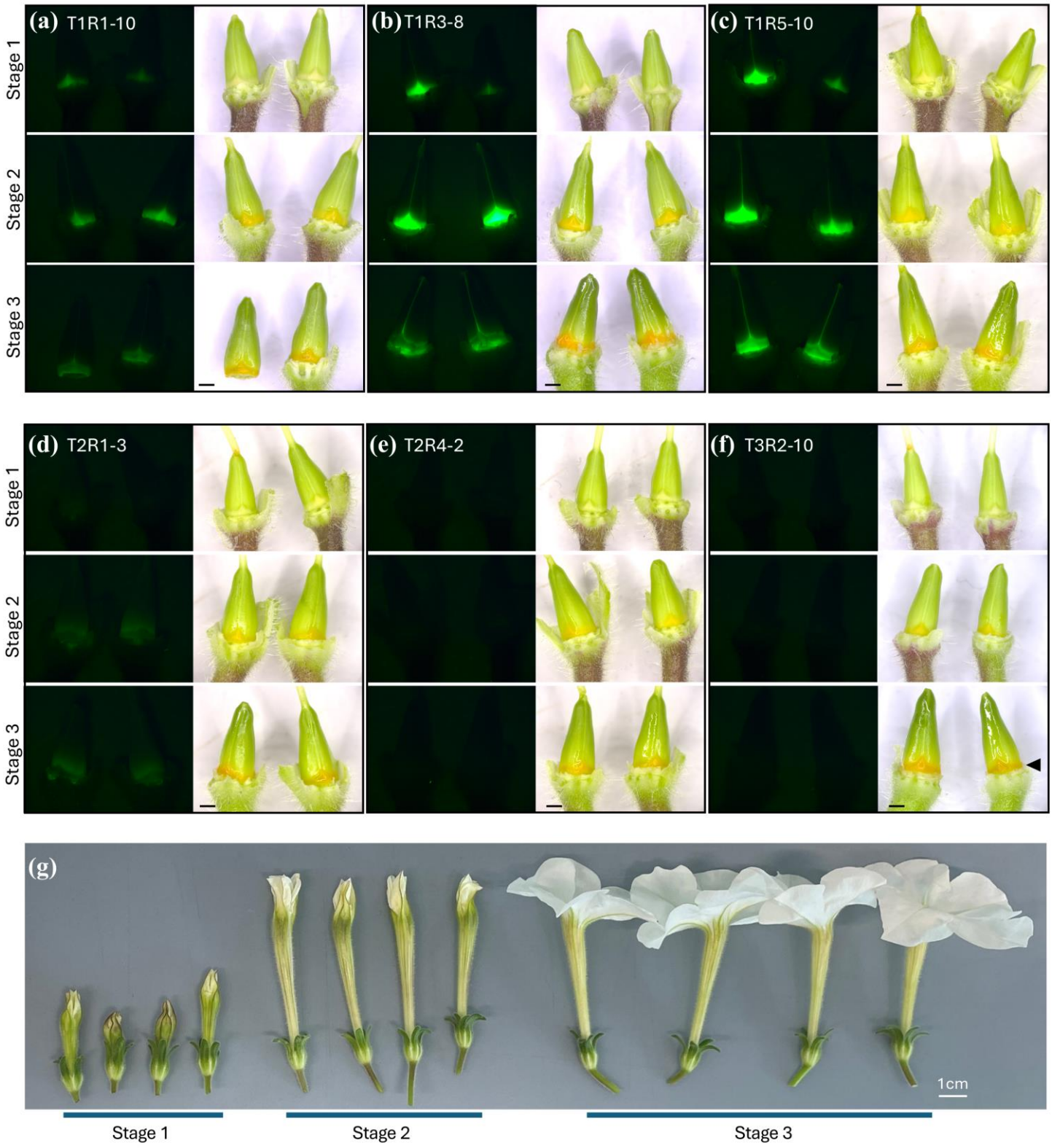


Figure 4.6: Nectaries of transformed *Petunia* at three developmental stages under white and blue light (with YFP filter). (a, b and c) *LmSgpp2 promoter – Venus* transformed lines at stages 1, 2 and 3, scale bar 1mm. (d,e) *LmSweet9 promoter – Venus* transformed lines at stages 1, 2 and 3. scale bar 1mm (f) Untransformed regeneration control line at stages 1, 2 and 3. (g) Example of floral developmental stages, scale bar 1cm. Nectary tissue (orange) indicated with arrow in (f).

4.4 Discussion

4.4.1 Presence of *Sgpp2* probably ancestral in Myrtaceae – with multiple loss of function events

Within the Leptospermeae tribe, *L. morrisonii* was found to have a *Sgpp* locus structure and gene expression patterns similar to *L. scoparium* (Grierson *et al*., 2024, Chapter 2), while *G. laevigatum* was found to contain only two *Sgpp* copies, and the *GlSgpp2* was not highly expressed (Fig. **4.1** and **4.4**). The absence of a highly expressed *Sgpp2* in *G. laevigatum* could explain the lack of DHA in that species. Establishing that *Sgpp2* expression correlates with DHA amounts in two further Leptospermeae species provides further evidence that this gene is responsible for DHA production. *Sgpp* gene phylogeny indicates that presence of the *Sgpp2* gene is the ancestral state of the Myrtaceae lineage (Fig. **4.3**). Furthermore, a third *Sgpp* copy occurs within the *Leptospermum* and *Eucalyptus* (Fig. **4.2**). Each *Sgpp3* clades more closely with the *Sgpp2* within its genus, so is therefore likely to have resulted from independent duplication events within each of these lineages.

Leptospermeae and Chamelaucieae are strongly supported sister tribes (e.g Wilson *et al.*, 2001), and the presence of the nectar DHA trait has so far been detected within members of seven genera (e.g. Fig. **4.3**). We hypothesise that the last common ancestor of these sister tribes probably had a highly expressed *Sgpp2*, and this high expression has been subsequently lost repeatedly throughout these clades. We do not currently have genomic sequence for many members of these DHA producing species, and none for those outside the *Leptospermum* genus. Further investigation into *Sgpp* gene number and expression in these species could help to validate this hypothesis. Interesting genera to examine would be *Verticordia* and *Aggreflorum*, where nectar testing to date indicates both the presence and absence of nectar DHA within each genus (Williams *et al.*, 2018; Obeng-Darko *et al.*, 2023). The repeated loss of this trait is the most parsimonious explanation and makes sense if selection for it varies in its pressure in different environments. While we have hypothesised that DHA in nectar is a byproduct of phosphate homeostasis, there may be other adaptive significance that is still unknown.

4.4.2 bZIP transcription factor could be involved in the high expression of *Sgpp2*

The promoters of *Sgpp2* and *Sgpp3* share many conserved regions (Fig. **4.5**), so the elements contributing to the high nectary-specific expression of *Sgpp2* are likely to be in the few regions unique to its promoter. The identification of two C-box motifs within these sequences could explain the high expression of these *Sgpp2*s. C-boxes are a type of ACGT box bound by bZIP type transcription factors, which have diverse roles that include involvement in photomorphogenesis,

signal transduction, and specialised metabolism (Han *et al.*, 2023). The bZIP family is one of the largest transcription factor families in plants and is involved in a range of pathways, including those involved in light response and hormone signaling. The differentially expressed bZIP transcription factor identified (gene model g19850.t1) is annotated with an OCS element binding factor that has been experimentally demonstrated to bind C-box motifs (Singh *et al.*, 1990) identical to those identified in the *LsSgpp2* promoter region. Whether the bZIP11 transcription factor does indeed interact with the C-box motif would need to be demonstrated experimentally, but the fact it is significantly differentially expressed and located in a QTL for *L. scoparium* nectar DHA suggests it could be involved.

The transcription factor identified is a homolog of AT4G34590 (bZIP11/ocs1) and its highest expression in *Arabidopsis* is in flowers (Arabidopsis Information Resource - TAIR). It belongs to the S1 subgroup of bZIPs that have roles in plant growth and development – especially seed maturation, root growth and floral development (Wang *et al.*, 2022). S1-bZIP overexpression induces sugar related gene expression and increased sugar content (Thalor *et al.*, 2012; Wang *et al.*, 2022). *Sucrose synthase* transcripts increased in bZIP11 overexpression lines (Thalor *et al.*, 2012), and expression of bZIP11 has been shown to be involved in activation of genes involved in metabolism of minor carbohydrates such as myo-inositol and raffinose (Ma *et al.*, 2011). In the RNAseq dataset from Grierson *et al*., (2024, Chapter 2), *Sweet9*, *CWIN* and *SUS4* – *Sucrose Synthase* are among the top 10 most highly expressed genes– demonstrating that carbohydrate metabolism pathways are highly active in this tissue. S1-bZIPS are unique as their 5'-mRNA leader contains a conserved upstream open reading frame (uORF). The translation of the main open reading frame (mORF) is mediated by sucrose concentrations: high sucrose levels enhance ribosome stalling on the uORF, which results in reduced translation of the mORF (Rahmani *et al.*, 2009). Our RNAseq analysis measures the transcript abundance within the hypanthium tissue, but we do not know how translation of the bZIP11 transcripts would be affected by sucrose concentrations within the hypanthium tissue, or more specifically, within the nectary tissue itself.

4.4.3 Highly similar promoter sequences within *Leptospermum* genotypes with range of DHA phenotypes

Within the *Leptospermum* genotypes with varying DHA phenotypes we found very little variation in the *Sgpp2* promoter, or the remainder of the gene. However, all promoter sequences in the analysis are from DHA producing lines – varying in the amount of DHA present rather than the binary presence or absence of DHA. Expression of *Sgpp2* is relatively high even within low DHA lines – average normalised counts were still higher than that of *Sweet9* within the low-DHA transcripts

(Supplementary Table **S2.3** within Grierson *et al*. 2024, Chapter 2). Further attempts to sequence the promoters in species producing no DHA would help to establish if the regions and motifs we have identified are essential for the expression of *Sgpp2*.

The lack of variation in the promoter coupled with the co-location of the *Sgpp* locus with a QTL for nectar DHA suggests more complex transcriptional regulation of *Sgpp2* expression based on DNA elements located beyond the promoter region analysed here. Fine tuning of expression may be conferred by another element within the QTL but not within the proximal promoter, such as a distal *cis* enhancer contributing to regulation through chromatin looping. Another possibility is that the colocation of the *Sgpp* with a QTL is coincidental – and there could be another gene within the QTL involved that we have not identified.

4.4.4 *Sgpp2* promoter able to be induced in diverse lineages

The *LmSgpp2* promoter strongly expressed in the nectaries of *Petunia* and *Nicotiana* transgenic lines, despite them being from a plant family relatively distant to the Myrtaceae. This indicates that whatever elements are conferring this strong and nectary specific expression are conserved across species. We see variability in the intensity of fluorescence, but there appears to be a pattern of a peak in intensity in stage 2 nectaries, which coincides with nectar production beginning just prior to flower opening (Ge *et al.*, 2000). bZIPs contribute to sugar metabolism and upregulate genes such as *Sucrose Phosphate Synthase* (*SPS*, Wang *et al.*, 2022) and genes required for sucrose biosynthesis are known to be upregulated in nectaries (Kram *et al.*, 2009), and could therefore be expressed in *Petunia* nectaries at the start of nectar secretion. If the bZIP11 transcription factor and C-box motifs identified are a key driver of expression, this could explain why the promoter can function in a distant lineage. The *LmSweet9* promoter did not work well as a control – this could have been due to a lack of conservation between the distant plant families, or a larger part of the promoter may be required due to structural differences compared with the endogenous *Sweet9*/*Nec1*.

4.4.5 Complex regulation of *Sgpp2* expression

These results may further explain why nectar DHA is present or absent in Leptospermeae species and reveal some of the mechanisms behind *Sgpp2* gene expression. The presence of the C-box motifs may provide the potential for expression, which is then induced by the bZIP11 transcription factor. These elements may be essential for expression of *Sgpp2* and the presence of DHA, and elements such as the P1BS motifs may be part of the fine tuning of gene expression, potentially in response to fluctuating Pi demands. The strong nectary fluorescence in the *LmSgpp2promoter-Venus* transgenic lines support the idea that a strong base amount of *Sgpp2* expression is turned on by elements

within the 3.3kb of promoter, and the C-box motif helps explain the activation of the *Sgpp2* promoter in distantly related lineages – as bZIP transcription factors are evolutionarily conserved and are involved in sucrose metabolism in many species (Wang *et al.*, 2022).

The lack of variation in the promoter among DHA-producing *Leptospermum* genotypes indicates that other elements beyond this region must also be affecting gene expression. The identification of multiple QTLs and the consistency of relative DHA amounts within genotypes across seasons indicates genetic control of variation in the DHA trait (Grierson *et al*., 2024, Chapter 2), but the specific mechanisms behind this genotypic variation are unexplained by the present data. Potentially a *cis* element beyond the proximal promoter but within the Ls4 QTL could contribute to regulation thorough chromatin looping. Another possibility could be variation in the bZIP11 transcription factor and its expression or expression induction capabilities. There could also be regulatory factors not explored here – such as chromatin accessibility or microRNA-triggered degradation.

The presence of the high affinity P1BS motif within the first 1kb or the promoter as well as within the first intron of *Sgpp2* supports the proposed role of *Sgpp2* maintaining Pi homeostasis during nectar production and secretion in a photosynthesising nectary (see Grierson *et al*. 2024, Chapter 2). All five *Sgpp2* and *3* sequences contain 2 or 3 P1BS elements further upstream, which further supports this and suggests that they may be part of fine-tuning expression rather than baseline expression. As this locus is enriched for the GT-1 motif involved in light responsive pathways (Terzaghi & Cashmore, 1995), this could support this proposed role.

4.4.6 Concluding comments

These results provide further evidence that *Sgpp2* is responsible for DHA accumulation in *Leptospermum*. At least two *Sgpp* gene copies are present within most members of the Myrtaceae family examined here. A highly expressed *Sgpp2* appears to be linked to the presence of DHA and is likely the ancestral trait within the Leptospermeae and Chamelaucieae tribes. The nectar DHA trait is dispersed throughout these tribes and therefore may have been repeatedly lost, potentially due to changes in selection pressure. The combination of two C-box transcription binding sites within the proximal promoter of highly expressed *Sgpp2*s, along with the presence of a significantly differentially expressed bZIP11 type transcription factor co-locating with the Ls9 QTL, could be responsible for the expression of *Sgpp2* and subsequently the presence of DHA. The lack of variation in the proximal promoter within DHA producing *Leptospermum* genotypes indicates that further variation of expression levels is likely conferred by other distal regulatory elements. Future work should confirm that the *Sgpp2* enzyme is capable of dephosphorylating DHAP in comparison with the other *Sgpps*. Apart from *G. laevigatum*, all the genotypes sequenced here produce DHA in varying

quantities – thus the inclusion of a wider range of true no-DHA lines could help to clarify the role of the C-box motifs. It would be interesting to further investigate the *Aggreflorum* and *Verticordia* genera – as they contain both species with some of the highest nectar DHA levels observed, and species with no nectar DHA (Williams *et al.*, 2018; Obeng-Darko *et al.*, 2023). Further transgenic experiments with a *LmSgpp2promoter-Venus* construct with the C-box elements removed could also confirm the role of the C-box motifs in strong baseline expression.

4.5 References

Bartel DP. **2009**. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**: 215–233.

Beernink BM, Vogel JP, Lei L. **2025**. Enhancers in Plant Development, Adaptation and Evolution. *Plant and Cell Physiology* **66**: 461–476.

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**: 348–359.

Clearwater MJ, Noe ST, Manley-Harris M, Truman G-L, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. **2021**. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. **2013**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

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

Gambino G, Perrone I, Gribaudo I. **2008**. A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* **19**: 520–5.

Ge Y-X, Angenent GC, Wittich PE, Peters J, Franken J, Busscher M, Zhang L-M, Dahlhaus E, Kater MM, Wullems GJ, *et al.* 2000. NEC1, a novel gene, highly expressed in nectary tissue of *Petunia hybrida*. *The Plant Journal* **24**: 725–734.

Grierson ERP, Thrimawithana AH, van Klink JW, Lewis DH, Carvajal I, Shiller J, Miller P, Deroles SC, Clearwater MJ, Davies KM, *et al.* 2024. A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*). *New Phytologist* **242**: 2270–2284.

Han H, Wang C, Yang X, Wang L, Ye J, Xu F, Liao Y, Zhang W. **2023**. Role of bZIP transcription factors in the regulation of plant secondary metabolism. *Planta* **258**: 13.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. **1999**. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* **27**: 297–300.

Kram BW, Xu WW, Carter CJ. **2009**. Uncovering the *Arabidopsis thaliana* nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. *BMC Plant Biology* **9**: 92.

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. **2002**. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* **30**: 325–327.

Ma J, Hanssen M, Lundgren K, Hernández L, Delatte T, Ehlert A, Liu C-M, Schluepmann H, Dröge-Laser W, Moritz T, *et al.* 2011. The sucrose-regulated *Arabidopsis* transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism. *New Phytologist* **191**: 733–745.

Marand AP, Eveland AL, Kaufmann K, Springer NM. **2023**. *cis*-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annual Review of Plant Biology* **74**: 111–137.

Obeng-Darko SA, Sloan J, Binks RM, Brooks PR, Veneklaas EJ, Finnegan PM. **2023**. Dihydroxyacetone in the Floral Nectar of *Ericomyrtus serpyllifolia* (Turcz.) Rye (Myrtaceae) and *Verticordia chrysantha* Endl. (Myrtaceae) Demonstrates That This Precursor to Bioactive Honey Is Not Restricted to the Genus *Leptospermum (*Myrtaceae). *Journal of Agricultural and Food Chemistry* **71**: 7703–7709.

Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J. **2009**. Sucrose Control of Translation Mediated by an Upstream Open Reading Frame-Encoded Peptide. *Plant Physiology* **150**: 1356–1367.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. **2012**. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* **61**: 539–542.

Schmitz RJ, Grotewold E, Stam M. **2022**. *Cis*-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell* **34**: 718–741.

Singh K, Dennis ES, Ellis JG, Llewellyn DJ, Tokuhisa JG, Wahleithner JA, Peacock WJ. **1990**. OCSBF-1, a maize ocs enhancer binding factor: isolation and expression during development. *The Plant Cell* **2**: 891–903.

Smale ST, Kadonaga JT. **2003**. The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry* **72**: 449–479.

Smallfield BM, Joyce NI, van Klink JW. **2018**. Developmental and compositional changes in *Leptospermum scoparium* nectar and their relevance to mānuka honey bioactives and markers. *New Zealand Journal of Botany* **56**: 183–197.

Srivastava AK, Lu Y, Zinta G, Lang Z, Zhu J-K. **2018**. UTR dependent control of gene expression in plants. *Trends in Plant Science* **23**: 248–259.

Stewart CN, Via LE. **1993**. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* **14**: 748–750.

Terzaghi WB, Cashmore AR. **1995**. Light-Regulated Transcription. *Annual Review of Plant Biology* **46**: 445–474.

Thalor SK, Berberich T, Lee SS, Yang SH, Zhu X, Imai R, Takahashi Y, Kusano T. **2012**. Deregulation of Sucrose-Controlled Translation of a bZIP-Type Transcription Factor Results in Sucrose Accumulation in Leaves. *PLOS ONE* **7**: e33111.

Wang H, Zhang Y, Norris A, Jiang C-Z. **2022**. S1-bZIP Transcription Factors Play Important Roles in the Regulation of Fruit Quality and Stress Response. *Frontiers in Plant Science* **12**.

Williams SD, Pappalardo L, Bishop J, Brooks PR. **2018**. Dihydroxyacetone Production in the Nectar of Australian *Leptospermum* Is Species Dependent. *Journal of Agriculture and Food Chemistry* **66**: 11133–11140.

Wilson PG, Heslewood MM. **2023**. Revised taxonomy of the tribe Leptospermeae (Myrtaceae) based on morphological and DNA data. *TAXON* **72**: 550–571.

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

Chapter 5: Enzymatic characterisation of the *Leptospermum morrisonii* SGPP2

5.1 Introduction

Species within the *Leptospermum* genus are woody perennials, many of which accumulate DHA in their nectar – a trait that is unique to some species within the Myrtaceae family (Obeng-Darko *et al.,* 2023; Wilson & Heselwood, 2023). The accumulation of DHA forms the basis of mānuka honey industry in Aotearoa-New Zealand, as it accumulates in the nectar of *Leptospermum scoparium* and results in formation of the antimicrobial compound methylglyoxal in the honey.

A haloacid dehalogenase (HAD) superfamily phosphatase protein, SGPP2, has recently been linked to nectar DHA accumulation in *Leptospermum* (Grierson *et al.*, 2024, Chapter 2). It was hypothesised that this enzyme was responsible for DHA accumulation in *Leptospermum* nectar through the dephosphorylation of dihydroxyacetone-phosphate (DHAP), an intermediate of primary metabolic pathways, such as glycolysis, gluconeogenesis and the Calvin Cycle. Nectar DHA in mānuka has been recently demonstrated to be derived from cellular pools of DHAP – mediated by photosynthesis within the nectary (Clearwater *et al.*, 2021).

HAD superfamily members are largely phosphatases, and are distributed across Bacteria, Arachaea and Eukaryota (Burroughs *et al.*, 2006). They are defined by presence of an α/β-core domain with four loops containing conserved residues (motifs 1-4) that make up the active site (Lahiri *et al.*, 2004; Burroughs *et al.*, 2006). Many HADs also contain a cap domain which acts as a dynamic lid over the active site and contains a substrate specificity loop (motif 5, Lahiri *et al.*, 2004). Diversification within this cap domain is largely responsible for expansion of the range of substrates this family can act on (Lahiri *et al.*, 2004; Burroughs *et al.*, 2006). Cap types are classified intro three types – C0 (no cap), C1 (inserted between motif 1 and 2) and C2 (inserted between motifs 2 and 4) (Lu *et al.*, 2005). This ability to diversify substrate specificity provides a potential basis for neofunctionalisation during evolution (Kuznetsova *et al.*, 2015; Pandey *et al.*, 2017; Du *et al.*, 2021). Gene duplication is commonly the first step underpinning neofunctionalisation – and *Leptospermum* such as *L. scoparium* and *L. morrisonii* have three *Sgpp* paralogues with different transcript abundance patterns (Grierson *et al*., 2024, Chapter 2).

The *Arabidopsis* orthologue of the *Leptospermum* SGPPs (AtSGPP) has been identified as a HAD hydrolase phosphomonoesterase of subfamily I (C1 type cap) and has had extensive enzymatic

characterisation (Caparrós-Martín *et al.*, 2013, 2014, 2019). Caparrós-Martín *et al.* (2013) showed AtSGPP exhibits dephosphorylating activity on a broad range of substrates, however DHAP was not among substrates assayed. Subsequent work investigated how changes to the substrate specificity motif (Caparrós-Martín *et al.*, 2014) and other sequence determinants (Caparrós-Martín *et al.*, 2019) affected substrate specificity and enzyme kinetics.

AtSGPP has been proposed to have roles in stress tolerance and Pi homeostasis (Caparrós-Martín *et al*., 2013). The *Leptospermum* genus is centred in Australia, which has a high proportion of nutrient poor soils (Orians & Milewski, 2007). *Leptospermum scoparium* can tolerate low phosphate soils (Gutiérrez-Ginés *et al.*, 2019; Noe, 2022), and Noe (2022) found that nectar production and growth was not affected by low phosphate treatments. Phosphate homeostasis is crucial for cell function, especially under stress and during photosynthesis – to protect photosynthetic machinery (Bustos *et al.*, 2010; Nilsson *et al.*, 2012). Based on the proposed function of AtSGPP, we hypothesised that LmSGPP2 would have similar function of Pi homeostasis during photosynthesis and sugar production and secretion in the nectary, and was potentially of benefit under low phosphate conditions (Grierson *et al*., 2024, Chapter 2). Confirmation of SGPP2s ability to dephosphorylate DHAP to produce DHA and phosphate would support this proposed role. Further evidence that this gene is responsible for DHA production in mānuka would also give more confidence in developing markers and other biotechnologies.

We report here enzymatic characterisation of the three *Leptospermum* SGPPs, LmSGPP1 LmSGPP2 and LmSGPP3. These three enzymes, along with the *Arabidopsis* SGPP, were produced and purified in microbial systems, and assayed for phosphatase activity on three substrates, including DHAP. We aimed to analyse in detail the activity of SGPP2 on DHAP and compare the phosphatase activity and substrate specificity between the *Leptospermum* SGPPs and AtSGPP. We found that LmSGPP2 was able to dephosphorylate DHAP, and we hypothesise that shifts in kinetic parameters between the SGPPs could be due to changes in the substrate specificity loop.

5.2 Methods

5.2.1 Expression constructs

Coding sequences were obtained for AtSGPP (accessed from TAIR, Gene ID - AT2G38740), and the three SGPPs from *L. morrisonii* genome (Chagné, Deng & Grierson – Chapter 3). The *L. morrisonii* genes were used, as we do not have approval to use Aotearoa-New Zealand native gene sequences in transgenic experiments. *L. morrisonii* also produces nectar DHA and has high *Sgpp2* expression (Chapter 4). Furthermore, *L. morrisonii* and *L scoparium* SGPPs deduced amino acid sequences have

high amino acid conservation, with 98.37% (SGPP1), 95.92% (SGPP2) and 97.14 % (SGPP3) shared identity, so findings here should apply to *L. scoparium* as well as *L. morrisonii*. Coding sequences were synthesised by Genscript (USA) and ligated into pMAL-c4x – to enable fusion proteins with maltose-binding protein (MBP) to be expressed. pMAL-c4x constructs were cloned into BL21(DE3) *Escherichia coli* lines. The expected molecular weights for each fusion protein were 69.70kDa for MBP-AtSGPP, 69.73kDa for MBP-LmSGPP1, 69.76kDa for MBP-LmSGPP2 and 69.73kDa for MBP-LmSGPP3.

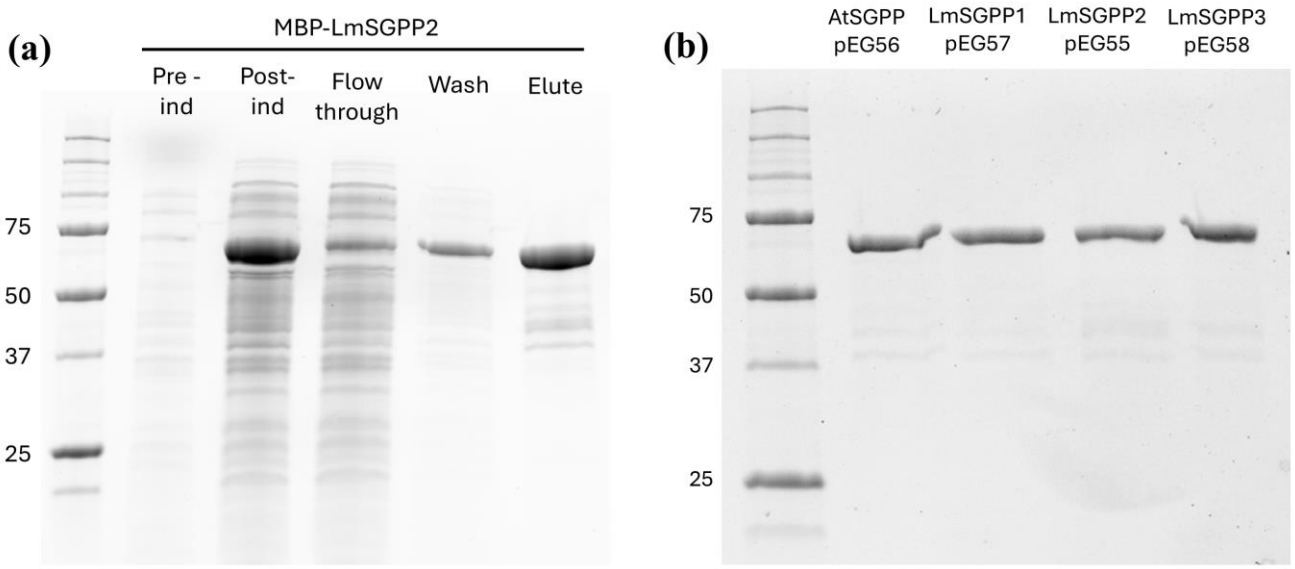


Figure 5.1: Purification of SGPP fusion proteins. (a) MBP-LmSGPP2 production and purification. Contains pre-induction cells and post induction cells, followed by post sonication supernatant flow through, wash, and elution fractions. (b) Purified fusion proteins MBP-AtSGPP, MBP-LmSGPP1, MBP-LmSGPP2 and MBP-LmSGPP3. Approx 2µg of each. Molecular weight markers are in kDa.

5.2.2 Purification of recombinant proteins

For each of the fusion protein plasmids, starter cultures were initiated from beadstocks with one bead in 5ml of LB + Amp selection (50µg/mL) and grown at 37°C for approximately 16hrs. 300ml cultures in 2L flasks were inoculated with 900µL of each of the overnight cultures. Cultures were then grown for 6-8 hours at 28°C until A_{600} was approximately 0.5. Cultures were then induced with 150µL of 1M IPTG for a final concentration of 500µM and induced for ~18 hrs at 28°C. Cells were then harvested at 10,000xg and pellets frozen directly at -20°C.

Fusion proteins were released from the harvested cells by resuspension in 24ml of column buffer (20mM Tris-HCl, 200mM NaCl, 1mM EDTA, pH7.4), and sonication in an icebath with a probe sonicator (Cgoldenwall) for approx. 1.5 mins total (3 mins of 1-2 second bursts and 1-2 second rests).

The fusion proteins were purified using 1ml MBP-Trap columns (Cytiva) as per manufacturers' instructions, and eluted in column buffer + 10mM maltose, and 200µL aliquots frozen directly at -80°C. Proteins were quantified using the Nanodrop 1000 (Thermo Fisher), and size and purity assessed using SDS-PAGE electrophoresis (Biorad). Miniprotein gels (Biorad), Laemelli loading buffer (Biorad) and unstained protein standards (Precision Plus, BioRad) were used and gels were run at 200V 0.35A for 30 mins. To assess protein induction and purification protocols, pre induction samples (~$4x10^7$ cells) and post induction samples (~$8x10^7$cells) were loaded in each well and 10ul of the flow through, wash and elution was also loaded (Fig. **5.1a**). When all final fusion proteins were run together to compare (Fig. **5.1b**), approx. 2µg of each was loaded.

5.2.3 Activity assays

To assess the phosphatase activity of each protein, the first substrate screening used D-ribose-5-phosphatye (R5P), DL-glycerol-3-phosphate (G3P), and dihydroxyacetone phosphate (DHAP) (R7750, 94124 and D7137 Sigma respectively). Reactions contained 25µg/mL purified protein, 20mM Tris-HCl, 5mM $MgCl_2$ and 10mM of substrate (R5P, G3P or DHAP) in a volume of 80µL at 32°C. At 15, 30 and 60 minutes 8µL was taken from each reaction and diluted 200-fold to assay the phosphate released. Phosphate released was assayed using Malachite Green Phosphate Assay Kit (Sigma-Aldrich, MAK307) as per manufacturer's instructions in a Nunclon™ Delta surface 96 well plate (Thermo Fisher). Samples were run in duplicate within each plate, and the plate was incubated at room temperature for 30 minutes before reading absorbance at 620nM on a Spectramax ABS plate reader (Molecular Devices). Phosphate standards between 0-40µM phosphate were run with each plate, as well as denatured protein and substrate only controls. Due to the instability of DHAP and G3P and the trace presence of phosphate in the substrate (0.03 mol mol^{-1} in the DHAP), the value from the substrate only control was subtracted from the calculated phosphate value for those samples at each time point. Each plate was run twice.

To establish Michaelis-Menten kinetic parameters, eight substrate concentrations between 0.63 and 160mM were used (exact concentrations varied depending on protein-substrate combination), and the reaction sampled between 4-5 minutes. If a sample was outside the linear range, the reactions were repeated with a higher dilution factor. Dilution factors for the assay between 120-200x were used depending on protein-substrate combination. Each plate was run twice. Initial velocities for LmSGPP2 and LmSGPP3 were calculated between 4-5 minutes at 10mM. A_{620} values were converted to µM phosphate using the standard curve, and the concentration of phosphate in the initial reaction was calculated by using the dilution factor for each reaction.

5.2.4 Analysis of assay results

Substrate screening data were log transformed for statistical analysis to equalise variance and the data were analysed with ANOVA using a split-plot in time structure account for the repeated measures of each sample. Results from 15 and 30 minutes were analysed, and the 60 minute values excluded because they were inconsistent – potentially due to the long incubation. Post-hoc comparisons were made using 5% Fisher's LSDs in Genstat (24th edition). The untransformed values were used for figures.

Michaelis-Menten kinetic analysis was done in Origin (2022b) using non-liner curve fitting with Levenberg Marquardt iteration, until converged. The equations used were as follows:

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

$$k_{cat} = \frac{V_{max}}{[ET]}$$

Where V_0 is the initial velocity, V_{max} the maximum velocity, [S] is the substrate concentration, K_m is the Michaelis-Menten constant for the substrate, [ET] is the protein concentration, and k_{cat} is the turnover rate. K_{cat}/K_m – catalytic efficiency was calculated from those values.

5.3 Results

5.3.1 Fusion proteins produced

The four fusion proteins were successfully produced and purified and were the expected molecular weights (Fig. **5.1**).

5.3.2 SGPP phosphatase sequence comparisons

The four SGPPs have between 66.39 to 95.51% identical sequence (Fig. **5.2**). Motifs 1-4 contain conserved catalytic elements (Burroughs *et al.*, 2006; Caparrós-Martín *et al.*, 2019). Loops 1-4 are responsible for binding and catalysis, and loop 5 is involved in substrate recognition. We see amino acid changes in the substrate specificity motif at positions 68 (LmSGPP2) and 69 (LmSGPP2 and 3) (Fig. **5.2**). At position 68 – the shift from isoleucine (I) to methionine (M) does not alter the charge, polarity or hydrophobicity and has a shift in size from 131.09Da (I) to 149.05Da (M). The position 69 shift from alanine (A) to serine (S) is a shift from a polar hydrophobic residue to a polar hydrophilic uncharged residue – and a size change from 89.05Da (A) to 105.04Da (S).

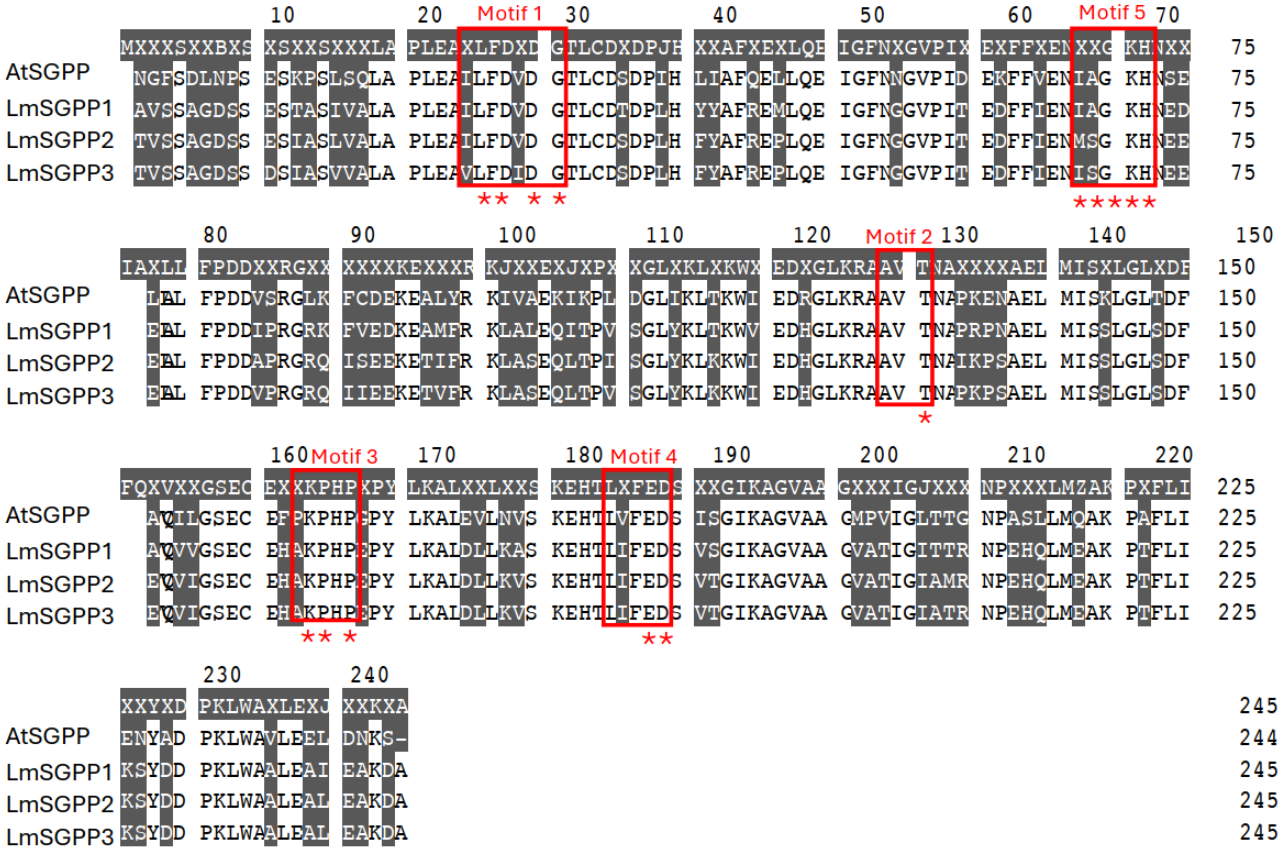


Figure 5.2: Alignment of *Arabidopsis* AtSGPP, and *Leptospermum morrisonii* LmSGPP1, LmSGPP2 and LmSGPP3 amino acid sequences. Motifs 1-5 are indicated by red boxes, and structural elements are indicated by red stars. Non-conserved amino acids are highlighted in grey. Figure modified from Geneious (v2022.0.1) output.

Between the two pairs of SGPPs that are predicted to be most similar in activity based on sequence similarity (AtSGPP and LmSGPP1, LmSGPP2 and LmSGPP3, Fig. **5.2**), there are nine amino acid changes that are shared within each pair, and differ between the pairs (positions 69, 90,97,107,117,137,153,192 and 208, Fig. **5.2**), but only one (position 69) is in a key motif. LmSGPP2 and 3 are 95.51% identical at the amino acid level and have three changes between them in key motifs (motif 1 positions 25 and 29, motif 5 position 68, Fig. **5.2**), but only the position 68 change in motif 5 is considered to be in a structural residue (Fig. **5.2**, Caparrós-Martín *et al.*, 2019).

5.3.3 LmSGPP2 can dephosphorylate DHAP

Phosphate was produced when LmSGPP2 was incubated with DHAP, indicating it is able to dephosphorylate DHAP to produce DHA and phosphate (Fig. **5.3**). The amounts of phosphate produced by LmSGPP2 were not significantly different to LmSGPP3 (p>0.05) and both LmSGPP2 and SGPP3 released more total phosphate off DHAP than either LmSGPP1 or AtSGPP (p<0.05). There was no significant difference between AtSGPP and LmSGPP1 (p>0.05) but there was a significant difference between the two pairs (AtSGPP and LmSGPP1 vs LmSGPP2 and LmSGPP3, p<0.05).

5.3.4 Activity on R5P and G3P substrates

To further assess the differences in activity between the SGPPs, they were screened on two further substrates – R5P and G3P (Fig. **5.4**). AtSGPP, LmSGPP1 and LmSGPP3 had similar activity on R5P – which was significantly higher than LmSGPP2 activity on R5P ($p<0.05$). There was no significant difference between the activity of LmSGPP2 and LmSGPP3 on G3P, and they were both higher than AtSGPP and LmSGPP1.

5.3.5 Kinetic analysis

Initial screening results were then followed up with more detailed kinetic analysis where possible. For the substrate R5P - AtSGPP and LmSGPP1 and LmSGPP3 were found to have similar Michaelis-Menten kinetics (Table **5.1**). They have similar V_{max} (9.91-12.17 µM/s) and k_{cat} (27.64-33.93s^{-1}) values but LmSGPP3 has a higher K_m (approx. twice that of AtSGPP), resulting in a lower catalytic efficiency (k_{cat}/K_m approx. half of AtSGPP). LmSGPP2 has a lower V_{max} (approx. half that of the other three), a higher K_m (approx. 4 times AtSGPP), and consequently a lower turnover rate (k_{cat}) and overall catalytic efficiency (k_{cat}/K_m approx.10% of AtSGPP and 20% of SGPP3).

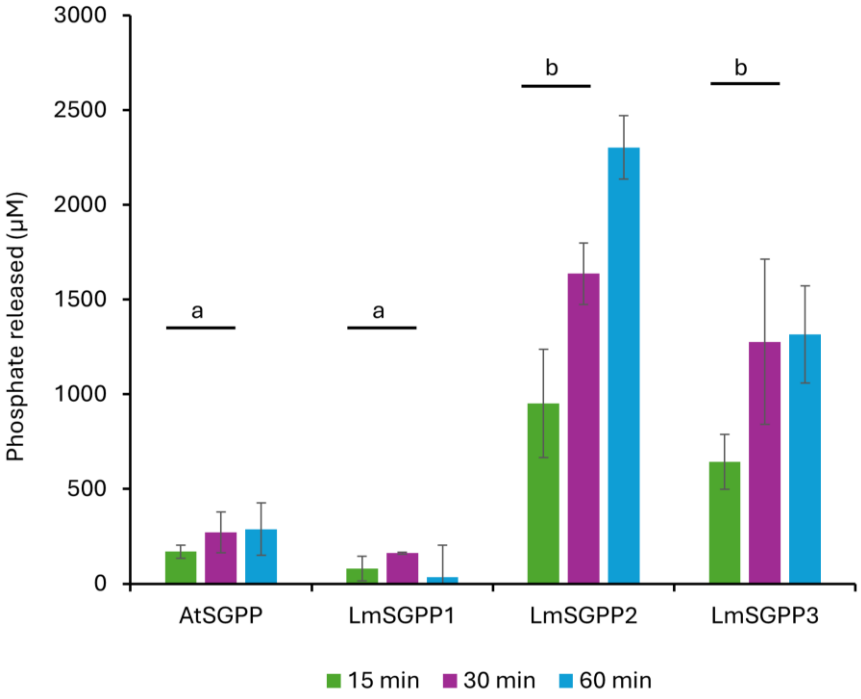


Figure 5.3: Phosphate released (µM) from dihydroxyacetone phosphate (DHAP) at three time points. The reaction mixture containing 10mM DHAP, 25µg/mL of purified protein (MBP-AtSGPP, MBP-LmSGPP1, MBP-LmSGPP2 or MBP-LmSGPP3), 20mM Tris-HCl (pH 7) and 5mM MgCl2. Error bars indicate standard deviation. There was a main effect of protein (p = 0.003) for the combined 15 and 30 min results. Proteins that do not share a letter are significantly different (5% LSD comparison).

For the substrate DHAP, LmSGPP2 was found to have similar kinetics on DHAP as it does on R5P (Table **5.1**), which is consistent with initial screening total phosphate results (Fig. **5.4b**). SGPP2 activity on DHAP has a relatively low catalytic efficiency compared with AtSGPP on R5P (approx. 10%, Table **5.1**). The K_m is also almost four times higher, indicating it has lower affinity for DHAP than AtSGPP does for R5P. DHAP is quite an unstable substrate. It underwent some non-enzymatic dephosphorylation under our reaction conditions and also contained trace amounts of phosphate before use. This increased the background error values in the Michaelis-Menten assays – especially at high substrate concentrations. Full kinetic analysis could not be done for LmSGPP3, but initial velocities (V_0) at 10mM were calculated (Table **5.2**). The V_0 of LmSGPP3-DHAP was significantly lower than that of LmSGPP2-DHAP (6.1 vs 10.3µmol/min, $p = 0.0007$).

Table 5.1: Kinetic parameters of phosphatase activity of purified MBP-AtSGPP, MBP-LmSGPP1, MBP-LmSGPP2 and MBP-LmSGPP3 on ribose-5-phosphate (R5P), and MBP-LmSGPP2 on dihydroxyacetone phosphate (DHAP). % catalytic efficiency column is relative to AtSGPP catalytic efficiency on R5P.

	Protein	V_{max} (µM/s)	K_m (mM)	$k_{cat}(s^{-1})$	k_{cat}/K_m ($M^{-1}s^{-1}$)	% catalytic efficiency
R5P	AtSGPP	12.17 ± 0.30	3.28 ± 0.32	33.93	10.35×10^3	100%
	LmSGPP1	9.91 ± 0.69	3.81 ± 1.03	27.64	7.25×10^3	~70%
	LmSGPP2	4.68 ± 0.28	12.16 ± 2.38	13.06	1.07×10^3	~10%
	LmSGPP3	11.25 ± 0.46	6.71 ± 0.94	31.36	4.68×10^3	~45%
DHAP	LmSGPP2	5.15 ± 0.25	12.55 ± 2.02	14.38	1.15×10^3	~10%

Table 5.2: Initial velocity at 10mM substrate MBP-AtSGPP, MBP-LmSGPP1, MBP-LmSGPP2 and MBP-LmSGPP3 on ribose-5-phosphate (R5P) and dihydroxyacetone phosphate (DHAP). Initial velocities that do not share a letter are significantly different ($p < 0.05$).

	R5P		DHAP	
Protein	V_o µmol/min	Sig.	V_o µmol/min	Sig.
AtSGPP	43.526	a	-	-
LmSGPP1	36.606	a	-	-
LmSGPP2	10.689	b	10.342	a
LmSGPP3	30.083	a	6.066	b

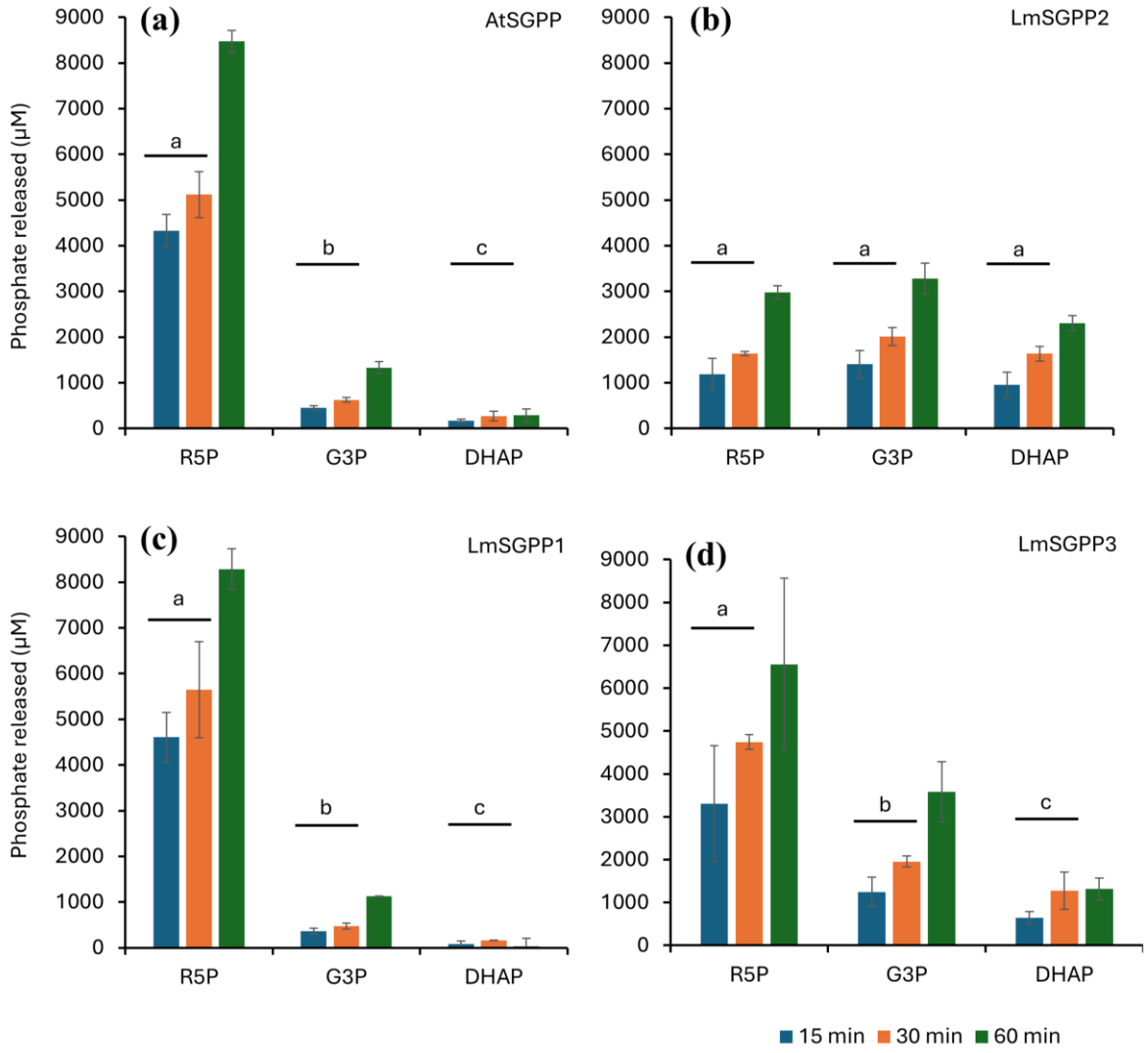


Figure 5.4: Phosphate released (µM) at three time points from the substrates Ribose-5-phosphate (R5P), glycerol-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). The reaction mixture contained 10mM substrate (R5P, G3P, or DHAP), 25 µg/mL of purified protein (MBP-AtSGPP, MBP-LmSGPP1, MBP-LmSGPP2 or MBP-LmSGPP3), 20mM Tris-HCl (pH 7), 5mM $MgCl_2$. Error bars indicate standard deviation. There was a main effect of protein (p = 0.003) for the combined 15 and 30 min results. Proteins that do not share a letter are significantly different (5% LSD comparison).

5.4 Discussion

5.4.1 SGPP2 can dephosphorylate DHAP – resulting in DHA

Confirming that SGPP2 can dephosphorylate DHAP provides further evidence that this protein is responsible for DHA accumulation in mānuka nectar. Kinetic analysis indicates that SGPP2 activity on DHAP has lower catalytic efficiency and affinity compared with AtSGPP on R5P (approx. 10% catalytic activity, K_m approx. four times higher, Table **5.1**). An analysis on the BRENDA database found that enzymes operating in specialised metabolism are generally far slower than those involved in core primary metabolism (~30-fold slower, Bar-Even *et al.*, 2011). If we compare LmSGPP2 to the

enzymes within the database, it has a below average catalytic efficiency, above average K_m, and an average turnover rate (Bar-Even *et al.*, 2011; Bathellier *et al.*, 2018). A bacterial HAD – BT4131 from *Bacteroides thetaiotaomicron* was found to have similarly low (~$1x10^3$) catalytic efficiency across a range of sugar phosphate substrates, and it was hypothesised to be an 'all purpose sugar phosphate phosphatase' – where high catalytic activity and substrate specificity is not necessarily compatible with its function (Lu *et al.*, 2005).

5.4.2 Comparison with previous studies

AtSGPP has been previously screened under similar conditions with the substrates R5P and G3P, and a similar pattern was observed of higher activity on R5P and lower on G3P (Caparrós-Martín *et al.*, 2013). The results for AtSGPP are comparable to those reported in Caparrós-Martín *et al.* (2013), as they found catalytic efficiency of $10.7x10^3$ for AtSGPP on R5P and it was $10.4x10^3$ here (Table **5.1**). Caparrós-Martín *et al.* (2014) previously assessed the impact of mutations at the two positions we see changes in here (positions 68 and 69) and found that AtSGPP-I68M had little impact on kinetics. AtSGPP-A69M however resulted in lower catalytic efficiency for all substrates screened, and different substrates were affected differently (e.g. R5P had ~50% and G3P ~25% lower catalytic efficiency compared with unaltered AtSGPP). The AtSGPP-A69M mutation also altered the pH dependency. In the present study we see that SGPP2 and 3 (where A is now S at position 69) have overall shifts in patterns compared to AtSGPP and SGPP1 (they have increased DHAP and G3P activity and decreased R5P activity, Fig. **5.4**, Table **5.1**). In Caparrós-Martín *et al.* (2014) the I68M change had little effect in AtSGPP, but in the present study for LmSGPP2, where position 69 is altered also, there is an effect for some substrates (R5P, Fig. **5.4**).

5.4.3 Amino acid changes in motif 5 likely alter substrate specificity

HAD phosphohydrolases, like all members of the HAD superfamily, have a conserved core catalytic domain formed by four loops, that correspond to sequence motifs 1-4 (Lahiri *et al.*, 2004; Burroughs *et al.*, 2006). This core domain positions substrate and cofactor binding residues and catalyses the phosphoryl group transfer to the active site aspartic acid (D) on loop 1, then to a water molecule (Lu *et al.*, 2005). This fundamental catalytic mechanism (Aspartic acid intermediate and Mg^{2+} cofactor) is conserved across the family, and diversification in substrate is mediated by the cap domain (Lu *et al.*, 2005; Burroughs *et al.*, 2006). SGPP2 has a C1 type cap containing the substrate specificity loop, consisting of a conserved glycine residue flanked by other residues that contribute to the catalytic site (Lu *et al.*, 2005; Caparrós-Martín *et al.*, 2014).

The key differences observed among the four SGPPs here are in the substrate specificity loop (motif 5 at positions 68 and 69, Fig. **5.2**). Based on these results (Table **5.1** and Table **5.2**, as well as Caparrós-Martín *et al*. (2014), it appears position I68M alone may not be sufficient to alter kinetics for some substrates, but when position 69 is also altered it may have more of an effect. Most of the cap residues are hydrophobic and serve to stabilise the deprotonated lysine (K) from loop/motif 3 and provide a solvent excluded active site in closed form (Lahiri *et al.*, 2004; Caparrós-Martín *et al.*, 2014). Isoleucine, alanine and methionine are non-polar/hydrophobic, and serine is polar/hydrophilic uncharged. An amino acid change like alanine to serine that alters hydrophobicity could certainly alter activity. Methionine is also unbranched, and more flexible than isoleucine (Aledo, 2019). Sequence comparisons across the whole proteins suggest that the motif 5 changes at positions 68 (LmSGPP2) and 69 (LmSGPP2 and 3) are the most likely candidates for the changes in substrate specificity observed, however there could be other changes that have an effect on activity. Caparrós-Martín *et al.* (2019) have demonstrated that motif 5 is essential but not the only determinant of substrate specificity, with residues in other positions potentially having a role.

Substrate ambiguity is common in members of the HAD superfamily, with 75% of its members able to act on at least five substrates (Huang *et al.*, 2015). AtSGPP specifically was found to be able to dephosphorylate eight of the 13 substrates screened (Caparrós-Martín *et al.*, 2013). Having flexible loops involved in the active site provides a good basis for diversifying substrate (Pandya *et al.*, 2014), and substrate ambiguity allows for evolution of new functions. Mutations that result in physiological relevance can occur within the gene, or in regulatory regions increasing the expression (Copley, 2020). In *Leptospermum*, *Sgpp2* is within a complex locus – three *Sgpp* paralogues that show different patterns of expression (Grierson *et al.*, 2024, Chapter 2; Chapter 4) and substrate specificity (Fig. **5.3**, Fig. **5.4**, Table **5.1**). The three *Leptospermum* SGPPs are potentially diverging in their function. For SGPP2, reaching physiological relevance may be due to a shift in substrate specificity and a mutation inducing high expression, rather than mutations causing optimisation for DHAP specifically as a substrate and achieving high catalytic efficiency.

5.4.4 Role of SGPP2 in *Leptospermum* nectaries

Grierson *et al.* (2024, Chapter 2) proposed the activity of SGPP2 may be related to a role in Pi homeostasis during sugar production and secretion by the nectary. Nectary cells have rapid sugar-phosphate turnover and experience oxidative and osmotic stress during nectar secretion (Nicolson & Thornburg, 2007). In *Leptospermum,* where the nectaries are also photosynthetic, maintaining appropriate Pi concentrations is likely crucial during periods of metabolic flux (Clearwater *et al.*, 2021; Grierson *et al.*, 2024).

Triose phosphates, and specifically DHAP, increase during active photosynthesis (Macrae & Lunn, 2006). Clearwater *et al.* (2021) demonstrated that nectar DHA is derived from these cellular pools of DHAP through ^{13}C labelled experiments. Clearwater *et al*. (2021) showed that increased light intensity resulted in increased nectar DHA, and inhibition of photosynthesis and triose phosphate transport resulted in decreased DHA. Our results strengthen the previously hypothesised role of SGPP2 contributing to phosphate homeostasis through demonstrating its ability to dephosphorylate DHAP, releasing DHA and Phosphate. The dephosphorylation of DHAP by SGPP2 is effectively irreversible (at least by SGPP2), which may explain the accumulation of DHA and subsequent diffusion out of the nectary parenchyma.

The relatively low catalytic efficiency of SGPP2 on DHAP may be due to its broad substrate specificity (Fig. **5.4b**). The high expression levels of SGPP2 could enable high levels of nectar DHA to be produced. SGPP2 may function as a generalist 'housekeeping' phosphatase – maintaining Pi homeostasis by being able to dephosphorylate a range of substrates. SGPP2 activity may occur largely when DHAP pools increase during photosynthesis. Its low affinity could mean that SGPP2 primarily acts on excess DHAP, ensuring that when only low concentrations of DHAP are present, it can still flow through primary metabolic pathways and preserve essential function. SGPP2 could be competing for substrate with Triose Phosphate Isomerase (TPI) which has often been regarded as a near perfect catalyst (Sharma *et al.*, 2012). Theoretical calculations to illustrate this competition indicate that if TPI and SGPP2 were in isolation competing for DHAP at a cytosolic concentration of 1mM, only around ~2% of total DHAP acted on by the enzymes would be converted to DHA.

AtSGPP is expressed in all plant organs but is high in flowers, developing siliques and under phosphate starvation (Caparrós-Martín *et al.*, 2013). In *Leptospermum* we see that SGPP1 and 3 are expressed in all plant tissues sampled but SGPP2 is restricted to the flower (Fig. **2.5a** within Grierson *et al.* 2024, Chapter 2), which is consistent with a specialised role in photosynthesising nectar producing tissues.

5.4.5 Future directions

Future work could further investigate the range of substrates that SGPP2 can work on, as it is possible that SGPP2 has a preferred substrate that has not yet been screened. Assessing its activity on glyceraldehyde 3-phosphate (the product of DHAP isomerisation by TPI), for example could help clarify its physiological role. This work was carried out at approximate (assumed) cytosolic pH (7.3) but it would be interesting to see how pH affects activity. To further assess the molecular basis of the shift in substrate specificity between the SGPPs, targeted mutagenesis of the substrate specificity loop (motif 5) could be informative. Generating variants of SGPP2 in which the motif 5 is replaced

with the corresponding sequence from LmSGPP1 and LmSGPP3 would allow us to assess if changes in this motif are sufficient to alter substrate specificity or catalytic efficiency and help to understand functional divergence between these SGPP paralogues.

5.5 References

Aledo JC. **2019**. Methionine in proteins: The Cinderella of the proteinogenic amino acids. *Protein Science* **28**: 1785–1796.

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. **2011**. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**: 4402–4410.

Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. **2018**. Rubisco is not really so bad. *Plant, Cell & Environment* **41**: 705–716.

Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. **2006**. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *Journal of Molecular Biology* **361**: 1003–1034.

Bustos R, Castrillo G, Linhares F, Puga MI, Rubio V, Pérez-Pérez J, Solano R, Leyva A, Paz-Ares J. **2010**. A Central Regulatory System Largely Controls Transcriptional Activation and Repression Responses to Phosphate Starvation in *Arabidopsis*. *PLoS Genetics* **6**: e1001102.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. **2013**. HAD hydrolase function unveiled by substrate screening: enzymatic characterization of *Arabidopsis thaliana* subclass I phosphosugar phosphatase AtSgpp. *Planta* **237**: 943–954.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. **2014**. The kinetic analysis of the substrate specificity of motif 5 in a HAD hydrolase-type phosphosugar phosphatase of *Arabidopsis thaliana*. *Planta* **240**: 479–87.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. **2019**. Sequence Determinants of Substrate Ambiguity in a HAD Phosphosugar Phosphatase of *Arabidopsis Thaliana*. *Biology* **8**: 77.

Clearwater MJ, Noe ST, Manley-Harris M, Truman G-L, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. **2021**. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

Copley SD. **2020**. Evolution of new enzymes by gene duplication and divergence. *The FEBS Journal* **287**: 1262–1283.

Du Z, Deng S, Wu Z, Wang C. **2021**. Genome-wide analysis of haloacid dehalogenase genes reveals their function in phosphate starvation responses in rice. *Plos One* **16**: e0245600.

Grierson ERP, Thrimawithana AH, van Klink JW, Lewis DH, Carvajal I, Shiller J, Miller P, Deroles SC, Clearwater MJ, Davies KM, *et al.* 2024. A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*). *New Phytologist* **242**: 2270–2284.

Gutiérrez-Ginés MJ, Madejón E, Lehto NJ, McLenaghen RD, Horswell J, Dickinson N, Robinson BH. **2019**. Response of a Pioneering Species (*Leptospermum scoparium* J.R.Forst. & G.Forst.) to Heterogeneity in a Low-Fertility Soil. *Frontiers in Plant Science* **10**: 93.

Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, Toews Keating S, Aono M, Love JD, Evans B, *et al.* 2015. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proceedings of the National Academy of Sciences* **112**: E1974–E1983.

Kuznetsova E, Nocek B, Brown G, Makarova KS, Flick R, Wolf YI, Khusnutdinova A, Evdokimova E, Jin K, Tan K, *et al.* 2015. Functional Diversity of Haloacid Dehalogenase Superfamily Phosphatases from *Saccharomyces cerevisiae*: BIOCHEMICAL, STRUCTURAL, AND EVOLUTIONARY INSIGHTS. *The Journal of Biological Chemistry* **290**: 18678–18698.

Lahiri SD, Zhang G, Dai J, Dunaway-Mariano D, Allen KN. **2004**. Analysis of the Substrate Specificity Loop of the HAD Superfamily Cap Domain. *Biochemistry* **43**: 2812–2820.

Lu Z, Dunaway-Mariano D, Allen KN. **2005**. HAD Superfamily Phosphotransferase Substrate Diversification: Structure and Function Analysis of HAD Subclass IIB Sugar Phosphatase BT4131. *Biochemistry* **44**: 8684–8696.

Macrae E, Lunn J. **2006**. Control of Sucrose Biosynthesis. In: Plaxton WC, McManus MT, eds. *Control of Primary Metabolism in Plants*. Blackwell Publishing, 234–257.

Nicolson SW, Thornburg RW. **2007**. Nectar chemistry. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Springer, Dordrecht, 215–264.

Nilsson L, Lundmark M, Jensen PE, Nielsen TH. **2012**. The *Arabidopsis* transcription factor PHR1 is essential for adaptation to high light and retaining functional photosynthesis during phosphate starvation. *Physiologia Plantarum* **144**: 35–47.

Noe ST. **2022**. The effects of soil nutrient addition on mānuka (*Leptospermum scoparium*) growth, flowering, and nectar production. PhD thesis, University of Waikato, Hamilton, New Zealand.

Orians GH, Milewski AV. **2007**. Ecology of Australia: the effects of nutrient-poor soils and intense fires. *Biological Reviews* **82**: 393–423.

Pandey BK, Mehra P, Verma L, Bhadouria J, Giri J. **2017**. OsHAD1, a Haloacid Dehalogenase-Like APase, Enhances Phosphate Accumulation. *Plant Physiology* **174**: 2316–2332.

Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. **2014**. Enzyme Promiscuity: Engine of Evolutionary Innovation. *Journal of Biological Chemistry* **289**: 30229–30236.

Sharma S, Mustafiz A, Singla-Pareek SL, Shankar Srivastava P, Sopory SK. **2012**. Characterization of stress and methylglyoxal inducible triose phosphate isomerase (OscTPI) from rice. *Plant Signaling & Behavior* **7**: 1337–45.

Chapter 6: Overexpression of *Leptospermum morrisonii Sgpp2* in nectaries of *Petunia* and *Nicotiana*

6.1 Introduction

Nectar is a sugar rich liquid produced by nectaries that is the primary interface between plants and their pollinators and has essential ecological and evolutionary functions. Nectaries have evolved multiple times independently and are structurally diverse (Bernardello, 2007). The composition of nectar also varies, both in the ratios of the sugars sucrose, glucose and fructose – and in the content of other compounds such as amino acids, proteins and specialised metabolites (Roy *et al.*, 2017). Despite nectar and nectary biology underlying the economic performance of varied crops, many aspects of their biology are not well understood (Liao *et al.*, 2025).

The molecular genetics of nectaries are variable because of their multiple evolutionary origins (Roy *et al.*, 2017). Most work has been done on short-lived flowers with mesophyllary nectaries and an eccrine mode of nectar secretion, such as model species *Arabidopsis*, *Petunia* and tobacco (e.g. Ren *et al.*, 2007; Kram *et al.*, 2009; Solhaug *et al.*, 2019). Some of the key genes and their functions have been characterised, including SPS which synthesises sucrose that is then exported into the apoplast via the uniporter SWEET9. The sucrose is then hydrolysed into fructose and glucose by CWINV4, creating a sucrose gradient which is required for SWEET9 as it is bidirectional (Ruhlmann *et al.*, 2010; Lin *et al.*, 2014; Minami *et al.*, 2021). Plants carrying mutations in the genes for *SPS*, *Sweet9* or *CWINV4* do not produce nectar, and conversely, overexpression of *Sweet9* leads to an increase in nectar (Ruhlmann *et al.*, 2010; Lin *et al.*, 2014). However, the role these genes play in nectar production in more diverse nectaries in non-model species is unknown (Liao *et al.*, 2025).

Many *Leptospermum* species accumulate dihydroxyacetone (DHA), a triose sugar, in their nectar, which is a trait unique to some species within the Myrtaceae (Obeng-Darko *et al.*, 2023). Clearwater *et al.* (2021) recently demonstrated that mānuka nectar DHA is probably derived from cellular pools of DHAP, and that photosynthesis within the nectary contributes to nectar sugars – including DHA. Grierson *et al.* (2024, Chapter 2) identified a phosphatase gene – *Sgpp2* – that was linked to nectar DHA through gene expression and QTL analyses. It was hypothesised that this gene was involved in phosphate homeostasis and was responsible for DHA accumulation through dephosphorylation of cytosolic pools of DHAP. Functional analysis *in vitro* has since demonstrated the ability of SGPP2 to dephosphorylate DHAP, supporting this hypothesis (Chapter 5).

Furthering our understanding of DHA accumulation in *Leptospermum* can inform breeding and replanting programmes for the honey industry, and also contribute to our understanding of non-model species nectaries and the accumulation of unique compounds in nectar. We aim to better understand the role of SGPP2 in generating DHA through its overexpression in the nectaries of model plants. We report here the generation of transgenic *Petunia* and *Nicotiana* lines overexpressing the *Leptospermum morrisonii Sgpp2* under the nectary specific promoter *Petunia axillaris Nec1*/*Sweet9*, and evaluation of nectar DHA in the resulting transgenic lines.

6.2 Methods

6.2.1 Construct design

The gene sequence for *LmSgpp2* was identified through blasting the *Leptospermum morrisonii* genome (Chagné, Deng & Grierson, Chapter 3) with the *LsSgpp2* sequence and manually annotating the locus based on the *Leptospermum scoparium* cv. 'Crimson Glory' genome (v2.0, Grierson *et al.* 2024, Chapter 2). The *PaNec1* promoter was obtained from the *Petunia axillaris* genome (v1.6.2) by blasting the *Petunia hybrida Nec1* sequence from Ge *et al.* (2000, AF313914.1) and using the length of promoter as described in Ge *et al.* (2000). The *PaNec1-LmSgpp2* expression cassette in pUC57 was synthesised by Genscript (USA) as a commercial service. *Venus* sequence was amplified from a Venus-N7 (Nuclear localised yellow fluorescent protein) based plasmid, and cloned into pEG47 in place of *LmSgpp2*. Both expression cassettes were then cloned into pART27 for *Agrobacterium*-mediated transformation. The resulting plasmids were termed pEG49 (*PaNec1promoter-LmSgpp2-OCSterminator*) and pEG50 (*PaNec1promoter-Venus-OCSterminator*) (Supplementary Fig. **S6.1**).

6.2.2 Agrobacterium

Both pEG49 and pEG50 were transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation using a Gene Pulser Xcell (BioRad) as per manufacturers' instructions, and the resulting bacterial cultures stored as beadstocks at -80°C. Two days prior to transformation, a 4mL LB starter culture was initiated directly from beadstock with antibiotic selection (spectinomycin 100mg/L). Cultures were grown overnight at 28°C with shaking. The day before transformation, a 25mL culture was inoculated from the starter culture and grown overnight.

6.2.3 Transformation protocol

We do not have Māori or regulatory approval to use Aotearoa-New Zealand native gene sequences in transgenic experiments or conduct transgenic experiments in native species. We also do not have an established transformation protocol for non-native *Leptospermum*. Therefore, transformation

experiments were conducted using *L. morrisonii* sequences in heterologous systems (*Petunia* and *Nicotiana*). *Petunia* cv. 'Mitchell' (*Petunia axillaris* x (*P. axillaris* x *P. hybrida*)) and *Nicotiana bethamiana* were grown in greenhouse conditions and young leaves were taken for *Agrobacterium*-mediated transformation. The leaves were sterilised by submersion in 10% bleach with Tween20 for 10 mins, swirling occasionally. The leaves were rinsed three times in sterile Milli-Q water and then blotted dry. Leaf discs were cut using a 5mm Rapid-Core biopsy punch (ProSciTech) and placed on cocultivation media for 4 days (Media contents listed in Supplementary material for Chapter 4, Table **S4.2**). When the *Agrobacterium* cultures reached an A_{600} of 0.6-08, they were pelleted and resuspended in 50mL minimal media + 200µM acetosyringone to an A_{600} of 0.2-0.3 and incubated with shaking for 2-4 hrs at 28°C. Leaf discs were incubated in 25mL *Agrobacterium* cultures for 5 mins, then blotted dry with sterile tissue paper and plated back onto cocultivation media. Uninoculated regeneration and kill controls were also plated on cocultivation media. Leaf discs were then cocultivated at 22°C for 3 days before transfer to selective media (for pEG49 and pEG50 treatments and kill controls) or regeneration media (regeneration controls). Each plate was transferred to new media every 3 weeks, and as shoots arose they were harvested into tubs of the same media to continue to grow. When shoots were at least 2-3cm, they were transferred onto selective rooting media (treatments) or regeneration rooting media (regeneration controls). After 3 to 6 weeks, rooted shoots were exflasked into the greenhouse.

At least 20 independent lines for each treatment were exflasked and an approximately 1kb portion of the T-DNA insert amplified by PCR to check they were transgenic.

6.2.4 Developmental series gene expression

To assess whether the transgene was being expressed, the *Petunia* transgenics nectaries were sampled at three developmental stages for reverse transcription-quantitative polymerase chain reaction (RT-qPCR) from two lines from each treatment and one control line. Floral developmental stages were based on Ge *et al.* (2000) and are defined as – Stage 1: short (2-4cm) closed corolla, white nectary; stage 2: elongated (5-6cm) closed corolla, yellow nectary; and stage 3: elongated (5-6cm) open corolla, yellow to orange nectary (Ge *et al*. 2000 stage 3 and 4 combined) (see Chapter 4, Figure **4.6g**). A sample consisted of nectaries from four flowers. Stages 1 and 3 were sampled in duplicate, and stage 2 in triplicate. The tissue sampled was nectary enriched – transverse cuts were made through the hypanthium/ovary either side of the nectary (Supplementary Fig. **S6.2**).

RNA was extracted from nectary enriched tissue based on the CTAB method from Gambino *et al.* 2008). For RT-qPCR, 700ng of RNA was used for cDNA synthesis. Genomic DNA was removed and cDNA was synthesised using the iScript™ gDNA Clear cDNA Synthesis kit as per the manufacturer's

instructions (Bio-Rad, USA). The RT-qPCR analyses were performed with a LightCycler 480 instrument. Reactions contained 3.75µL of LightCycler® 480 SYBR Green I Master (Roche Diagnostics, USA), 0.14µL of each primer (10µM), 2µL of diluted cDNA (1:20), and ultrapure water to a total volume of 7µL. The RT-qPCRs were as follows: one cycle of 5min at 95°C, followed by 45 cycles of 10s at 95°C, 20s at 55°C, and 10s at 75°C. Using the LightCycler 480 (Roche; version 1.5.0 SP4), data means were calculated from three technical replicates per biological replicate for each line and stage. Relative transcript abundance was then calculated for *Sgpp2*, *Venus* and *Nec1* using geometric means of the reference genes *Actin* and *EF1 alpha*. Primer efficiencies were calculated using serial dilution (primer pairs listed in Supplementary material for Chapter 2, Table **S6.2**).

Statistical significance was determined using a one-way ANOVA in Origin (2022b), followed by Fisher's LSD using a cutoff *p*-value of <0.05.

6.2.5 Nectar sampling

At least 14 lines per treatment had three replicates of nectar sampled. Regeneration controls had five lines with three replicates. Nectar was sampled from stage 3 flowers (described above) by centrifugation at 3000rpm for 1 min for each flower. The *Nicotiana* stage 3 was defined similarly to *Petunia* – elongated open corolla. The open petal portion of the corolla was removed so that the corolla tube could be inserted into the sieve (0.5mL Eppendorf tube with 1-3 holes in the base) inside a 1.5mL Eppendorf. Each nectar sample contained nectar from 3-4 flowers. Nectar samples were pooled from multiple days if fewer than four stage 3 flowers were present. At least three replicates were taken for each line.

6.2.6 Nectar DHA analysis

A preliminary subset of the nectar samples was analysed for the presence of DHA using GC-MS, and values were normalised to mg sugar using Brix. Methodology was based on Williams *et al.* (2014), with minor modifications to increase sensitivity. The analytical limits of detection (based on the smallest peak area that could be manually integrated) and quantitation of DHA were 0.3 and 0.7ng/mL respectively. For the *Petunia* nectar samples, the limits of detection and quantitation were 0.001 and 0.0024mg/g respectively, sugar based on 30% sugar content. For *Nicotiana* samples the limits of detection and quantitation were 0.004 and 0.0093mg/g sugar respectively, based on 22% sugar content.

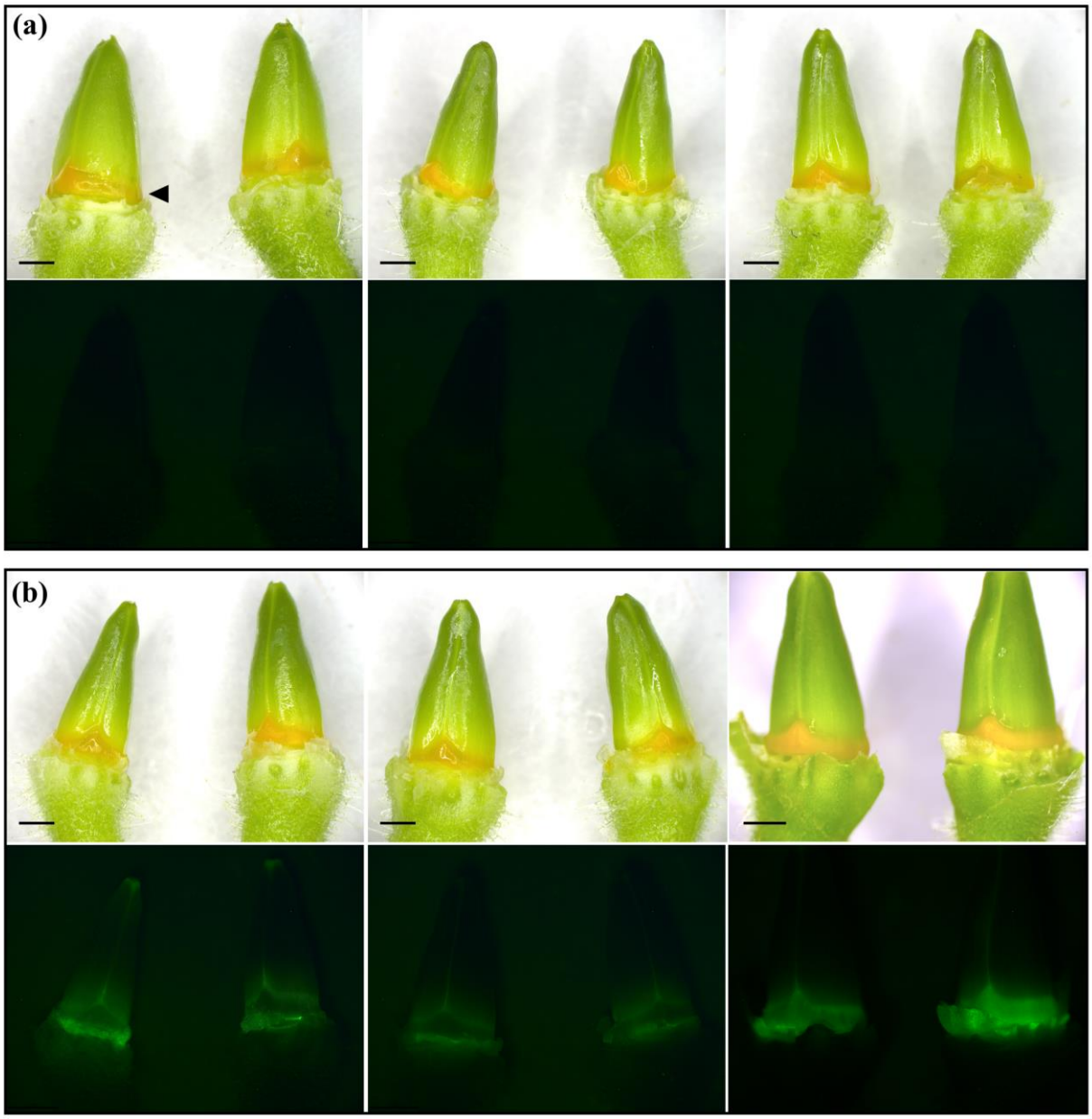


Figure 6.1: Ovary and nectary of overexpression lines of *Petunia cv* 'Mitchell' under white and blue light (with YFP filter). Petals and sepals have been removed. (a) Three independent *PaNec1-LmSgpp2* transformed lines (T1R1-2, T1R4-4, T1R6-5). (b) Three independent *PaNec1-Venus* transformed lines (T2R4-1, T2R6-5, T2R4-2). Scale bar 1mm. Nectary tissue (orange) indicated with arrow in (a).

6.2.7 Fluorescence screening and imaging

All lines were visually screened using a microscope for presence of fluorescence in the nectary and floral structures, and images were taken of the nectary for five lines at stage 3 for *Petunia* and three lines of the equivalent stage for *Nicotiana*. YFP fluorescence was viewed with a YFP filter and blue light. Paired white light and blue light images were taken on a Leica camera (DF550) on a Leica dissection microscope (M205FA).

6.3 Results

6.3.1 Plants were confirmed as containing transgenes

In the *PaNec1-LmSgpp2* transformed *Petunia* lines the transgene was detected in 18 out of 20 lines, and in the *PaNec1-Venus* transformed lines, 20 out of 20 had the transgene (Table **S6.1**). As expected, the *PaNec1-LmSgpp2* transformed lines were not fluorescent (Fig. **6.1a**). The transgene was detected in 5 out of 5 of the *PaNec1-LmSgpp2* transformed *Nicotiana* lines screened, and in *PaNec1-Venus* transformed lines, 5 out of 5 screened had the transgene. No transgenes were detected in the Regeneration controls for either species.

6.3.2 Nectar DHA was not increased in transgenic lines

For *Petunia*, of the 59 nectar samples measured for DHA, 40 were below limit of detection (0.001mg/g sugar), and another 13 below the limit of quantification (0.0024mg/g sugar). The remaining six samples ranged between 0.0025mg/g sugar to 0.0098mg/g with an average of 0.0039mg/g (data not shown). The presence of detectable DHA did not correlate with any transgene/treatment, and it was concluded that *LmSgpp2* transgene did not promote accumulation of DHA in nectar of the transgenics. The data for the *Nicotiana* nectar samples was similar – 15 samples were below the limit of detection and 30 below the limit of quantification. The samples above the limit of quantification had 0.012mg/g sugar and 0.01mg/g sugar.

6.3.3 The localisation of fluorescence was variable in *PaNec1-Venus* control lines

Out of the 20 pEG50 transformed *Petunia* lines, 17 flowered during screening and were scored for fluorescence at least twice, with four flowers examined each time. *Venus* fluorescence was observed in 12 out of 17 lines screened, however three of those only had fluorescence in some flowers (Table **S6.1**). Fluorescence was variable in its brightness and location – sometimes located surrounding the nectary, in the vasculature, and/or in the nectary (Fig. **6.1b**). Variation in brightness and location was also observed in *Nicotiana* (Supplementary Fig. **S6.3**).

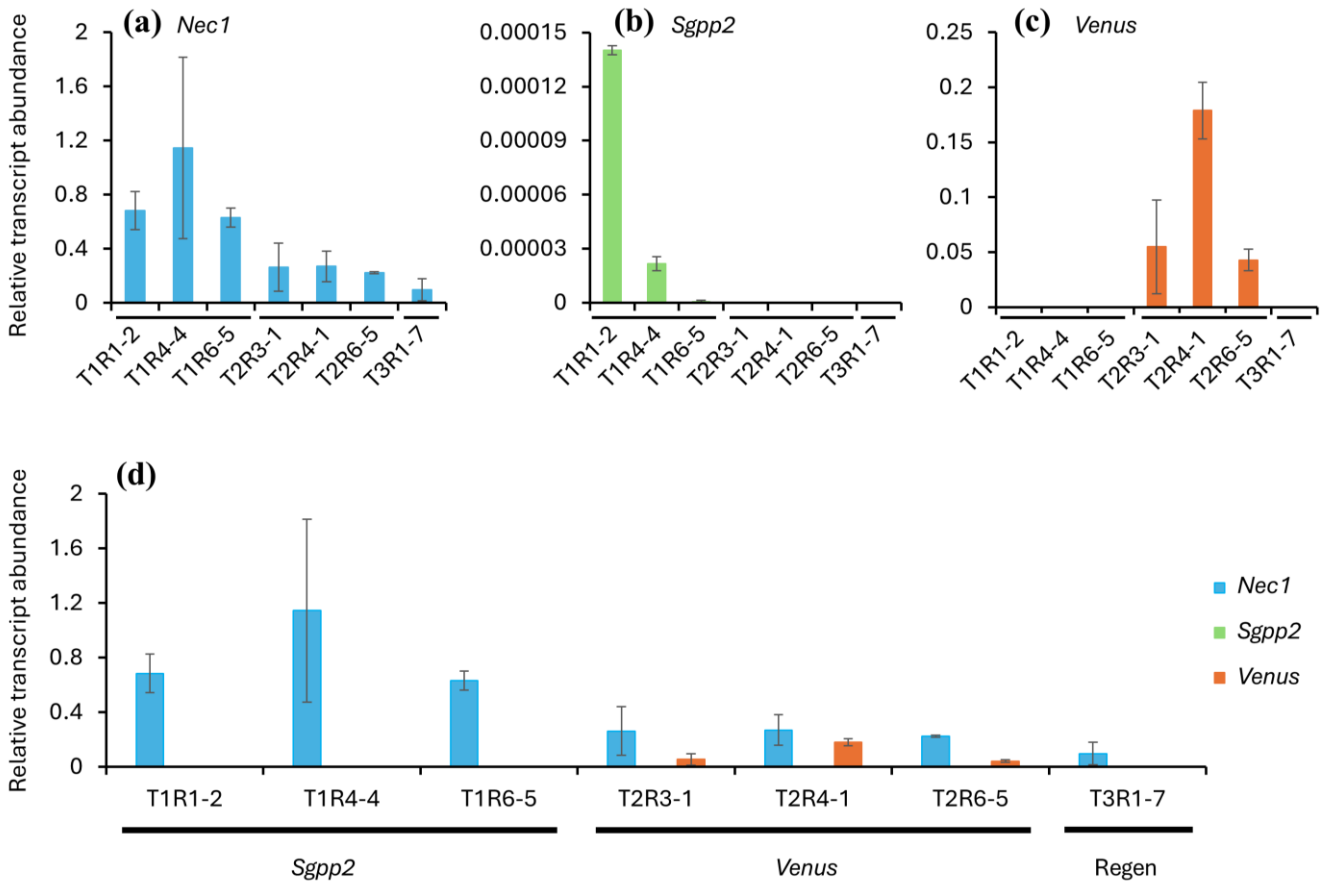


Figure 6.2: Relative transcript abundance of the *Nec1* gene, and the *Sgpp2* and *Venus* transgenes, in three independent lines of *Sgpp2* transformed *Petunia*, three independent lines of *Venus* transformed *Petunia*, and one untransformed regeneration control. (a) Transcript abundance of endogenous *Nec1*. (b) Transcript abundance of the transgene *Sgpp2*. (c) Transcript abundance of the transgene *Venus*. (d) Transcript abundance of endogenous *Nec1* and the transgenes *Sgpp2* and *Venus* (same data as a b and c on the same graph, to visualise with a common vertical axis for transcript abundance). All three genes are under *Nec1* promoter. Each line was sampled twice, with four flowers per sample.

6.3.4 Transcript abundance in transgenic *Petunia* lines

For stage 2 nectaries, endogenous *Nec1* transcript abundance was significantly higher in the *LmSgpp2* transformed lines ($p<0.05$) compared with *Venus* and the regeneration controls, and there was no significant difference between *Venus* transformed lines and the regeneration controls. As expected, *LmSgpp2* and *Venus* transcripts were restricted to the corresponding treatment (although very low values of non-replicated signal were detected in some *LmSgpp2* transformed samples using *Venus* primers, most likely representing non-specific amplification). Interestingly, even though they were under the same promoter, *LmSgpp2* transcript abundance was at least 900 times lower than *Venus* (Fig. **6.2b**, **c**, and **d**).

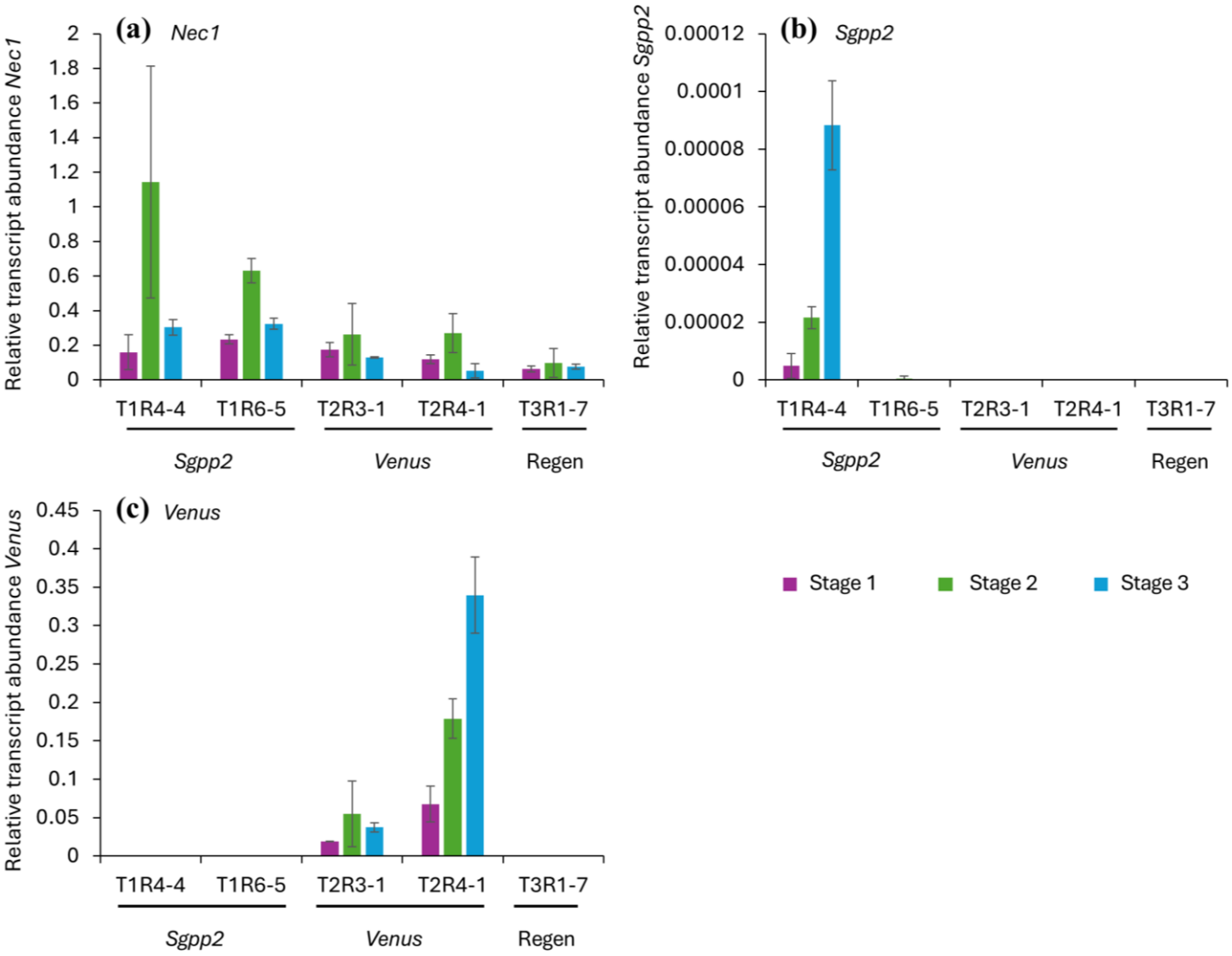


Figure 6.3: Relative transcript abundance of *Nec1* gene, and the *Sgpp2* and *Venus* transgenes, in two independent lines of *Sgpp2* transformed *Petunia*, two independent lines of *Venus* transformed *Petunia*, and one untransformed regeneration control, at three developmental stages. (a) Transcript abundance of endogenous *Nec1* at developmental stages 1 2 and 3. (b) Transcript abundance of the transgene *Sgpp2* at developmental stages 1 2 and 3. (c) Transcript abundance of the transgene *Venus* at developmental stages 1 2 and 3. Each line was sampled twice, with four flowers per sample.

In the *Petunia* developmental series, endogenous *Nec1* transcript abundance showed a similar pattern across all lines, with a peak at stage 2 (Fig. **6.3a**). There was no significant difference in transcript amounts between lines at stage 1, but at stage 2 and 3 transcript amounts in *LmSgpp2* lines were significantly higher than *Venus* and the regeneration control (p<0.05, Fig. **6.3b** and **c**). Similar to the stage 2 results, *LmSgpp2* and *Venus* transcripts were restricted to the corresponding treatment, with expression varying between lines (Fig. **6.3**).

6.4 Discussion

There was no increase in DHA found in the nectar of the transgenic lines, and the trace amounts present in some samples are expected as we have previously detected low levels of DHA in the nectar of untransformed *Petunia* (the authors' unpublished data). The lack of increased DHA is also not surprising given the *Petunia axillaris Nec1* promoter used to express *LmSgpp2* was not found to be consistently nectary specific and strongly expressed in the present experiment. In contrast to these results, Ge *et al.* (2000) used the *Nec1* promoter from *Petunia* cv. 'Mitchell' expressing the GUS reporter gene and showed strong nectary specific GUS signal, and subsequent studies have validated this expression pattern (e.g. Lin *et al.*, 2014). Ge *et al.* (2000) used the same experimental line as used for our experiments (Mitchell/W115), and the length of promoter used was also the same. We used the promoter sequence from *P. axillaris,* but Petunia cv. 'Mitchell' is a hybrid of *P. axillaris* and *P. hybrida*, and the promoter region is 99.72% identical between *P. axillaris* and *P. hybrida*. In the *Nicotiana PaNec1-Venus* transformed lines, fluorescence was again not consistently nectary specific, with many lines having expression in the ovary (Supplementary Fig. **S6.3**). *Nec1* expression has previously been found at low amounts in other tissues such as ovary and stamen, but that is usually in addition to higher expression in the nectary (e.g. Carter *et al.*, 1999; Kram *et al.*, 2009). *PaNec1-Venus* expression and protein fluorescence were variable (Fig. **6.1**, Fig. **6.2** and Fig. **6.3**) and were not consistent across lines or within flowers within a line. The reason for the variability in nectary specificity here is unknown, but it could be due to the location of the transgene insertion. It also cannot be ruled out that the transgene/promoter could have been mutated in the *Agrobacterium*, and more thorough sequencing of the transgene cassette should be undertaken to assess this. In contrast, Chapter 4 showed that *Petunia* and *Nicotiana* transgenics with the *LmSgpp2* promoter driving *Venus* had consistently high and largely nectary specific expression/fluorescence.

Despite being under the same *PaNec1* promoter, *Venus* transformed lines had far higher *Venus* transcript abundance than *LmSgpp2* transcript abundance in the *LmSgpp2* transformed lines (Fig. **6.2** and **6.3**). It is not uncommon to observe variability in expression when generating overexpression lines (e.g. Meyer, 1995), and it can be influenced by factors such as location of the transgene insertion and surrounding elements, transcriptional silencing due to homology with a gene being silenced and post-transcriptional regulation by e.g. miRNAs or transcript instability. The absence of introns in the *LmSgpp2* transgene could make it more of a target for degradation by small RNAs (Kiselev *et al.*, 2022). It has been demonstrated that LmSGPP2 is able to dephosphorylate a range of substrates including DHAP (Chapter 5), so another possible explanation for this difference in transcript abundance could be the role of DHAP in primary metabolism. As DHAP is an essential

intermediate in metabolic pathways such as glycolysis and gluconeogenesis, perhaps overexpression of LmSGPP2 had detrimental effects and its expression was somehow downregulated (discussed further below). Furthermore, as LmSGPP2 has broad substrate specificity it could also be dephosphorylating other sugar phosphates needed for other metabolic processes in the nectary and ovary, compounding detrimental effects.

Endogenous *Nec1* expression was highest in nectaries just before anthesis (stage 2) as observed previously (e.g. Ge *et al.,* 2000). Interestingly, endogenous *Nec1* expression was significantly increased in the *LmSgpp2* transformed lines at stage 2 and 3, compared with *Venus* transformed and the regeneration control. This could be related to the potential downregulation of the *PaNec1-LmSgpp2* transgene affecting endogenous *Nec1* expression. Perhaps in the attempt to restore homeostasis, *Nec1* expression was increased. We did not, however, see any significant difference in nectar sugar content across the nectar sampled lines (p= 0.083, data not shown).

DHA was not found to be increased in the *LmSgpp2* transformed lines, which was unsurprising given the minimal and potentially non-nectary expression of the *LmSgpp2* gene in the lines investigated for transcript abundance. Variability in *PaNec1-Venus* fluorescence – in both brightness and location – also provides evidence that the *PaNec1* promoter may not have been expressing the *LmSgpp2* gene in the nectary tissue. Due to the variability in location of fluorescence of *PaNec1-Venus* and the low amounts of *PaNec1-LmSgpp2* transcripts, we could not assess the effect of LmSGPP2 in nectary tissue. We aim to repeat these experiments with a strongly nectary specific promoter and a candidate for this would be the *LmSgpp2* promoter itself. Grierson *et al.* (2024, Chapter 2) demonstrated that *LsSgpp2* was more highly expressed than *Nec1* (within Supplementary material for Chapter 2, Table **S2.3**), and Chapter 4 showed that it is nectary specific – and that specificity is conserved across diverse lineages. Future experiments could also use codon optimisation to reduce any endogenous small RNAs potentially degrading the *LmSgpp2* transcript sequence. We chose to use model species with established transformation protocols for this experiment, however a nectar research model species with photosynthetic nectaries could more closely mirror *Leptospermum* nectaries and could be a better system to assess LmSGPP2 function.

6.5 References

Bernardello G. **2007**. A systematic survey of floral nectaries. In: Nicolson SW, Nepi M, Pacini E, eds. *Nectaries and Nectar*. Springer, Dordrecht, 19–128.

Carter C, Graham RA, Thornburg RW. **1999**. Nectarin I is a novel, soluble germin-like protein expressed in the nectar of *Nicotiana* sp. *Plant Molecular Biology* **41**: 207–216.

Clearwater MJ, Noe ST, Manley-Harris M, Truman G-L, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. **2021**. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

Gambino G, Perrone I, Gribaudo I. **2008**. A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* **19**: 520–5.

Ge Y-X, Angenent GC, Wittich PE, Peters J, Franken J, Busscher M, Zhang L-M, Dahlhaus E, Kater MM, Wullems GJ, *et al.* 2000. NEC1, a novel gene, highly expressed in nectary tissue of *Petunia hybrida*. *The Plant Journal* **24**: 725–734.

Kiselev KV, Suprun AR, Aleynova OA, Ogneva ZV, Kostetsky EY, Dubrovina AS. **2022**. The Specificity of Transgene Suppression in Plants by Exogenous dsRNA. *Plants* **11**: 715.

Kram BW, Xu WW, Carter CJ. **2009**. Uncovering the *Arabidopsis thaliana* nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. *BMC Plant Biology* **9**: 92.

Liao IT, Gong Y, Kramer EM, Nikolov LA. **2025**. The developmental basis of floral nectary diversity and evolution. *New Phytologist* **246**: 2462–2477.

Lin IW, Sosso D, Chen LQ, Gase K, Kim SG, Kessler D, Klinkenberg PM, Gorder MK, Hou BH, Qu XQ, *et al.* 2014. Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**: 546–9.

Meyer P. **1995**. Understanding and controlling transgene expression. *Trends in Biotechnology* **13**: 332–337.

Minami A, Kang X, Carter CJ. **2021**. A cell wall invertase controls nectar volume and sugar composition. *The Plant Journal* **107**: 1016–1028.

Obeng-Darko SA, Sloan J, Binks RM, Brooks PR, Veneklaas EJ, Finnegan PM. **2023**. Dihydroxyacetone in the Floral Nectar of *Ericomyrtus serpyllifolia* (Turcz.) Rye (Myrtaceae) and *Verticordia chrysantha* Endl. (Myrtaceae) Demonstrates That This Precursor to Bioactive Honey Is Not Restricted to the Genus *Leptospermum* (Myrtaceae). *Journal of Agricultural and Food Chemistry* **71**: 7703–7709.

Ren G, Healy RA, Klyne AM, Horner HT, James MG, Thornburg RW. **2007**. Transient starch metabolism in ornamental tobacco floral nectaries regulates nectar composition and release. *Plant Science* **173**: 277–290.

Roy R, Schmitt AJ, Thomas JB, Carter CJ. **2017**. Review: Nectar biology: From molecules to ecosystems. *Plant Science* **262**: 148–164.

Ruhlmann JM, Kram BW, Carter CJ. **2010**. CELL WALL INVERTASE 4 is required for nectar production in *Arabidopsis*. *Journal of Experimental Botany* **61**: 395–404.

Solhaug EM, Roy R, Chatt EC, Klinkenberg PM, Mohd-Fadzil N-A, Hampton M, Nikolau BJ, Carter CJ. **2019**. An integrated transcriptomics and metabolomics analysis of the *Cucurbita pepo* nectary implicates key modules of primary metabolism involved in nectar synthesis and secretion. *Plant Direct* **3**: e00120.

Williams S, King J, Revell M, Manley-Harris M, Balks M, Janusch F, Kiefer M, Clearwater M, Brooks P, Dawson M. **2014**. Regional, Annual, and Individual Variations in the Dihydroxyacetone Content of the Nectar of Mānuka (*Leptospermum scoparium*) in New Zealand. *Journal of Agricultural and Food Chemistry* **62**: 10332–10340.

Chapter 7: Synthesis

7.1.1 Thesis summary

My thesis aimed to elucidate the molecular basis of dihydroxyacetone (DHA) production in *Leptospermum* nectar. We aimed to answer the following questions:

- What genes and genomic regions are associated with nectar DHA in *L. scoparium*?
- What genetic diversity is present in contrasting DHA phenotypes within the Leptospermeae tribe?
- Does *Sgpp2* expression correlate with DHA amounts in other Leptospermeae species?
- How is *Sgpp2* expression regulated?
- Can SGPP2 dephosphorylate DHAP to produce DHA and phosphate *in vitro*?
- Can overexpression of LmSGPP2 in *Petunia* and *Nicotiana* nectaries produce DHA in nectar?

Chapter 2 aimed to identify genes and genomic regions associated with nectar DHA in *Leptospermum scoparium*. We used RNAseq analysis to identify nectary-associated genes differentially expressed between high and low nectar-DHA genotypes. We also used a mānuka high-density linkage map and quantitative trait loci (QTL) mapping population, supported by an improved genome assembly of *L. scoparium*, to reveal genetic regions associated with nectar DHA content. Expression and QTL analyses both pointed to the involvement of a phosphatase gene, *LsSgpp2*. The expression pattern of *LsSgpp2* correlated with nectar DHA accumulation in a range of genotypes, and it co-located with a QTL on chromosome 4. This chapter established that plant genetics is a key influence on DHA accumulation, and the identification of three QTLs indicated polygenic control of nectar DHA content. Based on the proposed function of the LsSGPP2 orthologue, we suggested that LsSGPP2 releases DHA from DHAP for the purpose of maintaining phosphate homeostasis in the photosynthesising nectary. We hypothesised that variability in *Sgpp2* gene expression was responsible for the variation in nectar DHA observed among different *Leptospermum* genotypes and species, and that this variation would be caused by sequence differences within the *Sgpp* locus.

To enable cross-species comparisons of the *Sgpp* locus, in **Chapter 3** we produced whole genome sequences for two more species within the Leptospermeae tribe. High-quality haplotyped *Gaudium laevigatum* and *Leptospermum morrisonii* genome assemblies were produced using high-fidelity long-read sequencing and synteny-based scaffolding against the *Leptospermum scoparium* genome. These genomes were selected to represent contrasting DHA phenotypes within the Leptospermeae as *L. morrisonii* has the presence of nectar DHA and *G. laevigatum* does not. These new genome

assemblies provide an important resource for studying traits related to nectar and foliage chemistry, and for understanding the genetic diversity within the Leptospermeae.

Following the identification of *Sgpp2* in Chapter 2 and the generation of more Leptospermeae genomes in Chapter 3, **Chapter 4** aimed to further understand the regulation of *Sgpp2* gene expression and the structure of the *Sgpp* locus within the Leptospermeae and wider Myrtaceae family. The *Sgpp* locus, gene sequences, and transcript abundances were compared among three Leptospermeae species– *L. scoparium*, *L. morrisonii*, and *G. laevigatum*. By comparing *Sgpp* gene promoters within individual genomes and between the different genomes, both conserved and unique regions were identified. Regions and binding motifs unique to the promoters of the highly expressed *Sgpp2*s were identified. To further characterise the *LmSgpp2* promoter we generated transgenic lines of *Petunia* and *Nicotiana* – using the *LmSgpp2* promoter driving expression of the reporter gene *Venus*. *Sgpp2* expression correlated with DHA amounts in the two further Leptospermeae species investigated, further supporting the hypothesis that *Sgpp2* is responsible for DHA production. Comparison of *Sgpp* gene sequences from throughout Myrtaceae indicated that that high *Sgpp2* expression was probably the ancestral state within DHA producing taxa, followed by repeated loss of the trait. Two C-box motifs were identified as being unique to the highly expressed *Sgpp2*s, and a bZIP11 transcription factor predicted to bind to them was also identified as significantly differentially expressed in the RNAseq dataset presented in Chapter 2. The bZIP11 gene was subsequently found to co-locate with another of the QTLs identified in Chapter 2. The promoter region across 13 DHA producing *L. scoparium* genotypes was found to be highly conserved in sequence, with the lack of variable proximal *cis*-elements indicating that there may be further complex transcriptional regulation fine tuning *Sgpp2* expression beyond the promoter region analysed. A limitation of this chapter was the inclusion of only one line that produces no DHA – and future analyses would be benefit from inclusion of even more extreme DHA phenotypes. The *LmSgpp2promoter-Venus* lines of *Petunia* and *Nicotiana* were found to have strong nectary specific expression indicating conserved regulatory elements in these diverse species. It was hypothesised that the presence of the C-box motifs may be essential for activation of *Sgpp2* expression by the bZIP11 transcription factor, and other regulatory elements within the QTL but outside of the promoter region studied may fine tune expression.

Chapter 5 characterised the activity of the *L. morrisonii* SGPPs and compared them with the previously characterised *Arabidopsis* SGPP. These three LmSGPP enzymes, along with AtSGPP, were produced and purified in microbial systems, and assayed for phosphatase activity on three substrates that included DHAP. It was demonstrated that LmSGPP2 can dephosphorylate DHAP to produce DHA

and phosphate. LmSGPP2 had higher activity on DHAP than AtSGPP and LmSGPP1, and similar activity to LmSGPP3. I hypothesised that shifts in the kinetic parameters between the SGPPs were due to changes in the substrate specificity motif. These assays provided preliminary insight into the range of substrates *Sgpp2* can dephosphorylate, and further substrate screening could allude to additional functions. These results validated our hypothesis that SGPP2 dephosphorylates DHAP and further supported its proposed role in phosphate homeostasis and the production of DHA in nectaries.

Following the *in vitro* characterisation of LmSGPP2 activity in Chapter 5, **Chapter 6** aimed to characterise LmSGPP2 *in vivo*. We generated *Petunia* and *Nicotiana* transgenic lines overexpressing the *L. morrisonii Sgpp2* under the nectary specific promoter *PaNec1*/*Sweet9* to assess the effect of *Sgpp2* expression in nectaries that do not produce DHA. No increase in DHA was found in the nectar of these transgenic lines. However, the *LmSgpp2* transgene transcript abundance was significantly lower than that of the control lines using the same promoter to control a reporter gene, which could explain the lack of DHA in the nectaries. Furthermore, the *PaNec1* promoter was found to be not as strongly expressed nor as nectary specific as previously described in the literature. A highly expressed nectary specific promoter (such as the *LmSgpp2* promoter used in Chapter 2) should be used for future work. It was hypothesised that as LmSGPP2 has been demonstrated in Chapter 5 to dephosphorylate DHAP and potentially a range of other sugar phosphate substrates, it may have been downregulated due to disruption of metabolism in these tissues. The effect of LmSGPP2 on nectar traits in *Petunia* and *Nicotiana* was unable to be assessed here due to low and mis-located expression of the transgene.

7.1.2 SGPP2 produces DHA in *Leptospermum* nectaries

This thesis has provided strong evidence that *Sgpp2* is responsible for DHA in *Leptospermum* nectar (Fig. **7.1**). *Sgpp2* expression levels correlated with DHA amounts in all species and genotypes we have looked at. It would be surprising if these two unique features of *Leptospermum* nectaries were unrelated, especially as I also demonstrated that SGPP2 is able to produce DHA from DHAP. DHAP is present in cells as part of primary metabolism and could have increased abundance due to photosynthesis in the nectary (Macrae & Lunn, 2006; Clearwater *et al.*, 2021; Mallén-Ponce *et al.*, 2025). Finally, the location of the *Sgpp2* gene within a QTL for nectar DHA provides even more support for this gene being responsible. However, as the data is correlative, confirmation of *Sppg2* function requires evidence from either loss-of-function or gain-of-function plant lines.

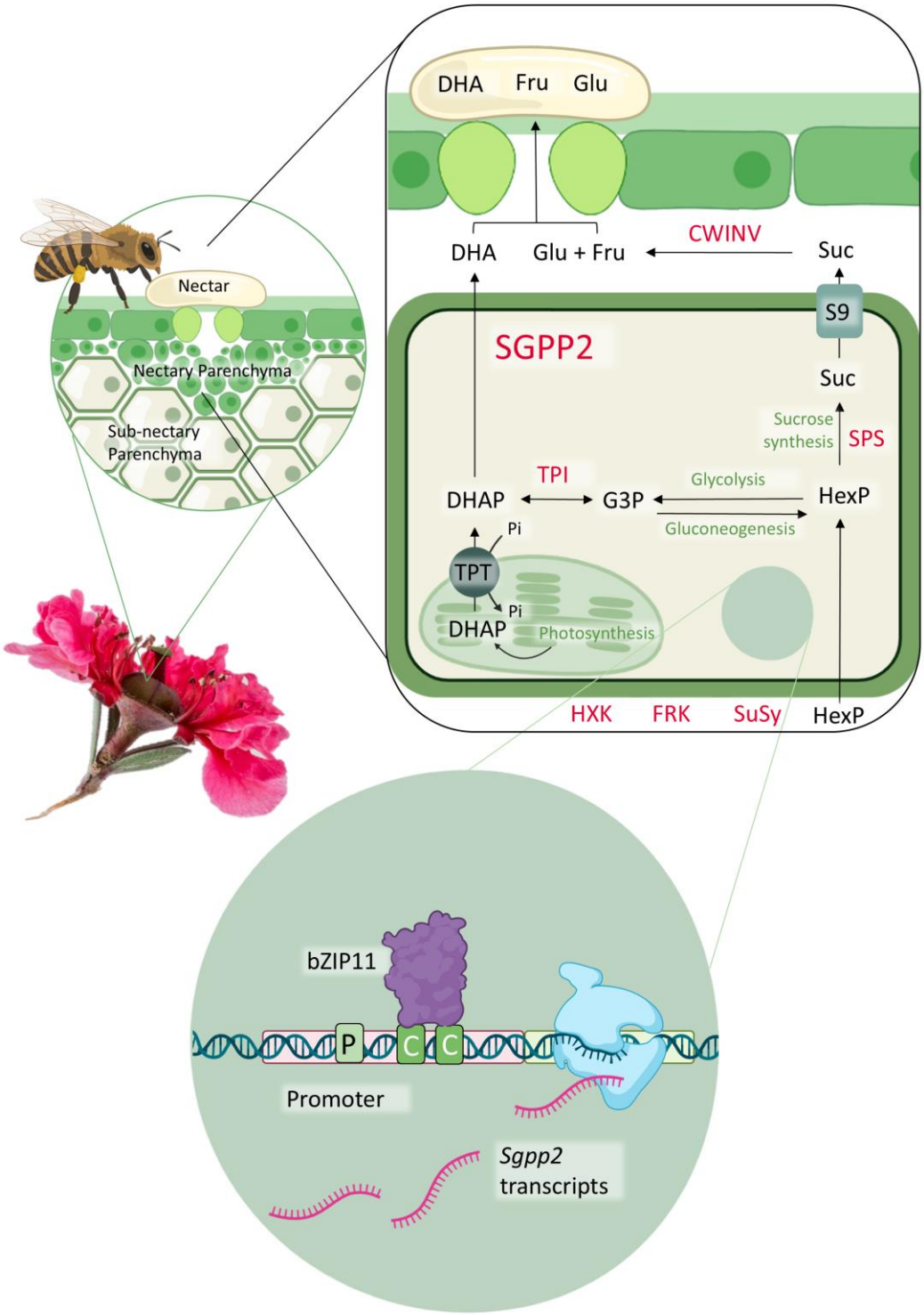


Figure 7.1: Simplified proposed mānuka (*Leptospermum scoparium)* nectar production model (based on Clearwater *et al.* 2021), with contributions from this thesis. In this model, the nectary sugars are unloaded into the subnectary parenchyma then enter the hexose phosphate (HexP) pool and primary metabolic pathways in nectary parenchyma via sucrose synthase (SuSy), hexokinase (HXK) or fructokinase (FRK). The nectary HexP pool is also contributed to by photosynthesis in nectary chloroplasts. Photosynthesis derived DHAP is exported via the Triose Phosphate Translocator (TPT) and is subsequently isomerised by Triose Phosphate Isomerase (TPI) to form glyceraldehyde-3-phosphate (G3P) and is further converted to HexP by gluconeogenesis. Some of this exported DHAP is dephosphorylated by an enzyme we have identified in the present work called SGPP2, to form DHA which then diffuses into the apoplast and into the nectar. *Sgpp2* gene expression is potentially regulated by a bZIP11 type transcription factor binding to C-box (Shown as C in promoter) motifs in its promoter, also identified in this present work. P in the promoter corresponds to the identified P1BS element. Figure created with some components from BioRender

7.1.3 Complex regulation of DHA production in *Leptospermum*

This thesis has identified multiple elements involved in the molecular basis of DHA production in *Leptospermum*. *Sgpp2* transcript abundance correlates with DHA amounts in nectar, therefore regulatory elements influencing its expression must be key in the production of DHA, along with the gene itself. The identification of two C-box motifs and a significantly differentially expressed bZIP11 transcription factor could be essential in the regulation of *Sgpp2* (Fig. **7.1**). These type of bZIPs often have diverse roles including involvement in sugar metabolism, photomorphogenesis and signal transduction (Wang *et al.*, 2022). The colocation of both the *Sgpp2* and the *bZIP11* in separate QTLs for nectar DHA provides further support of their significance. The ability of the 3.3kb of *Sgpp2* promoter to induce transgene expression in phylogenetically distant species demonstrates that key elements for strong nectary specific expression are within this region of the *LmSgpp2* promoter, and that whatever regulatory elements needed for *Sgpp2* to be expressed in nectaries are conserved among distant plant families.

The presence of multiple other motifs in the promoter could be indicative of complex recruitment of transcription factors. The P1BS motifs found throughout the promoter are usually involved in phosphate starvation responses, including adaptation to high-light stress and preserving photosynthesis (Müller *et al.*, 2007; Bustos *et al.*, 2010; Nilsson *et al.*, 2012). The GT-1 motifs are generally involved in light response pathways (Lam & Chua, 1990; Kaplan-Levy *et al.*, 2012). Even within an individual genotype, variation in nectar DHA is observed within and across seasons, indicating that environmental variables such as light could also be at play, as demonstrated by Clearwater *et al.* (2021). This *Sgpp2* expression could be further controlled though upregulation of light responsive transcription factors in response to changing light environments.

The lack of variation in the promoter observed among *L. scoparium* genotypes with varying DHA phenotypes indicates that other regulatory factors beyond the immediately adjacent promoter region such as distal *cis*-elements further upstream within the QTL, may be essential in further fine-tuning expression. Next steps to establish more specific regions within the QTL could include additional fine mapping (discussed further below) or technologies such as chromatin accessibility profiling (e.g. ATAC-seq). Variation was seen within the final intron of the *Sgpp2* gene which could also be part of the fine tuning of transcription, as demonstrated in other genes including AGAMOUS (Sieburth & Meyerowitz, 1997; Marand *et al.*, 2023).

The most probable origin of *Sgpp2* was by gene duplication then neofunctionalisation. It is currently unknown when exactly the *Sgpp2* duplication has occurred, but our analysis showed two Sgpp2 copies in *Punica granatum* – a species outside of the Myrtaceae but within the same order. The C-

box motifs could have evolved from a series of SNPs or insertions, enabling recruitment of the bZIP11 transcription factor potentially already active in carbohydrate metabolism within the nectary (Marand *et al.*, 2023). The high expression of *Sgpp2* suggests that it could have conferred an adaptive advantage (discussed further below). However the repeated loss of the nectar DHA trait observed among the Leptospermeae and Chamelaucieae tribes could indicate fluctuating selection pressures in different ecological contexts (Noe *et al.*, 2019; Obeng-Darko *et al.*, 2023). Analysis of the *Sgpp2* expression profiles and promoter sequences of more taxa with varying DHA phenotypes could help to address this hypothesis.

7.1.4 Function of *Sgpp2* in *Leptospermum* nectaries

Leptospermum nectaries are photosynthetic, and photosynthesis derived DHAP contributes to nectar DHA (Clearwater *et al.*, 2021). Changing light levels could create sudden increased in DHAP within the cell – potentially necessitating release of phosphate to ensure sufficient phosphate is available for exchange by the chloroplastic triose phosphate transporter (TPT) to preserve photosynthetic machinery (Tegeder & Weber, 2006). The *Arabidopsis* orthologue of SGPP2 has proposed functions relating to modulation of sugar-phosphate balance and Pi homeostasis (Caparrós-Martín *et al.*, 2013; He *et al.*, 2019), and Pi starvation induces high expression levels of *AtSgpp* (Caparrós-Martín *et al.*, 2013). Many of the motifs identified are associated with light responsive and/or phosphate starvation pathways – C-box (Song *et al.*, 2008), P1BS (Nilsson *et al.*, 2012; Liu *et al.*, 2017) and GT-1 (Lam & Chua, 1990; Terzaghi & Cashmore, 1995; Kaplan-Levy *et al.*, 2012), supporting a role related to phosphate homeostasis and photosynthesis.

Not all green photosynthesising nectaries have DHA in their nectar (e.g. van Delden, 2024), so what is unique to these DHA producing taxa? A possible answer to this could relate to their geographical distribution. Most of the Leptospermeae is centered in Australia, which has a high proportion of highly weathered and nutrient poor soils and specifically soils that are low in phosphorus (Orians & Milewski, 2007). The Australian flora has many unique features thought to have evolved in response to nutrient poverty, including some related to nectar traits (Orians & Milewski, 2007). Mānuka also maintains rapid growth with comparatively low requirements for key nutrients such as phosphate (Gutiérrez-Ginés *et al.*, 2019; Noe, 2022), behaviour that suggests efficient strategies to maintain sufficient Pi in its tissues. High *Sgpp2* expression and the resulting nectar DHA may have been retained within some lineages of the Leptospermeae and related clades by providing an advantage in maintaining phosphate homeostasis within floral tissues, in response to the combination of a photosynthetic nectary and low-phosphate availability.

It is still not entirely clear what fitness advantage, if any, *Sgpp2* expression and nectar DHA accumulation provides. We have hypothesised that DHA in nectar is a byproduct of phosphate homeostasis, however other possible explanations include DHA in the nectar having value to a pollinator or beneficial microbe, or as a signaling molecule. DHA as a root exudate has been found to be an attractant for a nematode (Wang *et al.*, 2020). DHAP has been demonstrated to be part of signaling pathways – such as activation of TOR (Mallén-Ponce *et al.*, 2025) and MGO – of which DHA is the precursor, has been demonstrated to induce a triose phosphate isomerase in rice (Sharma *et al.*, 2012). Sugars produced by photosynthesis can be sensed as signals (Petrillo *et al.*, 2014), and Clearwater *et al.* (2021) also hypothesised that photosynthesis in nectaries could have signaling purpose – potentially amplifying nectar production from phloem derived sources in response to a diurnal light signal. It cannot be excluded that DHA production in nectar may have no biological function: SGGP2 activity and nectary-expression may have evolved for another function that is no longer biologically relevant, and the sporadic occurrence reflects a trait gradually being lost from the taxa.

7.1.5 Significance for nectar biology

Most of our knowledge of nectar and nectaries is for model species, and research on non-model species like mānuka is crucial for widening our understanding. This thesis reports the first QTLs and genes linked to a minor abundance compound of nectar, where so far the identification of QTLs related to nectar and nectaries has been limited to effects on nectary size (Liao *et al.*, 2021) and nectar volume (Barstow *et al.*, 2022, 2025). This work also contributes to our developing understanding of photosynthetic nectaries. We have built on the work from Clearwater *et al.* (2021), by further characterising the proposed path from cellular DHAP to nectar DHA. Clearwater *et al.* (2021) hypothesised that photosynthesising chloroplasts in the nectary could have a role in regulating nectar flow and composition, as the contribution of nectary photosynthesis was less than the total increase in nectar sugar in response to light. We have identified a gene potentially linked to nectary chloroplasts, adding to our understanding of photosynthesising nectaries and the role of nectary plastids beyond amyloplasts (e.g. Ren *et al.*, 2007) or chromoplasts (e.g. Horner *et al.*, 2007). Further clarification of the unique role of nectary chloroplasts in *Leptospermum* contributes to our overall understanding in this understudied area.

7.1.6 Future impact for the mānuka honey industry

The methylglyoxal (MGO) levels in mānuka honey, and therefore nectar DHA amounts, directly underpin the mānuka honey industry and influence the export prices able to be achieved by honey producers. Increasing MGO levels of mānuka honey is a key goal of the industry, and the knowledge

generated within this thesis provides essential insights into the plant genetic basis of this trait. We have identified multiple genetic elements contributing to nectar DHA production, which can be further explored to develop gene-based tools to identify high-value germplasm. Such tools would enable efficient screening and selection of many lines, ensuring biodiversity can be maintained while increasing high-value honey production.

Ideally, identification of genomic sequence differences correlating with DHA production that could be converted to appropriate markers (e.g. for High Resolution Melt analysis or 'gene chips') would provide the basis for identification of high value mānuka germplasm. However, specific sequence differences have not been identified from the present work. Nevertheless, as high nectar-DHA is linked to high transcript abundance of the *Sgpp2* gene, the knowledge generated here could be used for a mRNA based phenotyping system. Such methodology, however, may not offer significant advantage over nectar-DHA phenotyping. Further development of genotyping methodology could help provide reliable identification of germplasm for breeding programmes and plantation planting, through clonal repropagation or selection of mother plants for harvesting seed capsules for replanting.

Future work is needed to identify specific sequence differences to underpin genomic tools. To clarify more specific genomic regions to focus on, further QTL mapping populations could be created. Limitations of our previous population include not drastically different DHA phenotypes of parental lines, and environmental variability potentially masking some of the QTL effects. New populations generated should expand the number of genotypes and use crosses between lines with more strongly contrasting DHA phenotypes. With further sequence information, crosses could be made between two high-DHA producing lines homozygous for a DHA marker, allowing us to then explore the other QTLs that may have been masked by it.

This work has been done as part of a government funded Smart Idea ('Elucidating key mānuka genes determining honey value') in partnership with Ngāti Porou Miere Rangahau, alongside a Sustainable Food and Fibre Futures funded project (Ngāti Porou mānuka R&D extension programme') to establish mānuka propagation capability in Tairāwhiti – East Cape. This has provided insight into how this knowledge can be utilised by the end users in the industry. This partnership has emphasised that being able to eco-source germplasm from within a rohe needs to be a key driver of future work. Eco-sourcing ensures plants will be well adapted to their environment, including flowering at appropriate times to be able to produce high-quality honey. It also ensures regional diversity can be maintained to ensure resilient populations. Respecting whakapapa and maintaining provenance is also important

for Ngāti Porou Miere Rangahau, as well as supporting the development of distinct provenance stories and products.

7.1.7 Future work

The discovery of multiple QTLs and genes involved in DHA production in *Leptospermum* species provides the basis for numerous future research directions. Some possibilities to expand this research space are discussed below.

What are the expression patterns of Sgpp2 *in wider taxa?*

We hypothesise that the ancestral trait is to have a highly expressed *Sgpp2* and nectar DHA, with repeated loss of the trait. To further test this, we could investigate *Sgpp*2s presence and expression in a wider range of taxa throughout the Leptospermeae and Chamelaucieae tribes. It would be interesting to see if the correlation between *Sgpp2* expression and nectar DHA continues in taxa with two to three times higher nectar DHA than we see in mānuka (e.g. *Aggreflorum whitei,* Williams *et al.*, 2018)

Is Sgpp3 *highly expressed in some taxa?*

As discussed above, we hypothesise that the ancestral state is high expressing copy of *Sgpp2* followed by *Sgpp3* arising through duplication – *Sgpp3* may have retained that high expression in some taxa. It would be interesting to establish if this could be a factor in the DHA levels seen within some of the highest DHA producing species (see Williams *et al.*, 2018).

How can we further investigate the regulation of Sgpp2 *expression?*

Further work could investigate the regulatory elements fine tuning *Sgpp2* expression. This could involve fine mapping of the QTLs to further refine the linked locus, or creation of further mapping populations described above. Other regulatory elements could be identified though technologies such as chromatin accessibility profiling (e.g. ATAC-seq) or validation transcription factor binding (e.g. CHIP-seq) (Marand *et al.*, 2023). Promoter dissection experiments could help narrow down the region of the promoter required for nectary specificity.

Are the C-box motifs key to high Sgpp2 *expression?*

Analysis of wider taxa within the Leptospermeae and Chamelaucieae should also focus on the C-box motifs identified – and establish if they are in fact crucial for the presence or absence of DHA. Ideal genera to start this investigation would be *Aggreflorum* and *Verticordia* which contain very high nectar DHA lines as well as lines with no nectar DHA. Further transgenic experiments with a

LmSgpp2promoter-Venus construct with one or both of the C-box elements removed could also confirm the C-box motifs role in strong baseline expression. Yeast one hybrid experiments could also confirm interaction of these motifs with the bZIP.

Can we demonstrate Sgpp2 *function* in vivo*?*

Ways to further demonstrate *Sgpp2* function include further overexpression experiments using the *LmSgpp2* promoter to drive expression of the *LmSgpp2* gene. These transgenics could initially be in *Petunia* and *Nicotiana* due to established transformation protocols, but it is unknown how crucial photosynthesis derived DHAP may be in DHA accumulation. Therefore ideally, a nectary model with photosynthesising nectaries would be used, such as ivy (*Hedera helix*). Stable transgenics to demonstrate function in non-DHA producing *Leptospermum* species would take significant time to develop transformation protocols and would take years to flower and assess nectar DHA. Transient expression assays in the flowers of Leptospermeae species could potentially be developed more quickly.

Can Sgpp2 *dephosphorylate other substrates?*

Additional substrate screening could help to further understand the activity of SGPP2, as there could be a currently unscreened substrate that SGPP2 has even higher phosphatase activity on. If one was found it could allude to another function, where the dephosphorylating of DHAP is not its primary purpose.

How does phosphate starvation affect nectar DHA?

Previous work has found that low phosphate soils did not affect nectar traits (Noe, 2022), however, perhaps even lower levels of phosphate are needed to see an effect in mānuka. Experiments with very low supplied phosphate could help to simulate the possible environments this trait could have evolved in and clarify the role of *Sgpp2* in phosphate homeostasis. Subjecting phosphate deprived flowers to varying light levels and measuring gene expression, nectar traits, and tissue Pi levels could also shed light on how *Sgpp2* expression might be influenced by both Pi starvation and photosynthesis.

Is the function of Sgpp2 *maintaining phosphate homeostasis in a photosynthesising nectary?*

Much of our understanding of photosynthesis is based on leaves, where as long as sucrose synthesis keeps up with CO_2 assimilation and triose phosphate export from the chloroplast, sufficient Pi should be released for exchange by the chloroplastic Triose Phosphate Translocator (Tegeder & Weber, 2006). Our understanding is limited in how photosynthesis and nectar production affect each other

in photosynthesising nectaries, and should be the subject of future research. We hypothesise that in *Leptospermum*, the combination of photosynthesis and nectar production in the nectaries combined with low phosphate availability is potentially causing a depletion in cytosolic Pi that is not being replenished by sucrose synthesis – necessitating SGPP2 to release Pi for exchange and preserve photosynthetic machinery. To test this hypothesis, we could look at gene expression changes in both photosynthetic and nectar production pathways in response to phosphate starvation (discussed above), varying light levels, and photosynthesis inhibitors such as DCMU (see Clearwater *et al.,* 2021).

What other factors could explain the evolution of the nectar DHA trait?

Understanding the scope of *Sgpp2* presence within the Myrtaceae could help us understand function. While we have hypothesised that DHA in nectar is a byproduct of phosphate homeostasis, there may be other adaptive significance that is still unknown. One possible advantage could be related to a pollinator interaction, that could potentially be explored through comparisons of distribution of DHA producing plants and pollinators.

Are microbes contributing to DHA production in nectar?

Investigating the potential role of microbes (e.g. *Pantoea agglomerans*, Kauffman-Bresciano, 2021) augmenting DHA production in nectar could also be explored further, especially as microbes can sometimes mimic the production of host-produced compounds (Digra & Nonzom, 2023), and (Larrouy *et al.*, 2023) characterised the floral microbiome of mānuka and found a wide diversity of taxa both endophytic and epiphytic.

7.1.8 Concluding comments

This thesis strongly indicated that *Sgpp2* is responsible for DHA in *Leptospermum* nectar, and that it is under complex regulation. We hypothesise the function of maintaining phosphate homeostasis in a photosynthesising nectary but further investigation is required. Future research can help us clarify the basis of this unique nectar trait in a non-model species and contribute to development of molecular tools for identification of high value germplasm to support the Aotearoa-New Zealand honey industry.

7.2 References

Barstow AC, McNellie JP, Smart BC, Keepers KG, Prasifka JR, Kane NC, Hulke BS. **2025**. Variant filters using segregation information improve mapping of nectar-production genes in sunflower (*Helianthus annuus* L.). *The Plant Genome* **18**: e70042.

Barstow AC, Prasifka JR, Attia Z, Kane NC, Hulke BS. **2022**. Genetic mapping of a pollinator preference trait: Nectar volume in sunflower (*Helianthus annuus* L.). *Frontiers in Plant Science* **13**: 1056278.

Bustos R, Castrillo G, Linhares F, Puga MI, Rubio V, Pérez-Pérez J, Solano R, Leyva A, Paz-Ares J. **2010**. A Central Regulatory System Largely Controls Transcriptional Activation and Repression Responses to Phosphate Starvation in *Arabidopsis*. *PLoS Genetics* **6**: e1001102.

Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. **2013**. HAD hydrolase function unveiled by substrate screening: enzymatic characterization of *Arabidopsis thaliana* subclass I phosphosugar phosphatase AtSgpp. *Planta* **237**: 943–954.

Clearwater MJ, Noe ST, Manley-Harris M, Truman G-L, Gardyne S, Murray J, Obeng-Darko SA, Richardson SJ. **2021**. Nectary photosynthesis contributes to the production of mānuka (*Leptospermum scoparium*) floral nectar. *New Phytologist* **232**: 1703–1717.

van Delden JM. **2024**. Nectar variation in Aotearoa New Zealand tree species. PhD thesis, University of Waikato, Hamilton, New Zealand.

Digra S, Nonzom S. **2023**. An insight into endophytic antimicrobial compounds: an updated analysis. *Plant Biotechnology Reports* **17:** 1–31.

Gutiérrez-Ginés MJ, Madejón E, Lehto NJ, McLenaghen RD, Horswell J, Dickinson N, Robinson BH. **2019**. Response of a Pioneering Species (*Leptospermum scoparium* J.R.Forst. & G.Forst.) to Heterogeneity in a Low-Fertility Soil. *Frontiers in Plant Science* **10**: 93.

He JZ, Dorion S, Lacroix M, Rivoal J. **2019**. Sustained substrate cycles between hexose phosphates and free sugars in phosphate-deficient potato (Solanum tuberosum) cell cultures. *Planta* **249**: 1319–1336.

Horner HT, Healy RA, Ren G, Fritz D, Klyne A, Seames C, Thornburg RW. **2007**. Amyloplast to chromoplast conversion in developing ornamental tobacco floral nectaries provides sugar for nectar and antioxidants for protection. *American Journal of Botany* **94**: 12–24.

Kaplan-Levy RN, Brewer PB, Quon T, Smyth DR. **2012**. The trihelix family of transcription factors – light, stress and development. *Trends in Plant Science* **17**: 163–171.

Kauffman-Bresciano CJ. **2021**. The role of microorganisms in shaping the nectar chemistry of Mānuka plants (*Leptospermum scoparium*). PhD thesis, Massey University, Palmerston North, New Zealand.

Lam E, Chua NH. **1990**. GT-1 binding site confers light responsive expression in transgenic tobacco. *Science* **248**: 471–474.

Larrouy JL, Dhami MK, Jones EE, Ridgway HJ. **2023**. Physiological stage drives fungal community dynamics and diversity in *Leptospermum scoparium* (mānuka) flowers. *Environmental Microbiology* **25**: 766–771.

Liao IT, Hileman LC, Roy R. **2021**. On the horizon for nectar-related research. *American Journal of Botany* **108**: 2326–2330.

Liu Y, Xie Y, Wang H, Ma X, Yao W, Wang H. **2017**. Light and Ethylene Coordinately Regulate the Phosphate Starvation Response through Transcriptional Regulation of PHOSPHATE STARVATION RESPONSE1. *The Plant Cell* **29**: 2269–2284.

Macrae E, Lunn J. **2006**. Control of Sucrose Biosynthesis. In: Plaxton WC, McManus MT, eds. *Control of Primary Metabolism in Plants*. Blackwell Publishing, 234–257.

Mallén-Ponce MJ, Quintero-Moreno AM, Gámez-Arcas S, Grossman AR, Pérez-Pérez ME, Crespo JL. **2025**. Dihydroxyacetone phosphate generated in the chloroplast mediates the activation of TOR by CO2 and light. *Science Advances* **11**: eadu1240.

Marand AP, Eveland AL, Kaufmann K, Springer NM. **2023**. *cis*-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annual Review of Plant Biology* **74**: 111–137.

Müller R, Morant M, Jarmer H, Nilsson L, Nielsen TH. **2007**. Genome-Wide Analysis of the *Arabidopsis* Leaf Transcriptome Reveals Interaction of Phosphate and Sugar Metabolism. *Plant Physiology* **143**: 156–171.

Nilsson L, Lundmark M, Jensen PE, Nielsen TH. **2012**. The *Arabidopsis* transcription factor PHR1 is essential for adaptation to high light and retaining functional photosynthesis during phosphate starvation. *Physiologia Plantarum* **144**: 35–47.

Noe ST. **2022**. The effects of soil nutrient addition on mānuka (Leptospermum scoparium) growth, flowering, and nectar production.

Noe S, Manley-Harris M, Clearwater MJ. **2019**. Floral nectar of wild mānuka (*Leptospermum scoparium*) varies more among plants than among sites. *New Zealand Journal of Crop and Horticultural Science* **47**: 282–296.

Obeng-Darko SA, Sloan J, Binks RM, Brooks PR, Veneklaas EJ, Finnegan PM. **2023**. Dihydroxyacetone in the Floral Nectar of *Ericomyrtus serpyllifolia* (Turcz.) Rye (Myrtaceae) and *Verticordia chrysantha* Endl. (Myrtaceae) Demonstrates That This Precursor to Bioactive Honey Is Not Restricted to the Genus *Leptospermum* (Myrtaceae). *Journal of Agricultural and Food Chemistry* **71**: 7703–7709.

Orians GH, Milewski AV. **2007**. Ecology of Australia: the effects of nutrient-poor soils and intense fires. *Biological Reviews* **82**: 393–423.

Petrillo E, Godoy Herz MA, Fuchs A, Reifer D, Fuller J, Yanovsky MJ, Simpson C, Brown JWS, Barta A, Kalyna M, *et al.* 2014. A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* **344**: 427–430.

Ren G, Healy RA, Horner HT, James MG, Thornburg RW. **2007**. Expression of starch metabolic genes in the developing nectaries of ornamental tobacco plants. *Plant Science* **173**: 621–637.

Sharma S, Mustafiz A, Singla-Pareek SL, Shankar Srivastava P, Sopory SK. **2012**. Characterization of stress and methylglyoxal inducible triose phosphate isomerase (OscTPI) from rice. *Plant Signaling & Behavior* **7**: 1337–45.

Sieburth LE, Meyerowitz EM. **1997**. Molecular dissection of the AGAMOUS control region shows that *cis* elements for spatial regulation are located intragenically. *The Plant Cell* **9**: 355–365.

Song YH, Yoo CM, Hong AP, Kim SH, Jeong HJ, Shin SY, Kim HJ, Yun D-J, Lim CO, Bahk JD, *et al.* 2008. DNA-Binding Study Identifies C-Box and Hybrid C/G-Box or C/A-Box Motifs as High-Affinity Binding Sites for STF1 and LONG HYPOCOTYL5 Proteins. *Plant Physiology* **146**: 1862–1877.

Tegeder M, Weber APM. **2006**. Metabolite Transporters in the Control of Plant Primary Metabolism. In: Plaxton WC, McManus MT, eds. *Control of Primary Metabolism in Plants*. Blackwell Publishing, 234–257.

Terzaghi WB, Cashmore AR. **1995**. Light-Regulated Transcription. *Annual Review of Plant Biology* **46**: 445–474.

Wang G, Wang Y, Abdelnabby H, Xiao X, Huang W, Peng D, Xiao Y. **2020**. Dihydroxyacetone of wheat root exudates serves as an attractant for *Heterodera avenae*. *PLOS ONE* **15**: e0236317.

Wang H, Zhang Y, Norris A, Jiang C-Z. **2022**. S1-bZIP Transcription Factors Play Important Roles in the Regulation of Fruit Quality and Stress Response. *Frontiers in Plant Science* **12**.

Williams SD, Pappalardo L, Bishop J, Brooks PR. **2018**. Dihydroxyacetone Production in the Nectar of Australian *Leptospermum* Is Species Dependent. *Journal of Agriculture and Food Chemistry* **66**: 11133–11140.

Supplementary material for Chapter 2

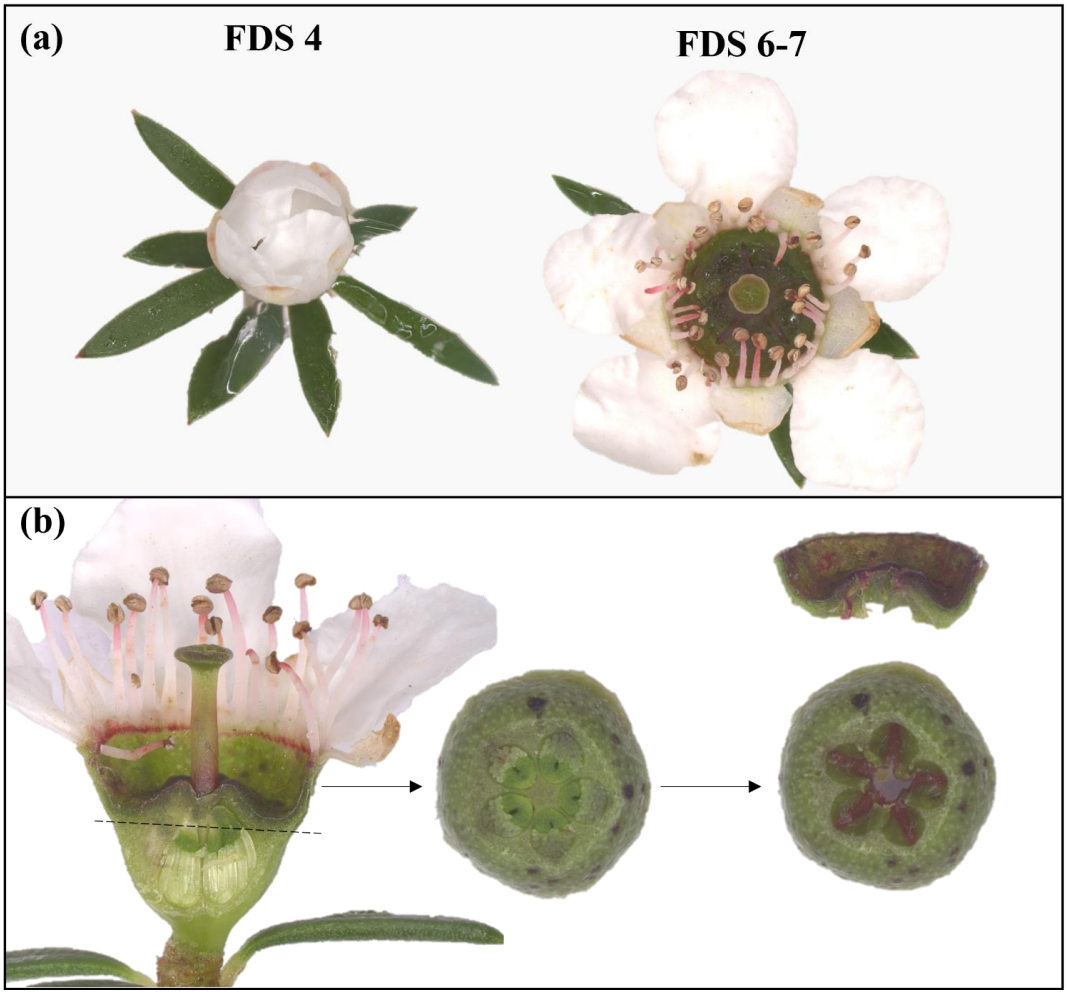


Figure S2.1: Developmental stages and nectary enriched tissue in *Leptospermum scoparium* line EC38. (a) Examples of FDS4 just prior to flower opening, and FDS6-7 nectar secreting flowers. (b) Cross section of mānuka flower showing the location of the transverse cut between the ovary and hypanthium surface, and the portion of ovules removed.

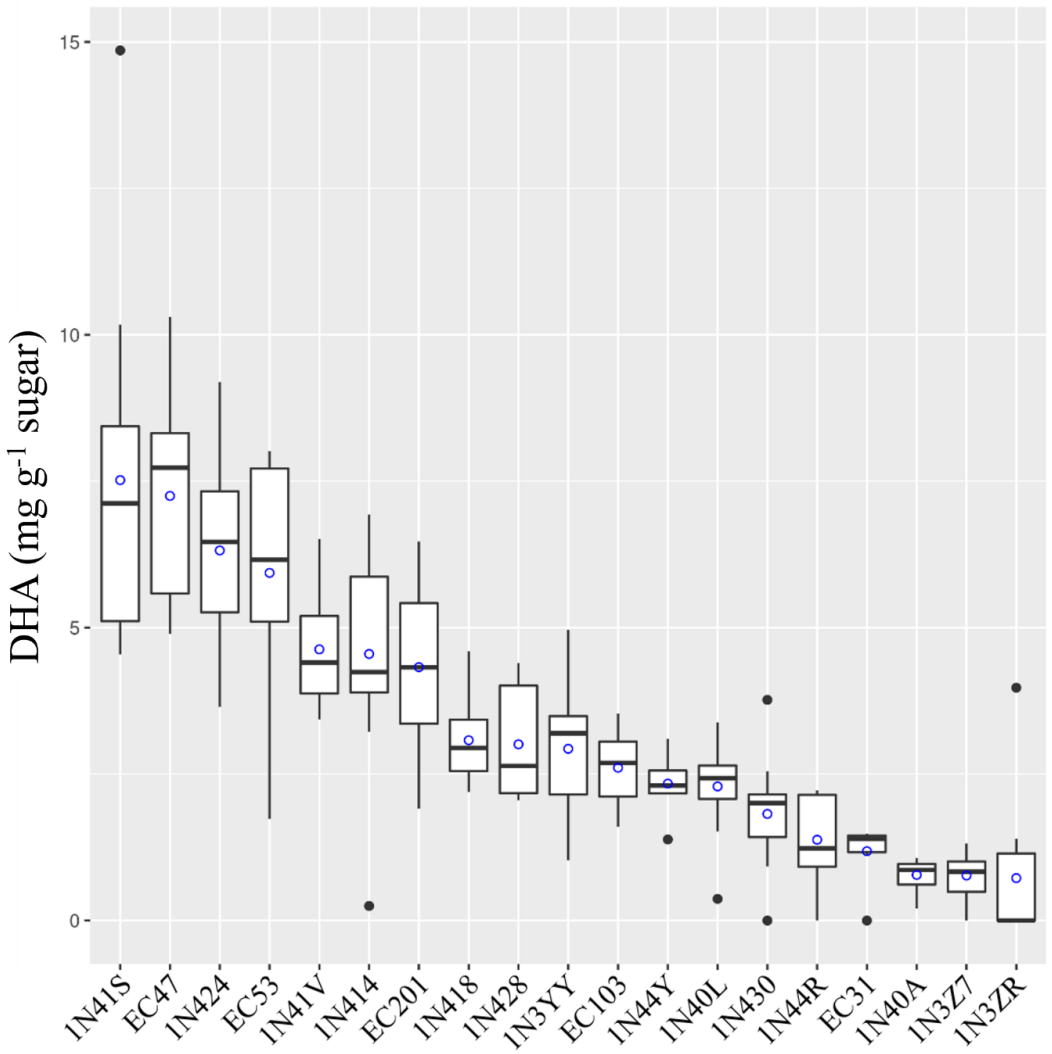


Figure S2.2: Dihydroxyacetone (DHA) nectar trait in a common garden design, for a range of mānuka (*Leptospermum scoparium*) genotypes across years DHA mg g^{-1} sugar in key lines, with at least 8 replicates across at least 3 seasons in total for each line. Blue circle shows the mean.

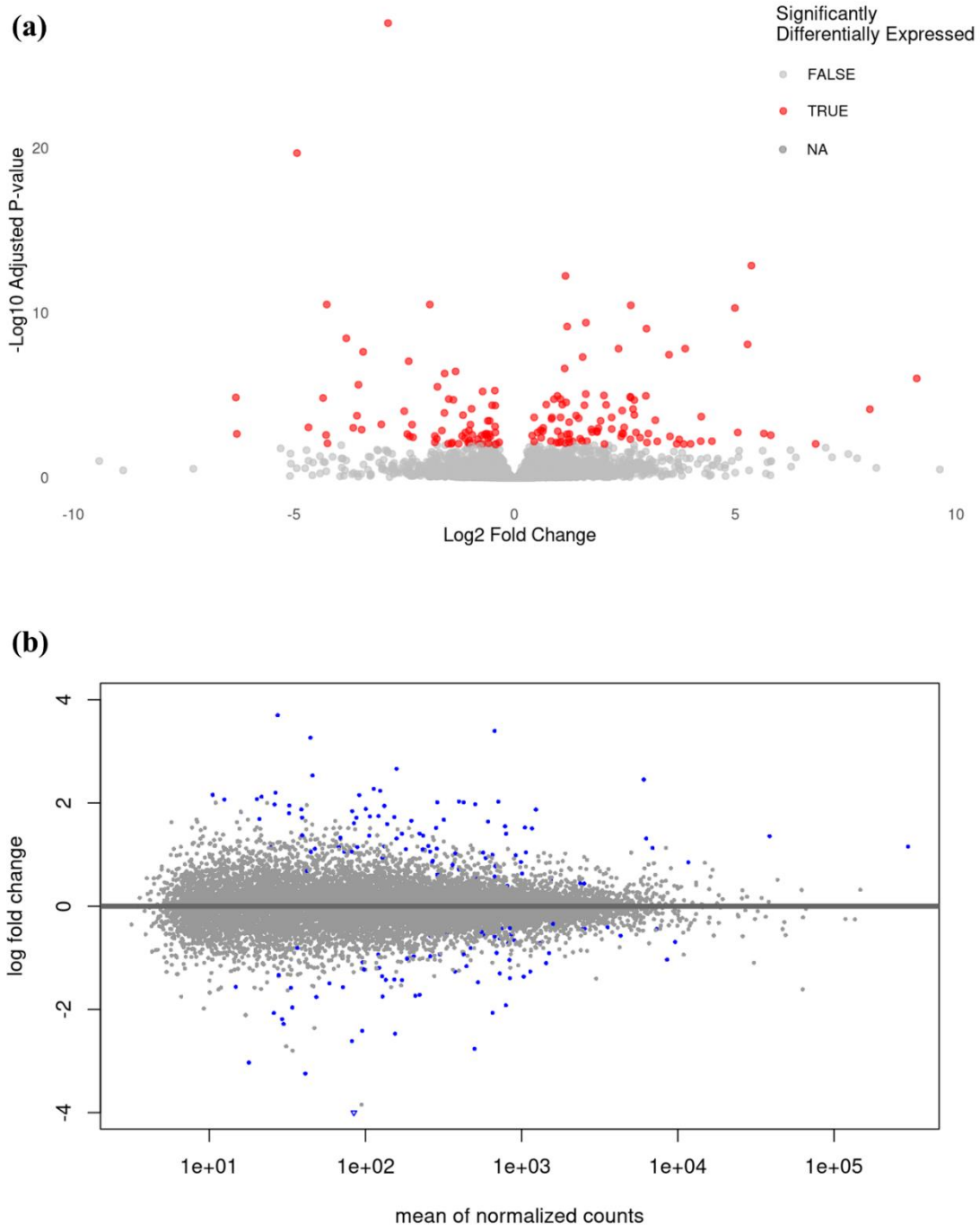
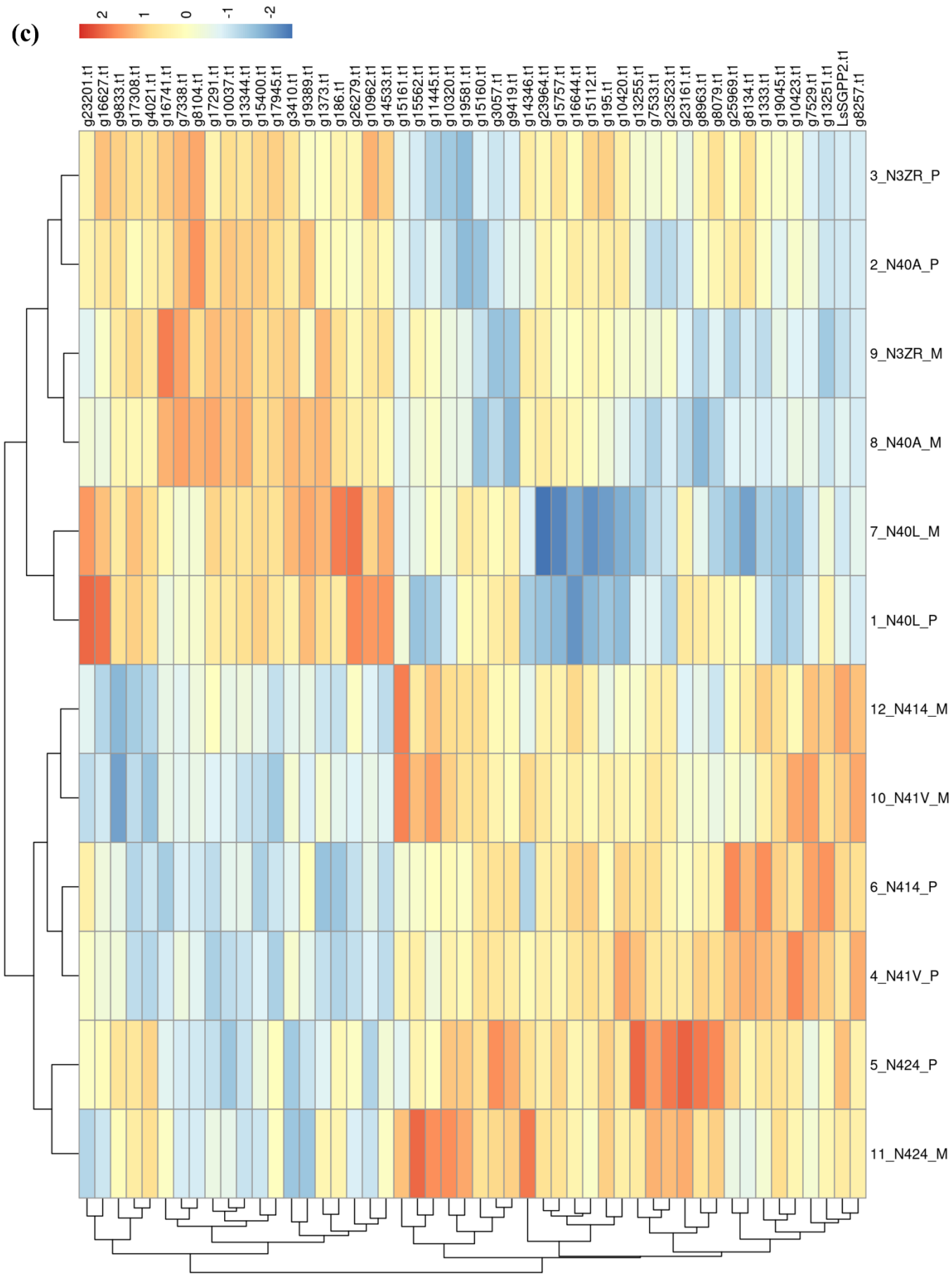


Figure S2.3: Visualisation of High DHA vs Low DHA RNAseq dataset. (a) Volcano plot of differentially expressed genes in mature tissue with an adjusted p-value of 0.01 or less highlighted in red. (b) MA plot of differentially expressed genes in mature tissue with an adjusted p-value of 0.01 or less highlighted in blue. (c) Heat map of top 50 differentially expressed genes from mature tissue, with both pre-secretion (P) and mature (M) samples shown. The corresponding counts were log transformed and the values were scaled by the row, and Euclidean distance was used for clustering of rows and columns. Continued on next page.


(c)
2
1
0
-1
-2
g23201.t1
g16627.t1
g9833.t1
g17308.t1
g4021.t1
g16741.t1
g7338.t1
g8104.t1
g17291.t1
g10037.t1
g13344.t1
g15400.t1
g17945.t1
g3410.t1
g19389.t1
g1373.t1
g186.t1
g26279.t1
g10962.t1
g14533.t1
g15161.t1
g15562.t1
g11445.t1
g10320.t1
g19581.t1
g15160.t1
g3057.t1
g9419.t1
g14346.t1
g23964.t1
g15757.t1
g16644.t1
g15112.t1
g195.t1
g10420.t1
g13255.t1
g7533.t1
g23523.t1
g23161.t1
g8963.t1
g8079.t1
g25969.t1
g8134.t1
g1333.t1
g19045.t1
g10423.t1
g7529.t1
g13251.t1
LsSGPP2.t1
g8257.t1
3_N3ZR_P
2_N40A_P
9_N3ZR_M
8_N40A_M
7_N40L_M
1_N40L_P
12_N414_M
10_N41V_M
6_N414_P
4_N41V_P
5_N424_P
11_N424_M

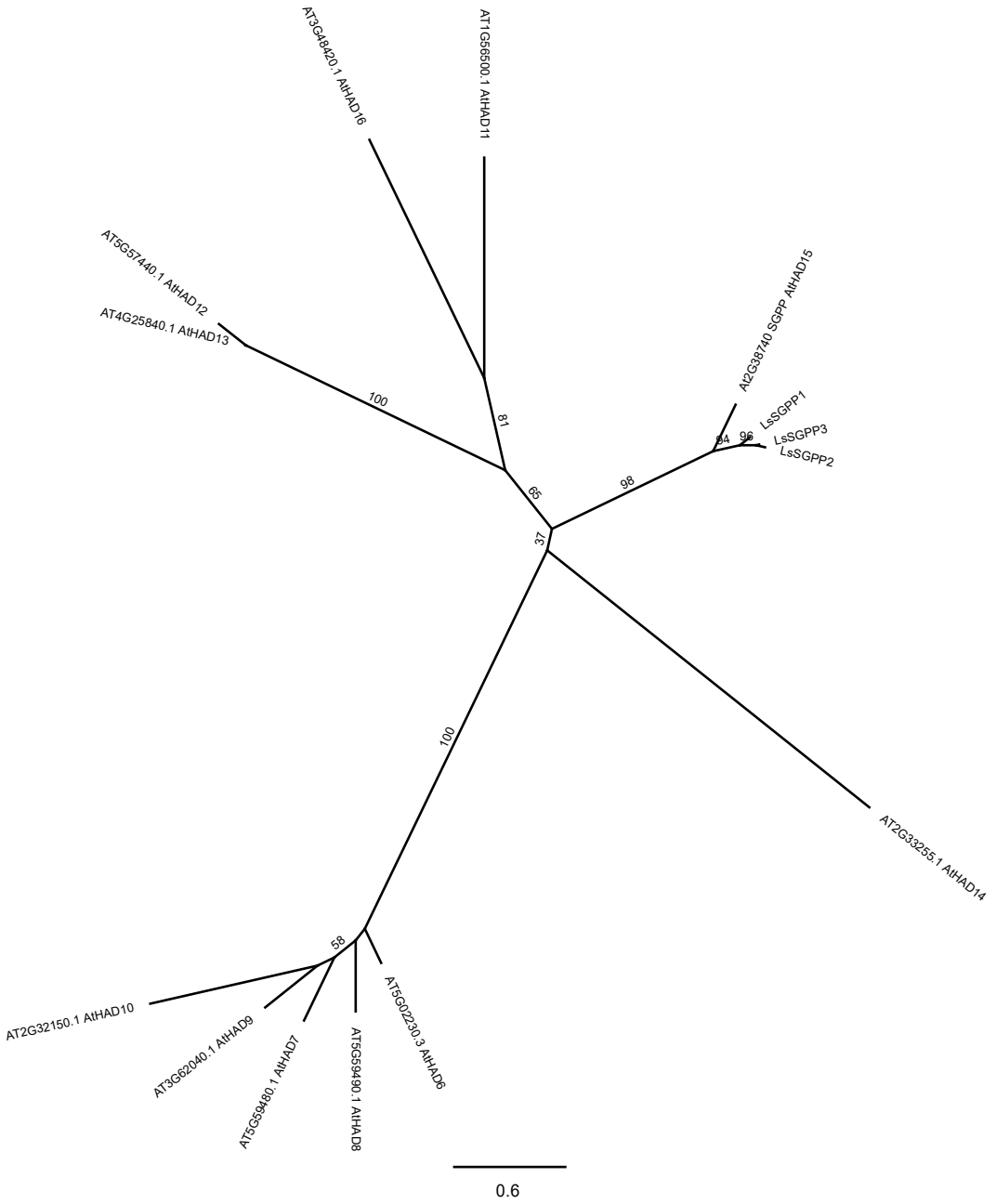


Figure S2.4: Phylogenetic relationships of LsSGPP1, LsSGPP2 and LsSGPP3 to related HAD family members in *Arabidopsis*. The full amino acid sequences for the proteins were aligned in Clustal Omega (Sievers *et al*., (2011) *Mol Syst Biol* 7:539) and then used to generate a phylogenetic tree using PHYML (Guindon *et al.,* (2010), *Syst Biol* 59:307-321) within the Geneious Prime bioinformatics software platform. Default parameters were used for alignment and tree-building, and the tree included a bootstrapping of 100. The *Arabidopsis* HAD proteins included in the analysis were chosen based on the HAD phylogenetic tree of Du *et al*. (2021) PLoS ONE 16(1): e0245600, and included all *Arabidopsis* proteins from subgroup Ib (containing AtSGPP). The secondary identifiers of the *Arabidopsis* proteins (AtHADx) are the same as those used in Du *et al*. (2021).

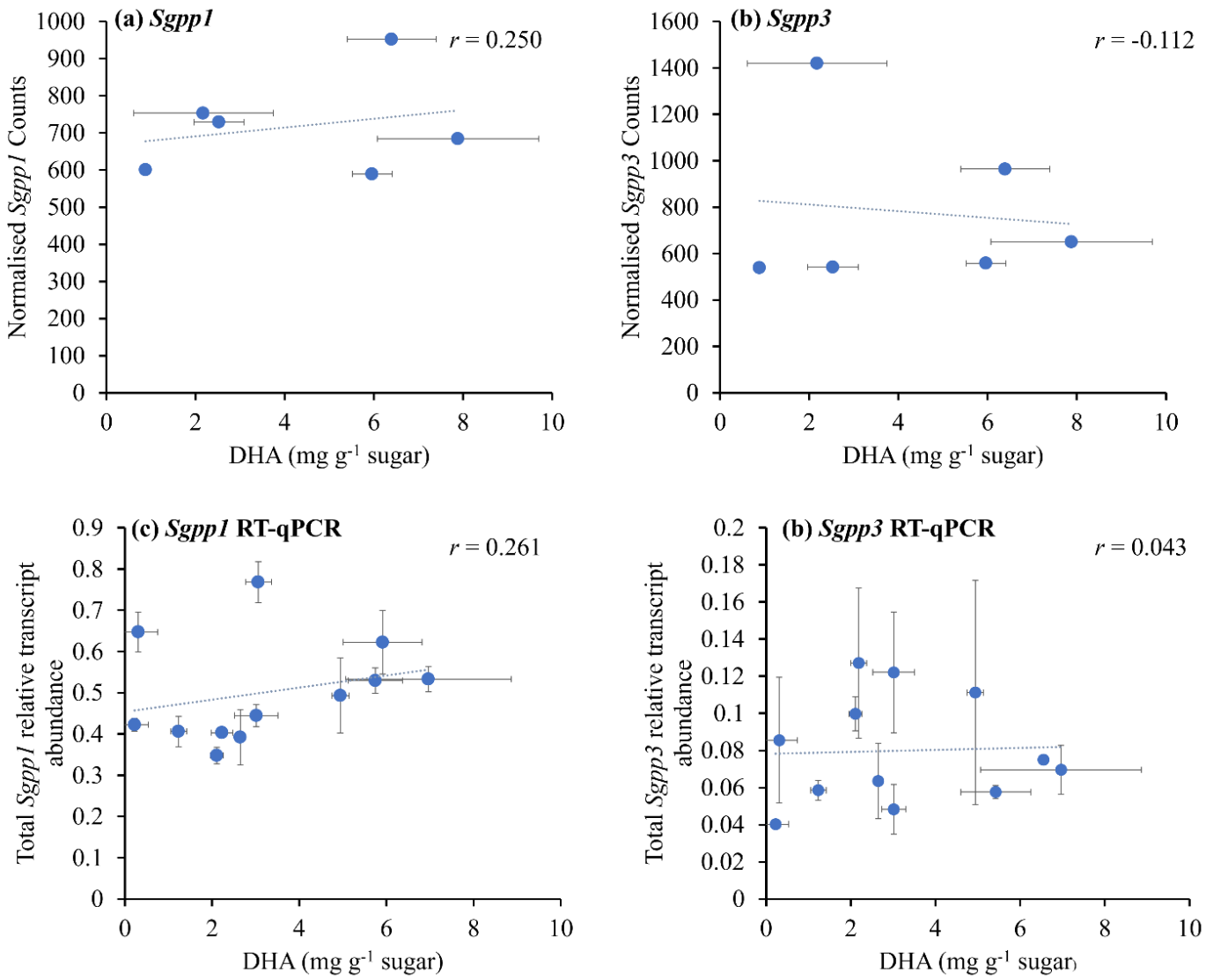


Figure S2.5: *Sgpp1* and *Sgpp3* expression in a range of *Leptospermum scoparium* genotypes. Normalised RNAseq counts of (a) *Sgpp1* and (b) *Sgpp3* from Low-DHA (N40A, N3ZR, and N40L) and High-DHA (N41V, N424 and N414) lines against DHA mg g^{-1} sugar in the corresponding samples at FDS 6-7. (c) Total *Sgpp1* expression and (d) Total *Sgpp3* expression (RT-qPCR) in 2020 with 10 genotypes (EC103 EC201 EC31 EC37 EC47 EC53 Electric red N40L N41V N428). Each genotype's nectar sample consists of combined nectar washes from ten flowers, each RNA sample consists of the hypanthium tissue of those same ten flowers. Error bars represent one standard deviation. Trendline fitted by linear regression, *r* values are the Pearson correlation coefficient.

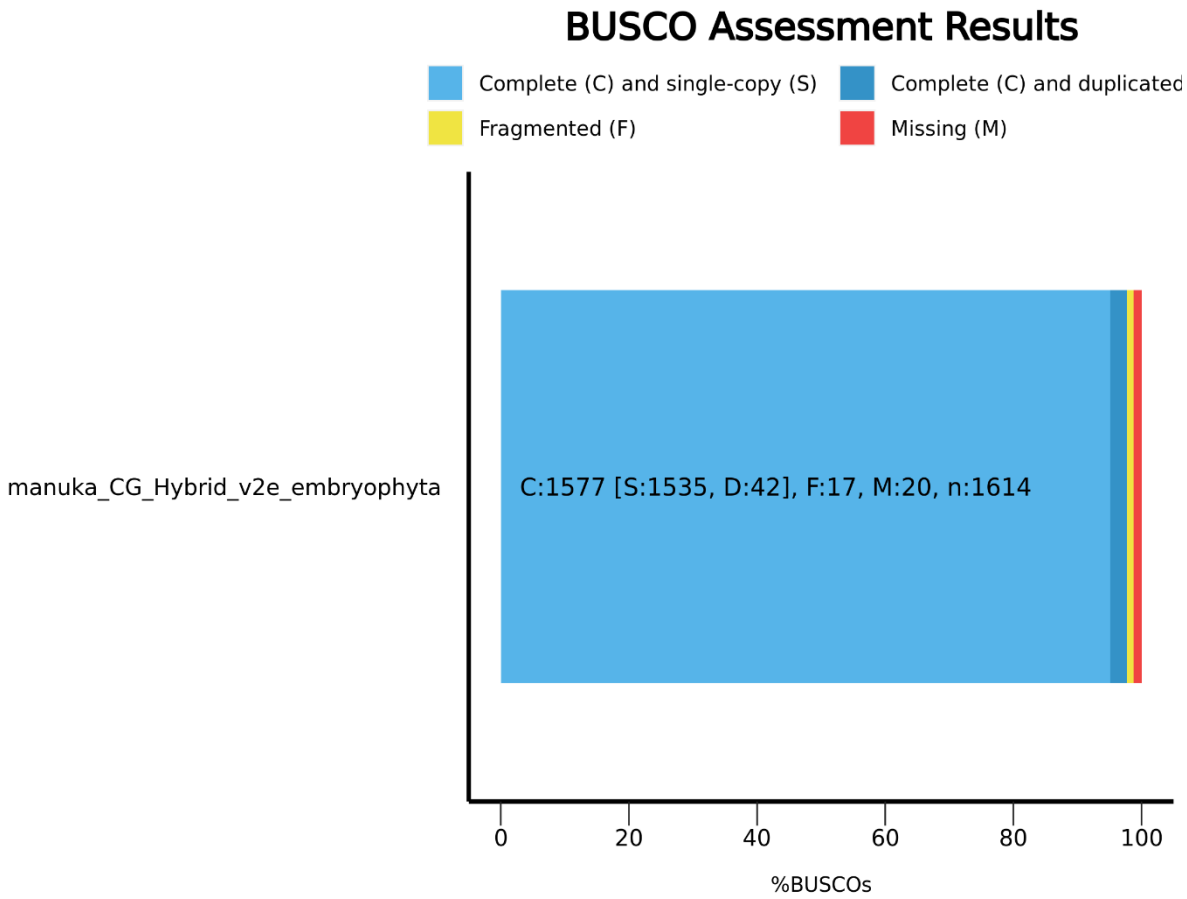


Figure S2.6: BUSCO completeness of the *Leptospermum scoparium* 'Crimson Glory' v2.0 genome. Lineage used is embryophyta_odb10.

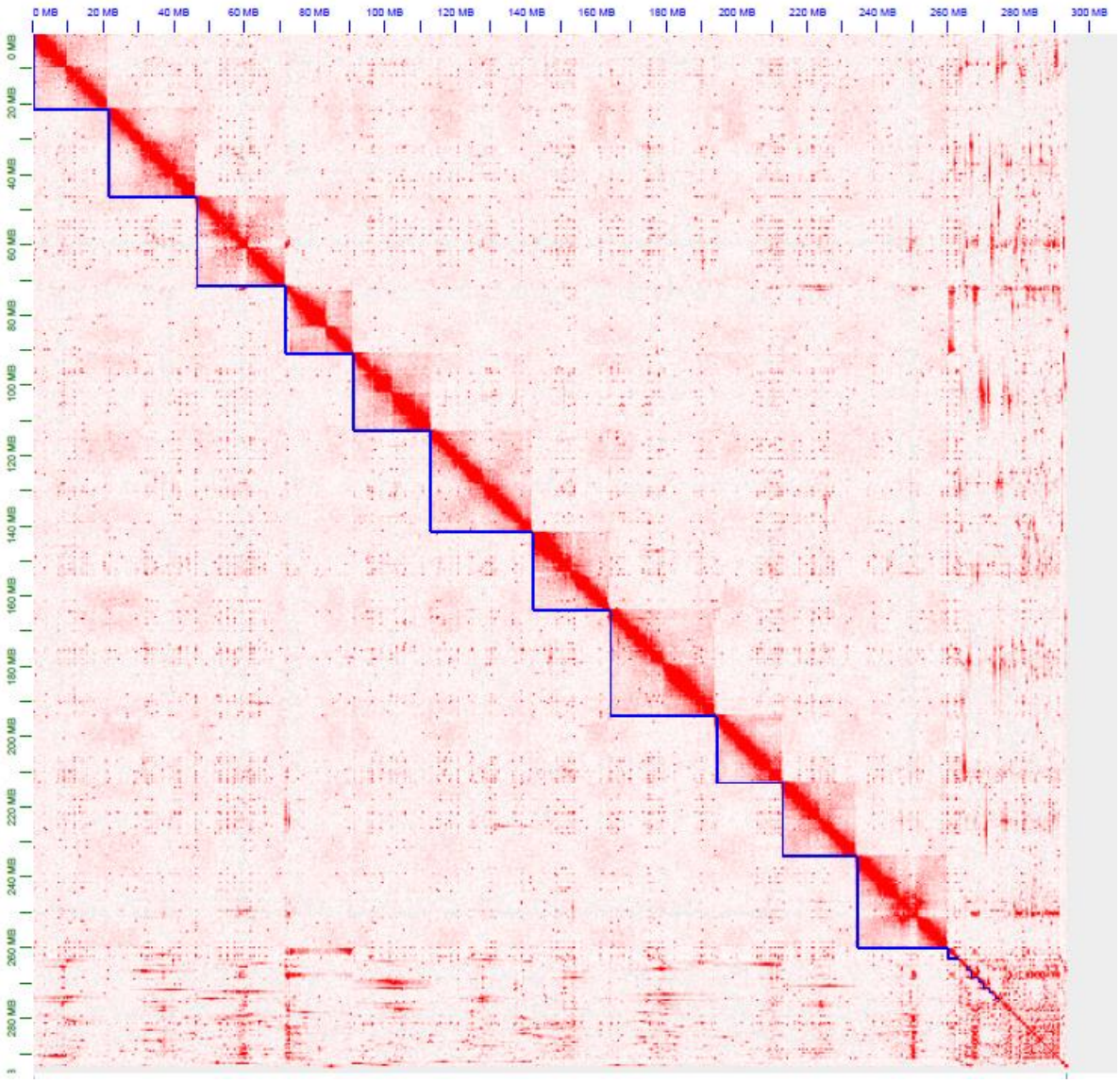


Figure S2.7: Hi-C contact map of the *Leptospermum scoparium* 'Crimson Glory' v2.0 genome.

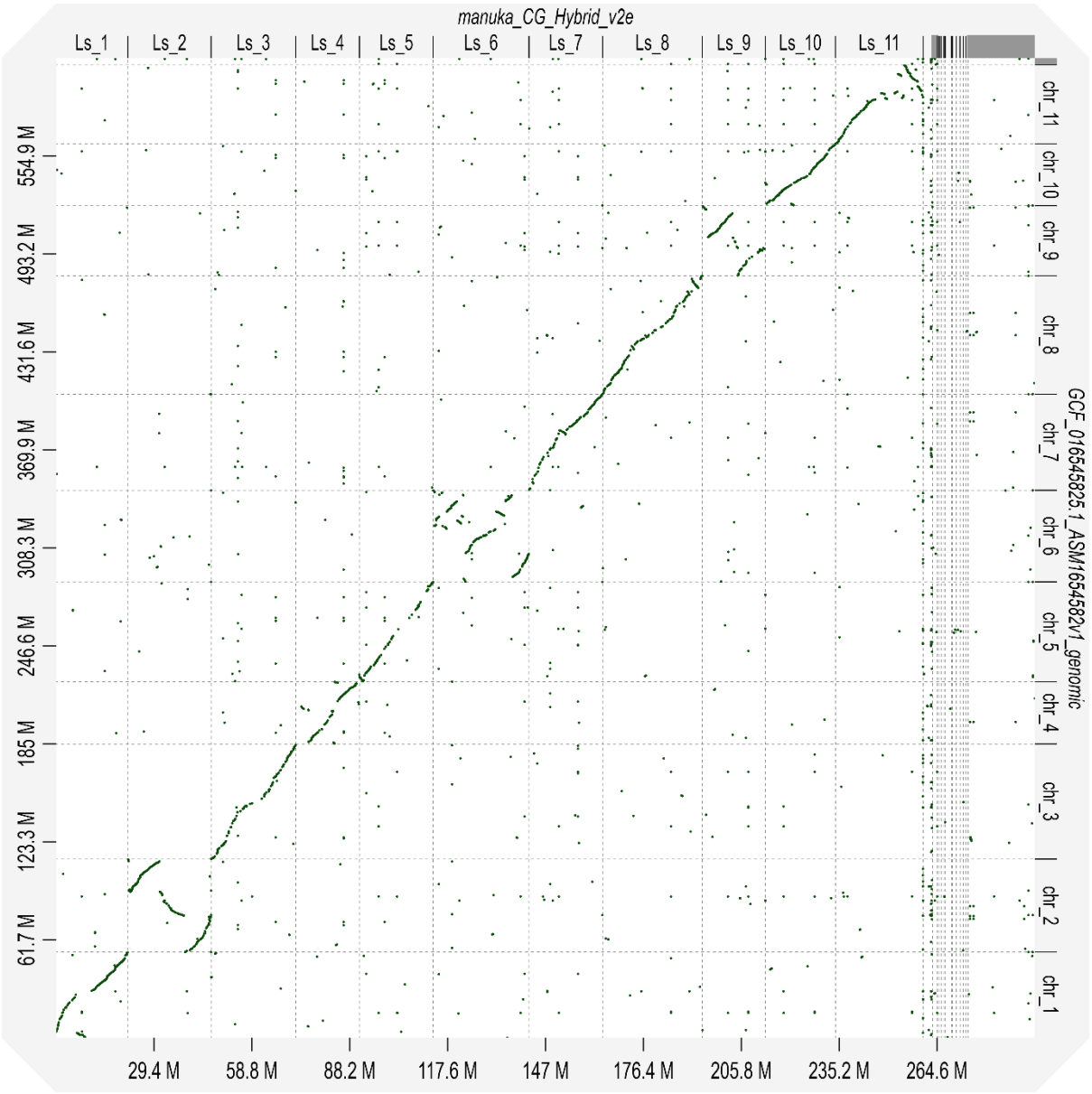


Figure S2.8: Conservation of synteny between the *Leptospermum scoparium* 'Crimson Glory' v2.0 genome and *Eucalyptus grandis*.

Table S2.1: *Leptospermum scoparium* lines used in each experiment.

Line	Common garden experiment	RNAseq	RT-qPCR	Parent of cross
EC53	Y		Y	
EC47	Y		Y	
EC31	Y		Y	
EC201	Y		Y	Y
EC103	Y		Y	Y
1N44Y	Y		Y	
1N44R	Y		Y	
1N430	Y		Y	
1N428	Y		Y	
1N424	Y	Y	Y	
1N41V	Y	Y	Y	
1N41S	Y		Y	
1N418	Y		Y	
1N414	Y	Y	Y	
1N40L	Y	Y	Y	
1N40A	Y	Y	Y	
1N3ZR	Y	Y		
1N3Z7	Y		Y	
1N3YY	Y		Y	
1N3Y7	Y		Y	
EC49			Y	
EC47			Y	
EC38			Y	
EC37			Y	
1N457			Y	
‘Electric Red’			Y	
‘Crimson Glory’		Y		
‘Martinii’			Y	

Table S2.2: Primer sequences and gene models used for RT-qPCR analysis.

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Gene model(s)
SGPPs	AGAGATAGGTTTCAATGGTGGA	GGAAGAGGAGCTCAGCTAT	LsSGPP1.t1, LsSGPP2.t1, LsSGPP3.t1
SGPP2	GGCAGGTAGACCGTAGAGTG	AGACACGACACATCTATCAGAATC	LsSGPP2.t1
Actin	CAGAAGGATGCCTATGTTGG	CCAGTTGCTGACAATACCAT	g20982.t2
Sand	TCAGGAGAGTCTGAAGGA GG	CGAAGCTACATACTG GTCCA	g7718.t1
rpl3	CTAGAGCTGGTCAGAATGGA	TCGGTAATTGCTGTGTGAGA	g745.t2

Table S2.3: Highly expressed genes in DESeq analysis of high and low DHA *Leptospermum scoparium* genotypes. In flowers secreting nectar *LsSgpp2* transcripts are more abundant in high DHA genotypes, and they are the most abundant transcripts in the nectary-enriched transcriptome. Shown are the top 10 genes with the highest transcript base means. Statistically significant values are indicated in bold text.

Gene ID	BaseMean	log2Fold Change	padj	Average L-DHA norm. count	Average H-DHA norm. count	*Arabidopsis* gene ID	TAIR description
LsSgpp2.t1	297,596	1.190	**6.87E-10**	180,556	423,512	AT2G38740.1	HAD-type phosphosugar phosphatase
g16350.t1	147,282	0.307	0.805	164,422	250,650	AT2G39060.1	Sucrose transporter expressed in nectaries and involved in nectar secretion
g9567.t1	136,074	-0.271	0.674	106,418	93,808	AT4G35160.1	Cytosolic *N*-acetylserotonin *O*-methyltransferase
g5295.t1	118,248	-0.310	0.865	120,791	114,174	NA	NA
g15578.t1	66,001	-0.0643	0.961	55,751	58,826	AT4G13940.1	*S*-adenosyl-L-homocysteine hydrolase
g12371.t1	63,223	-0.258	0.826	38,279	37,617	AT3G52600.1	Cell wall invertase expressed in flowers and ovary placental tissues
g5293.t1	63,005	-3.00	0.181	180,760	8,383	NA	NA
g22719.t1	62,024	0.316	0.762	39,147	57,696	AT1G02500.2	*S*-adenosylmethionine synthetase
g4943.t1	60,728	-0.003	0.999	47,396	49,443	AT3G43190.1	Protein with sucrose synthase activity
g22534.t1	49,002	-2.433	0.598	64,335	12,987	AT5G54160.1	Caffeic acid/5-hydroxyferulic acid *O*-methyltransferase

Table S2.4: Statistics of the *Leptospermum scoparium* 'Crimson Glory' v2.0 genome assembly.

Number of scaffolds	506
Total size of scaffolds	294,018,319
Longest scaffold	29,967,390
Shortest scaffold	1000
Number of scaffolds > 1K nt	490
Percentage of scaffolds > 1K nt	96.8
Number of scaffolds > 10K nt	471
Percentage of scaffolds > 10K nt	93.1
Number of scaffolds > 100K nt	61
Percentage of scaffolds > 100K nt	12.1
Number of scaffolds > 1M nt	15
Percentage of scaffolds > 1M nt	3.0
Number of scaffolds > 10M nt	11
Percentage of scaffolds > 10M nt	2.2
Mean scaffold size	581,064
Median scaffold size	25,249
N50 scaffold length	22,172,600
L50 scaffold count	6

Supplementary material for Chapter 3

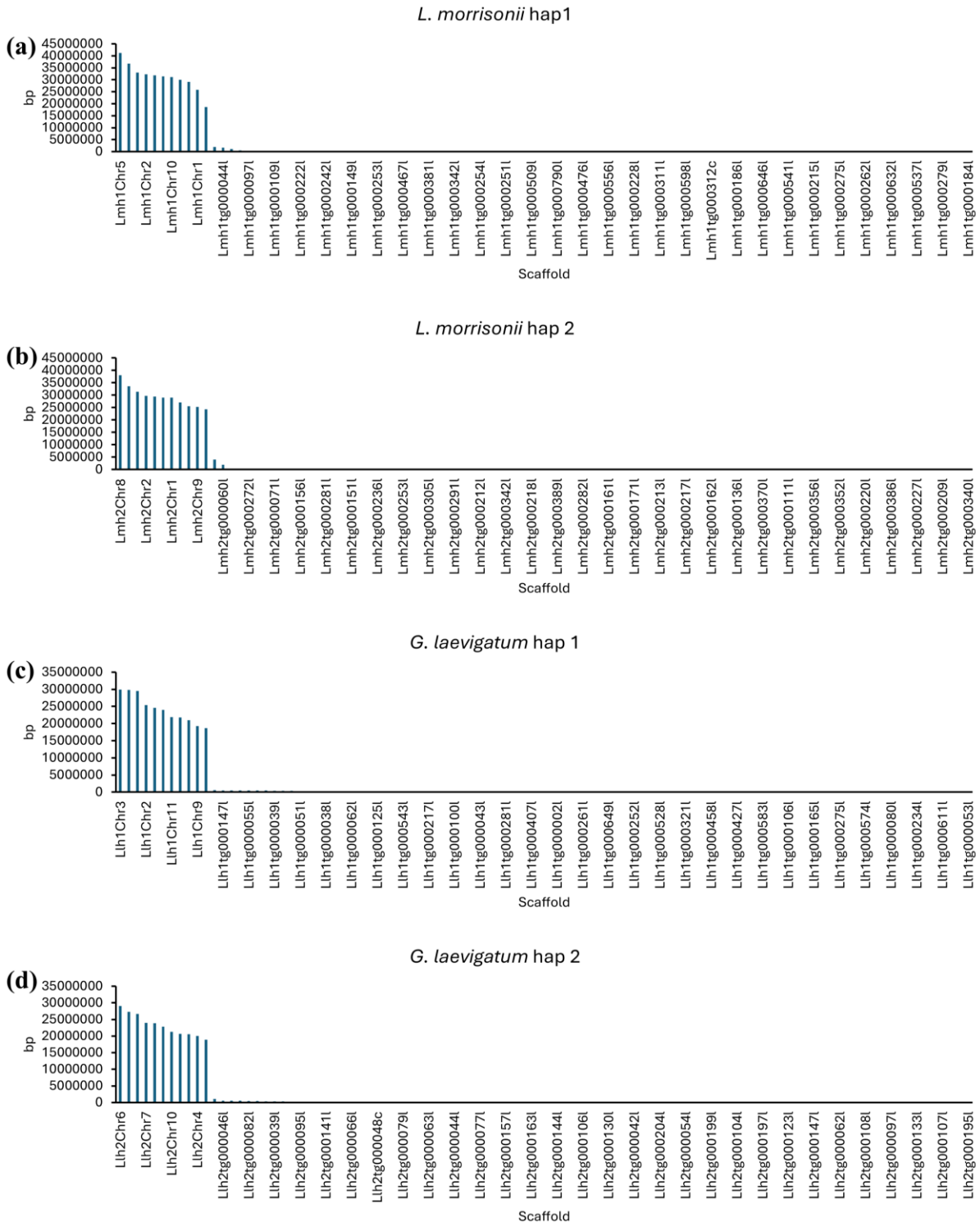


Figure S3.1: Assembled contigs statistics of the *Gaudium laevigatum* and *Leptospermum morrisonii* genomes. Length (bp) of top 100 longest scaffolds in each of (a) *L. morrisonii* hap 1, (b) *L. morrisonii* hap 2, (c) *G. laevigatum* hap 1 and (c) *G. laevigatum* hap 2.

Supplementary material for Chapter 4

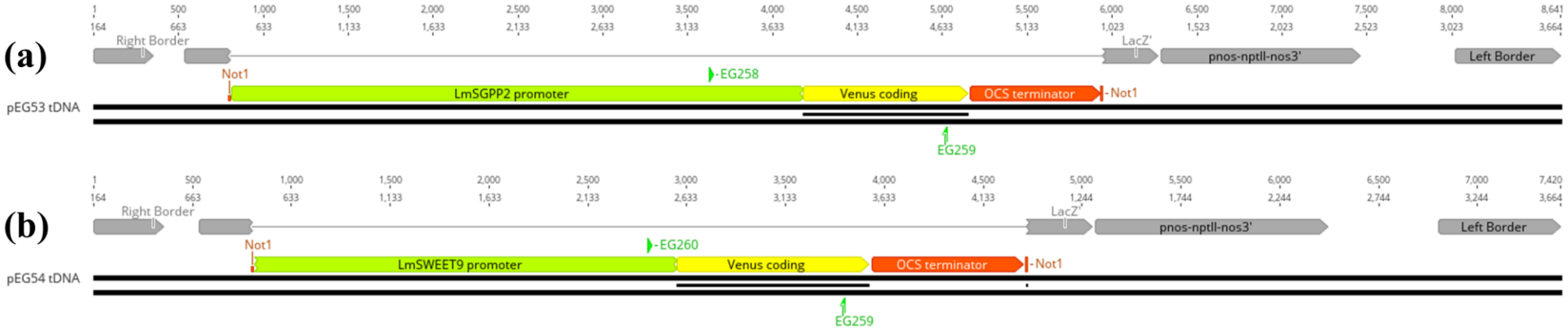


Figure S4.1: T-DNA inserts from (a) pEG53 (LmSGPP2promoter-Venus) and (b) pEG54 (LmSweet9promoter-Venus). Not1 restriction sites were used for cloning into pART27. Primers used for detecting transgene in transgenic plants were EG258 (GGCACAATAACAAAAGGTTTAGGA) and EG259 (AGTCCTTTCTTTTCAGATCCTTGA) for pEG53, and EG260 (ATCGATTCAAACTCAAGTTCTTCG) and EG259 for pEG54. Images created with Geneious version 2022.0 (Biomatters).

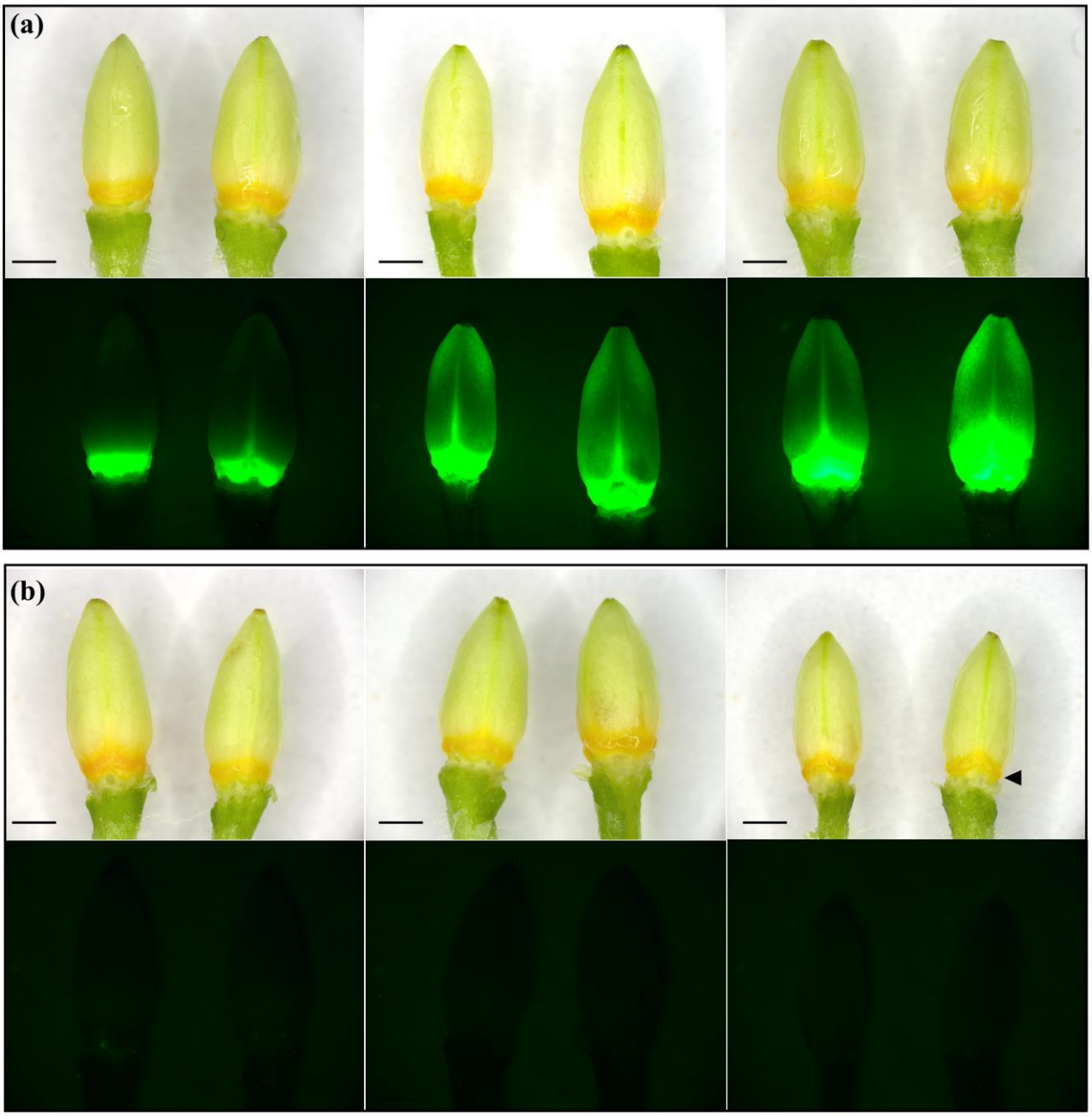


Figure S4.2: Ovary and nectary of overexpression lines of *Nicotiana benthamiana* under white and blue light (with YFP filter). Petals and sepals have been removed. (a) Three independent *LmSgpp2-Venus* transformed lines (T1R1-5, T1R3-2, T1R4-6). (b) Three independent *LmSweet9-Venus* transformed lines (T2R1-8, T2R5-3, T2R3-7) (Nectaries have minimal fluorescence in this treatment). Scale bar is 1mm. Triangle indicates nectary tissue.

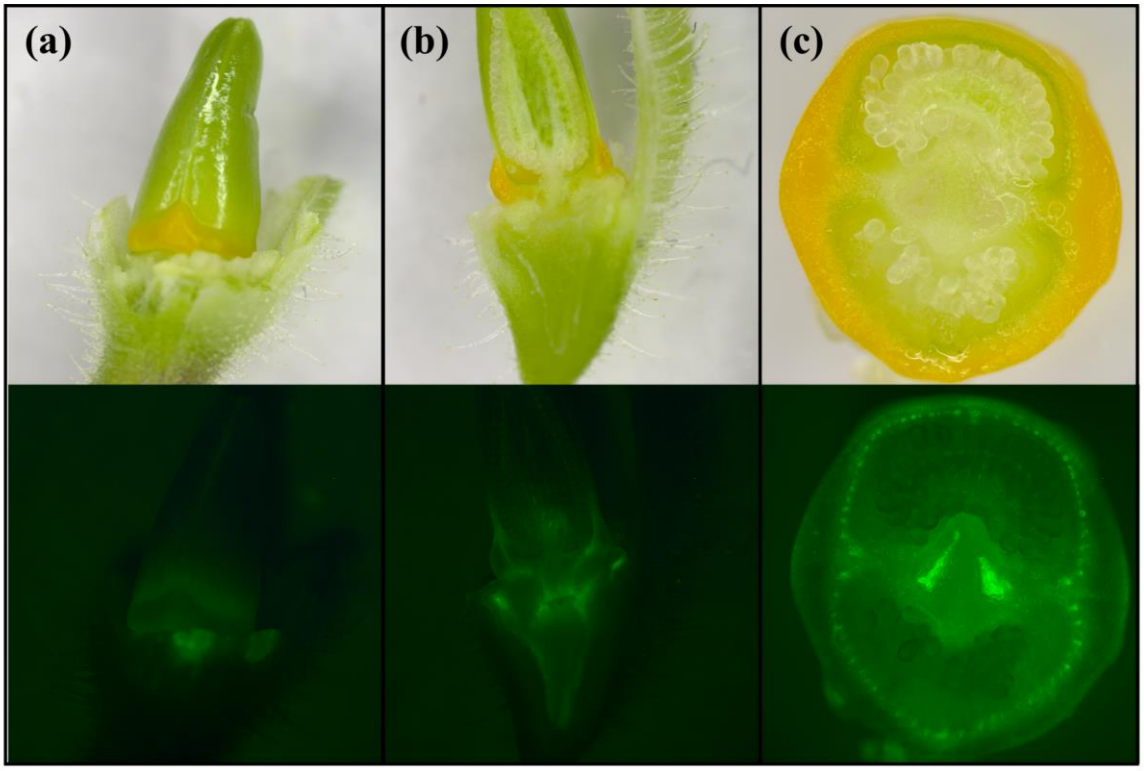


Figure S4.3: Nectaries of *LmSweet9-Venus* transformed *Petunia* under white and blue light (with YFP filter). (a) Nectary (b) longitudinal section of ovary nectary and receptacle (c) transverse cross section of nectary.

Table S4.1: Species and genotypes used in paper and what data has been generated for each, and that has been included in the present study.

Species	Genotype/ location	Average DHA mg/g sugar	Genome sequence	PCR Promoter sequence	PCR *Sgpp2* gene sequence	RNAseq	RT-qPCR
Leptospermum scoparium	cv 'Crimson Glory	3.7	Y	Y	Y		
Leptospermum morrisonii	PFR PN	4.2	Y			Y	Y
Gaudium laevigatum	Whanganui dunes	0.0	Y			Y	Y
Leptospermum scoparium	EC47	7.3		Y			
Leptospermum scoparium	EC31	1.2		Y	Y		
Leptospermum scoparium	EC201	4.3		Y	Y		
Leptospermum scoparium	EC103	2.6		Y	Y		
Leptospermum scoparium	EC53	5.9		Y	Y		
Leptospermum scoparium	1N424	6.3		Y	Y	Y	
Leptospermum scoparium	1N41V	4.6		Y	Y	Y	
Leptospermum scoparium	1N414	4.6				Y	
Leptospermum scoparium	1N41S	7.5		Y	Y		
Leptospermum scoparium	1N40L	2.3		Y	Y	Y	
Leptospermum scoparium	1N3ZR	0.7		Y	Y	Y	
Leptospermum scoparium	1N40A	0.8				Y	
Leptospermum scoparium	1N3Z7	0.8		Y	Y		
L. scoparium x L. rotundafolium	1N45G	0.7		Y	Y		
L. rotundafolium x L. macrocarpum	1N45R	0.0			Y		
L. scoparium x L. macrocarpum	'Electric Red'	0.2			Y		

Table S4.2: Media used for plant transformation.

Media	Number	Ingredients
Cocultivation	1828	1 × MS micro, 0.5 × MS macro, 50 µM Ferric sodium EDTA, 1 × B5 vitamins, 5 mg L^{-1} $CuSO_4$, 0.2 mg L^{-1} IAA, 3 mg L^{-1} BAP, 30 g L^{-1} sucrose, 7.5 g L^{-1} Davis Agar, 200 µM acetosyringone
Selective callus	2637	1 × MS micro, 0.5 × MS macro, 50 µM Ferric sodium EDTA, 1 × B5 vitamins, 5 mg L^{-1} $CuSO_4$, 0.2 mg L^{-1} IAA, 3 mg L^{-1} BAP, 500 mg L^{-1} cefotaxime, 30 g L^{-1} sucrose, 7.5 g L^{-1} Davis Agar. 200 mg L^{-1} kanamycin sulphate.
Regeneration callus	1861	1 × MS micro, 0.5 × MS macro, 50 µM Ferric sodium EDTA, 1 × B5 vitamins, 5 mg L^{-1} $CuSO_4$, 0.2 mg L^{-1} IAA, 3 mg L^{-1} BAP, 500 mg L^{-1} cefotaxime, 30 g L^{-1} sucrose, 7.5 g L^{-1} Davis Agar
Selective rooting	1609	1 × MS micro, 0.5 × MS macro, 50 µM Ferric sodium EDTA, 1 × B5 vitamins, 10 mg L^{-1} $CuSO_4$, 500 mg L^{-1} cefotaxime, 30 g L-1 sucrose, 7.5 g L^{-1} Davis Agar, 100 mg L^{-1} kanamycin sulphate
Regeneration rooting	1617	1 × MS micro, 0.5 × MS macro, 50 µM Ferric sodium EDTA, 1 × B5 vitamins, 10 mg L^{-1} $CuSO_4$, 500 mg L^{-1} cefotaxime, 30 g L-1 sucrose, 7.5 g L^{-1} Davis Agar,
Minimal	2216	1 × AB salts, 240 mg L^{-1} $NaH2PO_4$, 10 g L^{-1} glucose, 14.693 g L^{-1} morpholinoethane sulfonic acid. Media was supplemented with 200 µM acetosyringone after sterilisation.

Table S4.3: Genomes used for analysis of *Sgpps*. All were obtained from the NCBI database except for *Leptospermum scoparium*, *Leptospermum morrisonii* and *Gaudium laevigatum* (they were generated in the present study – Chapter 2) and *Syzigium maire* is from Genomics Aotearoa.

Species	Assembly	Number of Sgpps
Punica granatum	ASM765513v2	2
Melaleuca quinquenervia	NSW1078661	2
Syzigium maire	Ngā Hua o te Ia Whenua v1.0	2
Rhodamnia argenta	ASM2092103v1	2
Psidium guajava	guava_v11.23	2
Leptospermum morrisonii	V1 – Chapter 3	3
Leptospermum scoparium	V2 – Chapter 2	3
Gaudium laevigatum	V1 – Chapter 3	2
Corymbia calophylla	ANBG752993.2	1
Eucalyptus grandis	ASM1654582v1	3
Eucalyptus guilfoylei	CCA5924	3
Arabidopsis thaliana	TAIR10.1	1

Table S4.4: PCRs and primers used to sequence the *Sgpp* locus.

PCR	Primers
A1	EG229 EG223
A2	EG229 EG223
B9	EG232 EG248
B10	EG257 EG230
C5	EG231 EG237
PP2-1	EG220 EG221
PP2-3	EG220 EG213

Table S4.5: Primer sequences used to sequence the *Sgpp* locus.

Primer	Sequence
EG220	ATGACGGTTTCTTCCGCCGA
EG221	TCAAGCATTCTTTGCCTCAAGG
EG223	TCCGAGTCGCATAGAGTTCCAT
EG229	AGAAGCCATCAATGCTGAAACTG
EG230	AAATCTCCAACCGATCTAACGATG
EG231	CATCGTTAGATCGGTTGGAGATTT
EG232	ATGATTACAGAAGCAACTCGTTGT
EG237	TAGCATCGTTTCTGTTTCTCCTTA
EG248	GCTTAAGTTGACGTAAAGATGGGA
EG257	TTTTGGTGGCTTTTGACCAA

Supplementary material for Chapter 6

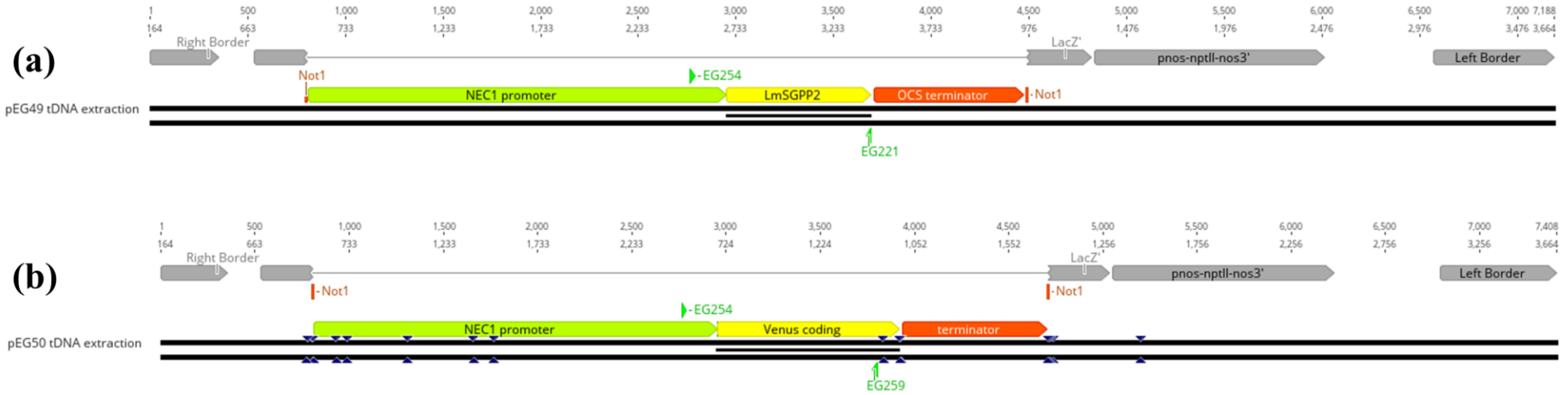


Figure S6.1: T-DNA inserts from (a) pEG49 (*Nec1-LmSgpp2*) and (b) pEG50 (*Nec1-Venus*). Not1 restriction sites were used for cloning into pART27. Primers used for detecting transgene in transgenic plants were EG254 (AAAGTCTGCATATTTCTTGCCAAA) and EG221 (TCAAGCATTCTTTGCCTCAAGG) for pEG48, and EG254 and EG259 (AGTCCTTTCTTTTCAGATCCTTGA) for pEG50. Images created with Geneious version 2022.0 (Biomatters).

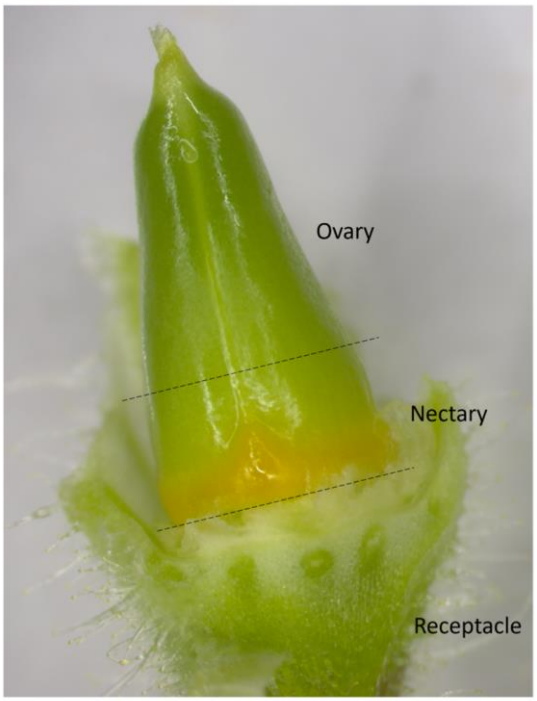


Figure S6.2: Example of nectary enriched *Petunia* tissue sampled for RNA extraction and qPCR. Transverse cuts indicated by dashed lines were made through the ovary either side of the nectary, and the nectary enriched tissue frozen for analysis.

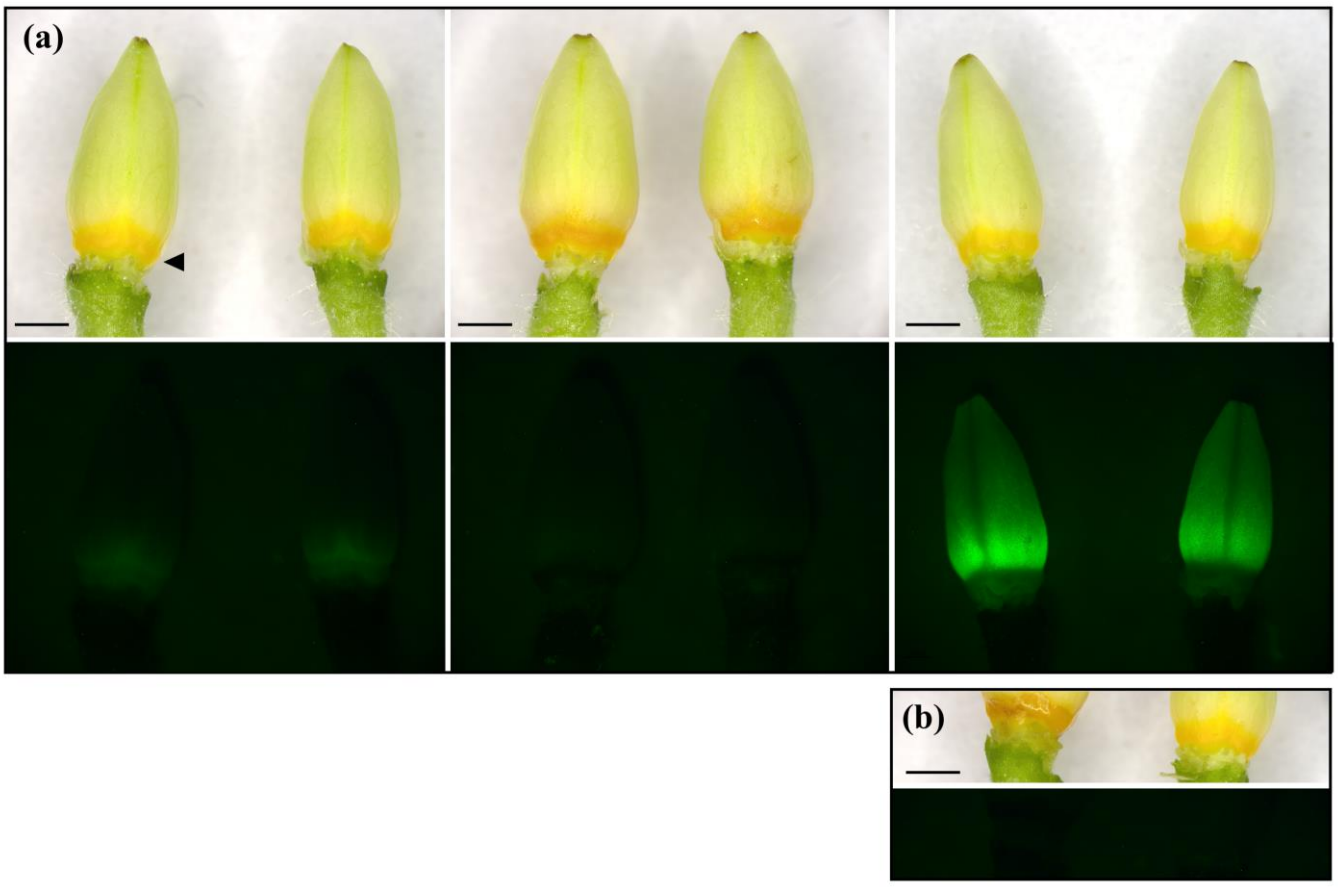


Figure S6.3: Ovary and nectary of overexpression lines of *Nicotiana benthamiana* under white and blue light (with YFP filter). Petals and sepals have been removed. (a) three independent Nec1-Venus transformed overexpression lines of *Nicotiana bethamiana* under white and green light (T2R2-8, T2R3-9, T2R5-6). (b) Nectaries of Nec1-LmSGPP2 transformed line (T1R6-9). Scale bar 1mm. Nectary tissue (orange) indicated with arrow in (a).

Table S6.1: *Petunia* lines that flowered during the experiment and whether they were transgenic and had observed fluorescence. Lines were verified for presence of transgene via PCR (for specific primers see Supplementary figure S6.1).

Treatment	Line	Transgenic	Fluorescence?
Nec1-LmSgpp2	T1R2-2	Y	No
	T1R3-1	Y	No
	T1R3-6	Y	No
	T1R4-1	Y	No
	T1R4-2	Y	No
	T1R4-4	Y	No
	T1R4-9	N	No
	T1R5-1	Y	No
	T1R5-4	Y	No
	T1R6-4	Y	No
	T1R6-6	Y	No
	T1R6-7	Y	No
	T1R1-2	Y	No
	T1R2-5	Y	No
	T1R3-2	Y	No
	T1R5-3	N	No
	T1R6-5	Y	No
Nec1-Venus	T2R1-2	Y	Yes - all
	T2R2-2	Y	Yes - all
	T2R2-4	Y	Yes - all
	T2R2-5	Y	Yes - some
	T2R3-1	Y	Yes - all
	T2R3-3	Y	Yes - all
	T2R4-1	Y	Yes - some
	T2R4-2	Y	Yes - all
	T2R4-4	Y	Yes - some
	T2R4-6	Y	Yes - all
	T2R5-1	Y	No
	T2R5-2	Y	Yes - all
	T2R3-2	Y	No
	T2R4-8	Y	No
	T2R6-1	Y	No
	T2R6-5	Y	Yes - all
	T2R6-10	Y	No

Table S6.2: Primers used for RT-qPCR on nectary enriched tissue of transgenic lines.

Gene	Primers	
Nec1	EG271	AGTGGGATGGATTTGTGCAG
	EG272	GGGCATGAACTCTACACTCTT
Sgpp2	EG269	TCAGTGAGGAAAAGGAAACCATAT
	EG270	CCGTGATCTTCTATCCACTTCTTC
Venus	EG263	GCACAAGCTGGAGTACAACT
	EG264	GATGTTGTGGCGGATCTTGA
Actin	PhACT2 F	CCTGATGAAGATCCTCACCGA
	PhACT2 R	CAAGAGCCACATAGGCAAGCT
EF1 alpha	NA160	TGTTCTCTGCCTTGTATGTCTGG
	NA162	TCAAAAGAGGCAGGCAGACAG

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website:
http://www.waikato.ac.nz/students/research-degre e

Co-Authorship Form

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Chapter 2 - A phosphatase gene is linked to nectar dihydroxyacetone accumulation in mānuka (*Leptospermum scoparium*) - New Phytologist (2024) 242: 2270–2284 doi: 10.1111/nph.19714

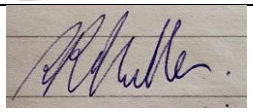
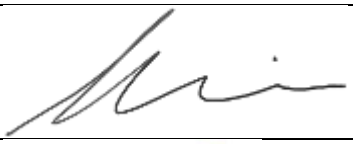
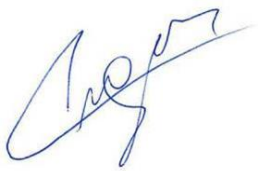
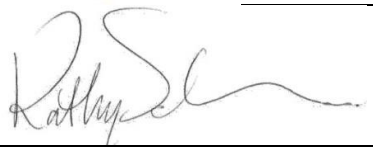
Nature of contribution by PhD candidate	Planning and designing the research, conducting the experiments, analysing and interpreting data. Led preparation of manuscript.
Extent of contribution by PhD candidate (%)	80%

7 CO-AUTHORS

Name	Nature of Contribution
Amali H. Thrimawithana	Performed experiments, analysed data, produced v2.0 of the 'Crimson Glory' genome
John W. van Klink	Performed experiments
David H. Lewis	Performed experiments
Ignacio Carvajal	Produced v2.0 of the 'Crimson Glory' genome
Jason Shiller	Produced v2.0 of the 'Crimson Glory' genome
Poppy Miller	Analysed data
Simon C. Deroles	Analysed data
Michael J. Clearwater	Interpreted data
Kevin M. Davies	Planned and designed the research, contributed to manuscript.
David Chagné	Conceived of project, planned and designed the research, analysed and interpreted data, produced v2.0 of the 'Crimson Glory' genome, contributed to manuscript
Kathy E. Schwinn	Conceived of project, planned and designed the research, analysed and interpreted data, contributed to manuscript

8 Certification by Co-Authors

The undersigned hereby certify that:
❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
Amali H. Thrimawithana		**19/08/25**
John W. van Klink		**18/08/25**
David H. Lewis		**18/08/2025**
Ignacio Carvajal		**18/08/25**
Jason Shiller		**18/08/25**
Poppy Miller		**18/08/25**
Simon C. Deroles		**20/08/25**
Michael J. Clearwater		**14/08/25**
Kevin M. Davies		**19/08/25**
David Chagné		**14/08/25**
Kathy E. Schwinn		**27/08/25**

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website:
http://www.waikato.ac.nz/students/research-degre e

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Chapter 3 - Chromosome-scale haplotyped genome assemblies of Australian tea trees *Leptospermum morrisonii* and *Gaudium laevigatum*

Nature of contribution by PhD candidate	Contributions to this paper were part of developing the idea and acquiring funding, contributions to the manuscript, and all field and lab-based work
Extent of contribution by PhD candidate (%)	40%

9 CO-AUTHORS

Name	Nature of Contribution
David Chagné	Developing idea and acquiring funding, preparation of manuscript
Cecilia Deng	Bioinformatic work assembling genomes and transcriptomes, contributions to manuscript

10 Certification by Co-Authors

The undersigned hereby certify that:
❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
❖ ~~that the candidate wrote all or the majority of the text.~~

Name	Signature	Date
David Chagné		**27.08.2025**
Cecilia Deng		**27.08.2025**

Contributions to Chapters 4-6

Chapter 4: Sarah Bailey assisted with minION sequencing analysis. John van Klink conducted the nectar analyses.

Chapter 5: Robert Simpson and Cyril Hamiaux advised on protein assays. Andrew McLachlan assisted with conducting statistical analyses.

Chapter 6: Rebecca Yorker assisted with tissue culture, nectar harvesting and analysis of transgenics. Catherine Sansom conducted the nectar analyses.