Australian Government
Rural Industries Research and Development Corporation

Genetic Tools for Improving Tea Tree Oil

RIRDC New ideas for rural Australia

Genetic tools for improving Tea Tree oils

by Andras Keszei, Hamish Webb, Carsten Kulheim and William Foley

October 2010

RIRDC Publication No 10/189
RIRDC Project No PRJ-002803

© 2010 Rural Industries Research and Development Corporation.
All rights reserved.

ISBN 978-1-74254-156-3
ISSN 1440-6845

Genetic Tools for Improving Tea Tree Oils
Publication No. 10/189
Project No. PRJ-002803

The information contained in this publication is intended for general use to assist public knowledge and discussion and to help improve the development of sustainable regions. You must not rely on any information contained in this publication without taking specialist advice relevant to your particular circumstances.

While reasonable care has been taken in preparing this publication to ensure that information is true and correct, the Commonwealth of Australia gives no assurance as to the accuracy of any information in this publication.

The Commonwealth of Australia, the Rural Industries Research and Development Corporation (RIRDC), the authors or contributors expressly disclaim, to the maximum extent permitted by law, all responsibility and liability to any person, arising directly or indirectly from any act or omission, or for any consequences of any such act or omission, made in reliance on the contents of this publication, whether or not caused by any negligence on the part of the Commonwealth of Australia, RIRDC, the authors or contributors.

The Commonwealth of Australia does not necessarily endorse the views in this publication.

This publication is copyright. Apart from any use as permitted under the *Copyright Act 1968*, all other rights are reserved. However, wide dissemination is encouraged. Requests and inquiries concerning reproduction and rights should be addressed to the RIRDC Publications Manager on phone 02 6271 4165.

Researcher Contact Details

Name: William J Foley
Address: Research School of Biology
Australian National University
Canberra 0200
Phone: 02 612525335
Fax: 02 61255573
Email: William.foley@anu.edu.au

In submitting this report, the researcher has agreed to RIRDC publishing this material in its edited form.

RIRDC Contact Details

Rural Industries Research and Development Corporation
Level 2, 15 National Circuit
BARTON ACT 2600

PO Box 4776
KINGSTON ACT 2604

Phone: 02 6271 4100
Fax: 02 6271 4199
Email: rirdc@rirdc.gov.au.
Web: http://www.rirdc.gov.au

Electronically published by RIRDC in October 2010
Print-on-demand by Union Offset Printing, Canberra at www.rirdc.gov.au
or phone 1300 634 313

Foreword

Plant essential oils were Australia's first export and remain a significant contributor to the economy particularly in rural and regional areas. Maintaining competitive advantages in the face of low labour costs elsewhere is a particular challenge and Australian producers need to have access to break-through technology that enables them to continue to maintain higher prices due a high quality product. This project has applied new advances in the understanding of the biosynthesis of terpenes, the major constituents of essential oils, in particular those from Tea Tree (*Melaleuca alternifolia*).

For the first time, this work has identified the genes that produce terpenes in *Melaleuca alternifolia* that provide a direct diagnostic test to ensure that only the commercially valuable chemotype is planted. This will enable the Tea Tree Industry to optimise oil profiles at a much-reduced cost for breeding programmes. Secondly the work has identified the genes and gene variants that are associated with higher yields of essential oils in different trees. This is a significant achievement and makes possible direct selection of high yielding plants without the need for extensive traditional breeding programmes. In addition, the work means that Australia can control the genetic resources for an important industry

This project was funded from industry revenue, which is matched, by funds provided by the Australian Government.

This report, an addition to RIRDC's diverse range of over 2,000 research publications, forms part of our Tea Tree Oil R&D program, which aims to support the development of sustainable and profitable production systems.

Most of RIRDC's publications are available for viewing, free downloading or purchasing online at www.rirdc.gov.au. Purchases can also be made by phoning 1300 634 313.

Craig Burns
Managing Director
Rural Industries Research and Development Corporation

Acknowledgments

Many people contributed to the success of this work. Ms Yasmin Hassan did the early work on identifying terpene synthase genes from *Melaleuca alternifolia*. Ms Anja Zeissler contributed significantly in isolating sabinene hydrate synthase and Mr Paolo Momigliano made an important contribution to understanding the expression of the different terpene synthase genes in the different chemotypes. Our collaboration with Professor Jonathan Gershenzon (Max Planck Institute of Chemical Ecology in Jena, Germany) and with Dr Jörg Degenhardt and Dr Tobias Köllner (Martin Luther University at Halle, Germany) enabled the terpene synthase genes to be characterized and their function verified. Their involvement was essential for this project.

In the second part of the study, we collected the samples in conjunction with Mr Gary Baker (NSW DPI Wollongbar) and Dr John Doran (CSIRO). We would like to thank Dr Alan Wade and Ms Samira Samtleben, for their efforts in collecting thee samples. Dr Gavin Moran's assistance with the association analysis was vital as was Dr Ian Wallis' help with numerous laboratory matters. Dr Rob Lanfear helped with the statistical analysis and writing the Python script that enabled us to separate barcodes when the commercial software proved unable to do so. Staff of both the Biomolecular Resource Facility at the ANU facilitated the 454 sequencing experiment and Professor Barry Pogson kindly provided access to Real-time PCR facilities.

Throughout this work we received strong support from both the tea tree industry and RIRDC staff. Their enthusiasm and interest in this work contributed significantly to the successful outcome.

Abbreviations

A	adenine
bp	base pairs
C	cytosine
cDNA	copy DNA constructed from RNA template
CYP	cytochrome P450
DMAPP	dimethylallyl diphosphate
EST	expressed sequence tag
FID	flame ionization detector
FPP	farnesyl diphosphate
G	guanine
GC-MS	gas chromatography coupled with mass spectrometry
GEP	gene expression programming
GPP	geranyl diphosphate
GGPP	geranylgeranyl diphosphate
HPLC	high-performance liquid chromatography
IPP	isopentyl diphosphate
MeJa	methyl jasmonate
MEP	2-C-methylerythritol-4-phosphate
ML	Maximum Likelihood
mRNA	Messenger RNA (mRNA is transcribed from a DNA template, and carries coding information to the sites of protein synthesis)
m/z	mass-to-charge ratio
NIST	National Institute for Standards and Technology
ORF	open reading frame
PCA	principal components analysis
PCR	polymerase chain reaction
qRT-PCR	Quantitative real time RT-PCR
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SIM	single ion monitoring (in GCMS)
SNP -	single nucleotide polymorphism
SRM	selected reaction monitoring
T	thymine
TIF	translation initiation factor EIF-5A
TPS	terpene synthase
tRNA	transfer RNA
UTR	untranslated region (of mRNA)

Contents

List of Figures

List of Tables

Executive Summary

What the report is about

This report describes the genes and gene variants that control both the yield and the chemical profile of commercially valuable essential oils in Australian tea tree *Melaleuca alternifolia.* It opens the door for molecular breeding, which will reduce reliance on long-term and costly breeding experiments and improve yield and profitability of the Australian industry.

Who is the report targeted at?

The report is aimed primarily at breeders and individual producers in the tea tree industry. It emphasises to the industry that its profitability is dependent on small random changes in a few genes, and so conserving the genetic variability in the species is vital.

The report also demonstrates to the agricultural R & D sector that molecular breeding is achievable using the high level skills of Australian science, and that recent technological advances make possible extensive progress on a modest budget.

Finally, the report is targeted at agricultural policymakers in Australia who are generally poorly informed about the opportunities provided by the revolution in genomics. Significant opportunities for Australia are being surrendered to other countries and this report shows how benefits of plant genomics can accrue rapidly even to small industries.

Background

In recent years there has been a revolution in understanding the biosynthesis of terpenes, which are the major components of essential oils from *Eucalyptus* and from tea tree (*Melaleuca alternifolia*). These breakthroughs have come from fundamental studies of the molecular genetics of terpenes and have involved identification of a number of genes of a family known as terpene synthases. These genes catalyse the last committed step in the biosynthesis of most of the essential oils and are largely responsible for the diversity of terpenes that are of interest to industry. No other industry in Australia is so reliant on a handful of genes for the production of their major products.

Aims/objectives

This project aims to improve the breeding of *M. alternifolia* to achieve better essential oil profiles and better yields of oil. To do so it will develop new knowledge on the genes that influence the yield and profile of terpenes. Specifically it aims to (i) validate a series of diagnostic tools based on gene variants associated with specific oil profiles (ii) develop diagnostic tests based on gene variants of four candidate genes (*ippi*, *dxs*, *dxr*, *gpps* as described above) associated with oil yields (iii) provide the opportunity for industry to protect IP associated with genes responsible for their major products.

Methods used

A combination of plant chemistry, molecular genetics and large-scale genomics including some highly innovative methods has been used in this work. These methods are applicable to improvement of many other non-traditional crops.

Results/key findings

Re-analysis of existing chemical data on the composition of *M. alternifolia* leaf oils identified previously undetected patterns. These patterns suggest that as few as three terpene synthase genes can make all the monoterpenes in the oil. This prediction was fulfilled and the three genes have been characterized. The gene that makes the commercially variable product arose from a chance gene

duplication event to an existing gene that made the low value product cineole. This duplicated gene experienced a small number of mutations, which led to the production of the valuable commercial, product. The profitability of the industry is thus based on very small number of genetic variants. These findings have been translated into diagnostic tools based on simple and cheap polymerase chain reactions so that the commercially valuable varieties can be identified at a very early stage.

We had predicted that quantitative variation in the yield of terpenes in *M. alternifolia* should be limited by the synthesis of substrates for the final terpene synthase enzymes. We found that this was indeed the case and showed that 50-60% of the variation in yield could be explained by variations in the expression of a set of enzymes that control the synthesis of substrates as predicted. We found that single nucleotide variations in these genes could in themselves explain significant amounts of variation. Therefore, these small genetic variants can be used directly to choose high yielding plants without the need for long-term breeding experiments. *M. alternifolia* is the only Australian crop for which the genetic control of factors affecting both qualitative profile and quantitative yield of the final products is so clear-cut.

Implications for relevant stakeholders for:

The long time involved in breeding crops from woody trees means that breeding programmes are expensive. In the case of tea tree, genetic factors are so overwhelming in influencing the profile and amount of the final product that the industry should continue to embrace molecular approaches to reduce the time and cost of breeding programmes. The diagnostic tools sought for assuring growers of the correct essential profile of their crop are clear and unambiguous. The diagnostic tools proposed for predicting oil yields are highly promising but need to be validated against more families.

Australian agricultural policy makers could be better informed about the potential for genomics in improving industries especially where Australia faces disadvantages in the cost of labour and production. The current project has shown that very significant advances in molecular breeding can be made on a modest budget and this project on tea tree has offered exceptional value for money.

Recommendations:

1. Diagnostic tools for verification of chemotypes can be made available to the Industry immediately.

2. Diagnostic tools identified for predicting yield should be validated and tested against a wider number of families.

3. Improvements in oil quality in particular the occurrence of limonene could be made by extending the approach described here.

4. The tea tree industry should ensure that it has a plan to conserve its genetic resources focussing on oil profile as well as yield.

5. The tea tree industry together with RIRDC and the researchers should give consideration to protection of the intellectual property arising from this work.

6. RIRDC as one of the few agencies that has shown the vision to engage with native plant genomics could play a significant role by sponsoring a workshop in which the basics of genomics and the benefits of the *Eucalyptus* genome are explained for other Australian government agencies and private industry**.**

Introduction

Over the past decade there has been a revolution in understanding the genes that control the production of essential oils. Most work has been conducted in the Lamiaceae especially in mint and has led to dramatic improvements in yield of oils and in other desirable traits (e.g. reduction in undesirable components and double-cutting). These new discoveries are being applied to a range of other crops including citrus.

This project aims to use these new technologies to improve the breeding of *Melaleuca alternifolia* to achieve better essential oil profiles and better yields of oil. To do so it will develop new knowledge on the genes that influence the yield and profile of terpenes in *M. alternifolia*.

Several tentative conclusions can be drawn from work on other species (i) Changes in the profile of essential oils are caused by variations in enzymes that complete the last step in terpene synthesis. These all belong to the gene family known as terpene synthases, and small sequence differences between these genes can be responsible for qualitatively different oils. (ii) Single terpene synthase genes can produce multiple essential oil components. Thus the variation of a single gene can affect many components of the oil. (iii) Changes in the yield of oils are due to the availability of precursors. This is determined by variation in a handful of genes that partition these precursors between different biochemical pathways.

The implications of these findings for the Australian tea tree oil industry are that (i) tea tree oil production is more reliant on the products of a handful of genes than any other Australian plant-based industry and (ii) the relationship between the final product and the known genes is much stronger and more direct for essential oils than for other plant products such as wheat, or wood. Therefore using a genetic approach to improving the resources available to the industry makes good sense.

In earlier work (ANU74A) we identified some of the genes that are responsible for the final step in synthesising the terpenes found in *M. alternifolia*. To be able to explain the complete oil profile, there are still further genes to discover, particularly those that make the major commercially valuable product, terpinen-4-ol. With these genes and an understanding of their expression patterns in different chemotypes of *M. alternifolia*, we can validate our diagnostic tools that will let us predict which chemotype will develop from early seedlings.

The second and more difficult issue is to identify genes and gene variants that influence the yield of oil from commercially valuable chemotypes. Based on work in other species we believe that we can identify the most likely candidates that control aspects of oil yield. These genes are deoxyxylulose synthase (DXS) and deoxyxylulose phosphate reducto-isomerase (DXR), isopentenyl diphosphate isomerase (IPPI), and geranyl pyrophosphate synthases (GPPS). All are involved in synthesising the universal substrates for all monoterpene production. We aim to clone these genes, and identify common variants that are associated with variations in oil yield. If this proves possible, a second series of diagnostic tests could be developed from these findings.

Background and review of terpene biosynthesis in plants

Variation in terpene yield and heritability of variation

Quantitative variation in terpene production in plants is common and the variation within species can often be large. In blue mallee (*Eucalyptus polybractea*), total foliar terpene concentration varies from 0.7% w/DM to 13% w/DW, almost a 20-fold difference (King *et al.* 2006). Quantitative variation in terpene traits also has important economic and ecological implications, e.g. in Lodge-pole pine (*Pinus contorta*), higher foliar concentrations of the monoterpene 3-carene confers increased resistance to

attack by the Douglas fir pitch moth (*Synanthedon novaroensis*) (Rocchini *et al.* 2000). There are many different factors that can produce quantitative variation in terpene concentration in plants such as environmental factors e.g. nutrient availability (Muzika 1993), water stress (Delfine *et al.* 2005), atmospheric CO_2 concentrations (Penuelas *et al.* 1997), seasonality and temperature (Peñuelas and Llusià 1997), or by induction via herbivory (Pare *et al.* 1999), which produces the plant hormone methyl jasmonate, which then induces the production of terpenes in some species (Martin *et al.* 2003) (but not eucalypts (Henery *et al.* 2008). However, in all species that have been examined (including both woody and herbaceous species), the genetic component of variation (expressed as the narrow sense heritability) has without exception been high (0.6-0.9).

High heritability has been found for specific terpenes e.g. foliar 1,8-cineole concentrations in *Eucalyptus kochii* and *Eucalyptus melliodora*, as well as total foliar terpene concentrations in *Eucalyptus camaldulensis*.

Terpene biosynthesis pathways

Terpene biosynthesis can be divided into three phases (Figure 1). Phase one involves the biosynthesis of isopentenyl pyrophosphate via two separate pathways. Phase two involves the conversion of isopentenyl pyrophosphate to dimethylallyl pyrophosphate and the conversion of these compounds to prenyl pyrophosphates, and phase three involves the conversion of prenyl pyrophosphates into terpenes catalysed by terpene synthase genes.

Phase one of terpene biosynthesis and its effects on yield.

The biosynthesis of isopentenyl pyrophosphate occurs via two spatially separated pathways in plants (Eisenreich *et al.* 1998). Both pathways synthesize isopentenyl pyrophosphate (IPP), the universal building block of all terpenes. The mevalonate (MVA) pathway obtains its precursor molecule acetyl-CoA from the Krebs cycle and is located in the cytosol (McGarvey *et al.* 1995). The deoxyxylulose phosphate pathway (MEP) gets its precursor molecule glyceraldehyde phosphate from the Calvin cycle and is located in the plastid (Rohmer 1999). As well as biosynthesising IPP, the MEP pathway also produces dimethylallyl pyrophosphate (DMAPP) at a 6:1 ratio (Rohdich *et al.* 2003). Recently, it has been shown that there is some transport of isopentenyl units across the plastid double membrane, suggesting that substrates are shared between the cytosol and the plastid although most of the precursors of each pathway are thought to remain in the compartment they were biosynthesized. The MVA pathway provides the substrate for the synthesis of sesqui- and triterpenes (McGarvey *et al.* 1995) whereas the MEP pathway provides substrate for the synthesis of monoterpenes, carotenoids, abscisic acid, gibberellic acid and diterpenes (Rohmer 1999).

During phase one, carbon from primary metabolism is directed to secondary metabolism and genes within this part of the pathway are thought to play an important role in determining terpene yield (Carretero-Paulet *et al.* 2002; Carretero-Paulet *et al.* 2006; Enfissi *et al.* 2005; Wildung *et al.* 2005). Work in peppermint (*Mentha piperita*) has shown that the early steps in this pathway may act as bottlenecks to terpene yield in plants. Over-expression of *dxr* and *dxs* the first two steps of the MEP

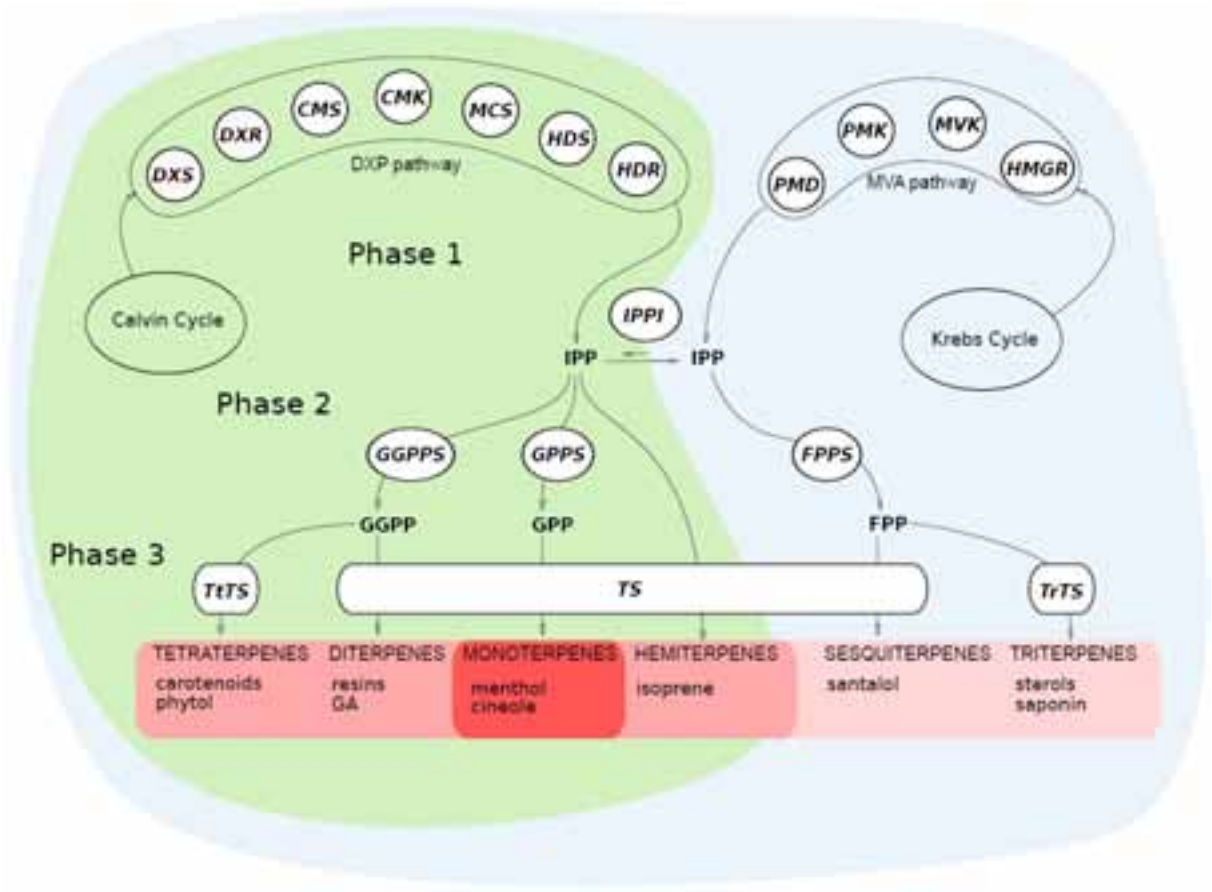

Figure 1: Simplified schematic of the terpene biosynthetic pathway in plants, showing the catalysing enzymes for each step. Phase 1: MEP pathway; the pathway is fed by D-glyceraldehyde-3-phosphate and pyruvate, DXS catalyses these into 1-deoxy-D-xylulose 5 phosphate, DXR catalyses this into 2C-methyl-D-erythritol 4 phosphate, CMS catalyses this into-diphosphcytidyl-2C-methyl-D-erythritol, CMK catalyses this into, 4-diphosphcytidyl-2C-methyl-D-erythritol 2-phophate, MCS catalyses 2C-methyl-D-erythritol 2,4-cyclodiphosphate, HDS catalyses this into 1-hydroxy-2-methyl-2-(E)-butenyl 4 phosphate which HDR catalyses into both isopentenyl diphosphate and dimethylallyl diphosphate at a 6:1 ratio (Rohdich et al. 2003). MVA pathway; acetyl-CoA is condensed to form 3-hydroxy-3-methylglutaryl-CoAm which is reduced to form mevalonic acid by HMGR, MVK converts this to 5-phosphomevalonate, which PMK converts to 5-pyrophosphomevalonate, which is converted to isopentenyl diphosphate (IPP) by PMD. Phase 2; IPPI isomerases IPP to DMAPP and vice versa, prenyl pyrophosphates convert IPP and DMAPP into prenyl pyrophospates (GPP, GPP and FPP). Phase 3: prenyl pyrophosphates are converted into terpenes by terpene synthases (McGarvey et al. 1995); Eisenreich et al. 1998; Rohmer 1999)

pathway increased foliar terpene concentrations by over 100% in transgenic peppermint (Wildung *et al.* 2005). Similar results were obtained by using transgenic *Arabidopsis* plants over-expressing *dxs* (Carretero-Paulet *et al.* 2002; Carretero-Paulet *et al.* 2006). Adding to this work, differential expression of both these genes is strongly correlated with differences in terpene yield between different cultivars of grape (Battilana *et al.* 2009), basil (Xie *et al.* 2008) and tomato (Enfissi *et al.* 2005). In the case of grape, there was also a significant QTL for yield that co-located with one of the three copies of the *dxs* gene in this plant (Battilana *et al.* 2009). This suggest that both *dxs* and *dxr* can act as bottlenecks to terpene production in plants and that both these genes may be involved in controlling variation in terpene production between individuals. The possible bottlenecks that may be expected to occur in the MVA pathway in plants are not well understood, for this study we have chosen to investigate the role of *mvk*, encoding a central enzyme of the MVA pathway. We wanted to investigate if it has any effect on yield and to study the transcript abundance of genes from both pathways and whether they are linked. From previous work found in the literature, we expect *dxr* and *dxs* to have a significant effect on terpene yield in *Melaleuca alternifolia*.

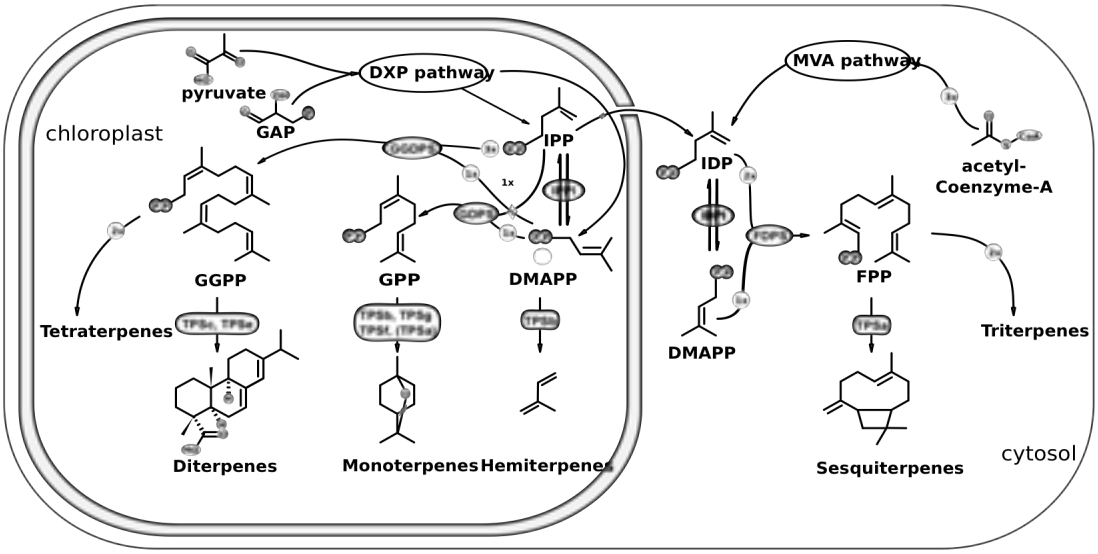

Figure 2: Schematic of the terpene biosynthesis pathways in plants with emphasis on the role of IPPI in converting IPP to DMAPP and the ratio at which each prenyl pyrophosphate is most active.

Phase two of terpene biosynthesis and its effect on yield and terpene composition:

Phase two of terpene biosynthesis begins with the enzyme isopentenyl pyrophosphate isomerase (IPPI), which catalyses the conversion of isopentenyl pyrophosphate (IPP) to its structural isomer dimethylallyl pyrophosphate (DMAPP) (Figure 2). Most plants have two copies of *ippi*, which are thought to have specialised but overlapping functions (Cunningham *et al.* 2000; Nakamura *et al.* 2001; Phillips *et al.* 2008). In *Arabidopsis,* one copy is expressed more strongly in the cytosol (termed *idi2*) and the other more strongly in the plastid (termed *idi1*). However, knockouts mutants of each gene in *Arabidopsis* have shown that each *idi* gene can compensate for the other in both parts of the cell and that there is no significant difference in the terpenes produced by either *idi1* or *idi2* knockouts when compared to the wild type (Phillips *et al.* 2008). IPP and DMAPP form the backbone of all plant terpenes. Prenyl pyrophosphate synthases catalyze the conversion of IPP and DMAPP to prenyl pyrophosphates that are the substrate for terpene synthases. The three prenyl pyrophosphates used in

terpene biosynthesis are geranyl pyrophosphate (GPP), synthesized by geranyl pyrophosphate synthase (*gpps*), that is used for monoterpene synthesis, geranylgeranyl pyrophosphate (GGPP), synthesized by geranylgeranyl pyrophosphate synthase (*ggpps*) that is used for di- and tetraterpene biosynthesis and farnesyl pyrophosphate (FPP), synthesized by farnesyl pyrophosphate synthase (*fpps*) that is used for sesqui- and triterpene synthesis (McGarvey *et al.* 1995). The ratio of IPP to DMAPP determines which of these prenyl synthases will be most active; *gpps* works most efficiently with a IPP: DMAPP ratio of 1:1 (Bouvier *et al.* 2000), *ggpps* works most efficiently with a ration of 3:1 (Allen *et al.* 1981; Ohnuma *et al.* 1989) and *fpps* works most efficiently with a ratio of 2:1 (Hugueney *et al.* 1990). The key intermediate gene *ippi* is responsible for controlling the IPP: DMAPP ratio and may have an important role in regulating the proportion of the different classes of terpene that are biosynthesised (Phillips *et al.* 2008). Wildung and Croteau (Wildung *et al.* 2005) showed that when both copies of *ippi* were down-regulated in peppermint, there was a large increase in the concentration of foliar sesquiterpenes. This shows that phase two of the terpene biosynthesis pathway is important in determining the relative proportion of each terpene class produced by plants. Wildung and Croteau (Wildung *et al.* 2005) also showed that *gpps* could act as a bottleneck to terpene production; when *gpps* was over-expressed in peppermint there was an increase in total foliar terpene concentration. This work suggests that firstly, *ippi* could be involved in the allocation of resources to the synthesis of mono- and sesquiterpenes in *M. alternifolia* and that second, the expression of *gpps* could be a significant influence on yield.

Phase three and terpene modification

Phase three of terpene biosynthesis is the conversion of the prenyl pyrophosphates into terpenes by a large family of enzymes know as terpene synthases (TPS), which determine the specific terpenes produced by the plant. Over 100 *tps* genes have been characterized to date and they are classified based on their primary substrate (GPP, GGPP or FPP). Some terpene synthase enzymes are capable of producing multiple products from a single precursor e.g. TPS1, a sesquiterpene synthase in *Zea mays* (corn) produces a mix of (*E*)-farnesene, (*E*,*E*)-farnesol, and (3*R*)-(*E*)-nerolidol when incubated with FPP and also linalool and geraniol when incubated with GPP although far less efficiently (Schnee *et al.* 2002). The ability of TPS enzymes to produce multiple products is one of the reasons there is so much diversity in terpene structures across the plant kingdom. Whereas many plant terpenes are direct products of TPS enzymes, e.g. 1,8-cineole (Chen *et al.* 2004) other terpenes are modified from the primary products of terpene synthases. The most common class of enzymes responsible for modifying terpenes are cytochrome P450 enzymes (Keeling *et al.* 2006) e.g. in peppermint six further steps catalysed by P450 enzymes are required for the biosynthesis of menthol from the monoterpene limonene (Croteau *et al.* 2005), adding further to the diversity of terpenes seen in plants. Phase three reactions should have only a small effect on total terpene yield but might play a greater role if substrates are limiting.

Regulation of flux through the terpene biosynthesis pathways

Regulation of flux through the terpene synthesis pathways is likely to be complex. A number of genes in the MEP pathway are present in multiple copies. So far, it has been demonstrated that *dxs* (Walter *et al.* 2002), *cmk* (Kim *et al.* 2008a), *hdr* (Kim *et al.* 2008b) and *ippi* (Blanc *et al.* 1995) can occur as multiple copy genes. Phillips *et al.* (2007) showed in Norway spruce (*Picea abies*), that when oleoresin production is induced via wounding and fungal treatment, the expression of *dxs2a*, *dxs2b*, *dxr* and *hdr* were all up-regulated but *dxs1* was not affected. When cell cultures were treated with methyl salicylate, chitosan and a beetle-associated fungal pathogen *Ceratocystis polonica*, *dxs2a* was up-regulated, but it was not up-regulated upon methyl jasmonate treatment. In contrast, d*xs2b* was up-regulated by methyl jasmonate and chitosan treatments but not the other treatments. This suggests that multi-copy genes in the terpene synthase pathways have distinct functions. It also shows that the expression genes in the MEP pathway are co-regulated. What regulates the expression of genes along the pathway is still poorly understood, but retrograde chloroplast to nuclear signalling is thought be

involved (Guevara-Garcia *et al.* 2005; Pogson *et al.* 2008). From the work outlined above I expect the expression of the genes in the MEP pathway to be correlated in *M. alternifolia*.

Thus far, it has not been investigated whether non-synonymous polymorphisms change the catalytic activity level of the enzymes in the terpene biosynthesis pathways. Functional analysis and fine-scale mapping of even apparently simple quantitative traits have turned out to be surprisingly complex. For example in *Drosophila melanogaster*, alcohol dehydrogenase (ADH), a classic fast-slow amino acid polymorphism, *Fast* homozygotes have a ~2.5 fold higher ADH activity level than *Slow* homozygotes. Some of this can be explained by catalytic differences between the *Fast* and *Slow* alleles caused by the amino acid change, but this does not explain the 1.5 fold difference in ADH protein levels between the alleles (Choudhary *et al.* 1991). The difference in protein levels is controlled by several polymorphisms in the first intron and 3' untranslated region (Stam *et al.* 1996). We believe that the possibility of enzymatic differences also playing a role in quantitative variation in terpene traits warrants investigation.

Most continuous traits, such as terpene yield in *M. alternifolia*, show substantial heritabilities (Houle 1992; Roff 1997). There is a lot of debate about the possible evolutionary forces that maintain variation in quantitative traits within wild populations. These include (i) the relative importance of mutation and balancing selection in maintaining variation in quantitative traits and (ii) whether regulatory or structural changes are most important in maintaining quantitative variation (Barton *et al.* 2002). With the advent of new sequencing technologies and the abundance of genetic markers available, identifying the genes and allelic variants of those genes is now a possibility, not just in model species but also in non-model species such as *M. alternifolia*. An approach that relies on a broader range of species, can contribute to a basic understanding of natural selection as well as improving our understanding of the evolution of variability within populations (Barton *et al.* 2002).

Variation in terpene yield in *Melaleuca alternifolia*

Most of the information about the molecular variation in terpene biosynthesis has come from studies of domesticated crop species, mutants or transgenic plants including those in *Arabidopsis* and peppermint described above. The ability to investigate the factors that control variation between individuals within wild populations opens the possibility of explicitly linking genotype, phenotype and ecology of quantitative terpene traits in the future as well as adding to the literature on standing quantitative variation within populations and how that variation is maintained.

In addition to this qualitative variation there is also significant quantitative variation in foliar terpene concentrations within the species, with oil yields varying from between 10 $mg \cdot g^{-1}$ DM to over 140 $mg \cdot g^{-1}$ DM (Hassan 2007). Although the oil profile is dominated by monoterpenes there is also variation in the foliar sesquiterpene concentrations, which can make up to 15% of total terpenes (Hassan 2007) and which is independent of the monoterpene chemotypes (Hassan 2007). This suggests that the biosynthesis of monoterpenes and sesquiterpenes is controlled independently.

Consistent with studies in other terpene-rich species, several lines of evidence argue that genetic factors are more important than environmental factors in determining oil yield in *M. alternifolia*. For example, the narrow sense heritability of total oil concentration was estimated as 0.8 (Doran *et al.* 2006) and a breeding programme achieved an increase in foliar terpene concentration of between 32 and 54% in just three generations. Similarly, differences in oil yield between seed lots were consistent between years and between sites (Doran *et al.* 2002). This suggests that the differences in terpene yield between individuals are under strong genetic control and that environmental effects on yield are secondary to genetic effects.

Structure of this report

In this report we consider both qualitative variation in the terpene profile of *M. alternifolia* as well as quantitative variation in yield. We start with a re-investigation of the chemical composition of terpenes in the belief that there is much about their biosynthesis that has been missed in earlier studies. Armed with predictions from that re-analysis, we then discover and characterize the genes that are responsible for the underlying patterns and show how they are controlled and from that data derive diagnostic tools that can be used to distinguish between different chemotypes. We then turn to the difficult issue of quantitative variation by first examining correlations between yields of terpenes and the expression of some key intermediate genes. We then examine the variation in these genes and identify gene variants that can be used diagnostically to identify high yielding plants without the need for long and expensive breeding experiments.

Objectives

This project aims to improve the breeding of *M. alternifolia* to achieve better essential oil profiles and better yields of oil. To do so it will develop new knowledge on the genes that influence the yield and profile of terpenes. Specifically it aims to (i) validate a series of diagnostic tools based on gene variants associated with specific oil profiles (ii) develop diagnostic tests based on gene variants of four candidate genes (*ippi*, *dxs*, *dxr*, *gpps* as described above) associated with oil yields (iii) provide the opportunity for industry to protect IP associated with genes responsible for their major products.

Chapter 1 - A biosynthetic interpretation of terpene chemotypes in *Melaleuca alternifolia*

Introduction

Identification of the genes that underlie phenotypic traits provides a crucial link to understanding evolutionary processes, such as adaptive evolution and genetic drift, since genetic variation forms the basis for selection and evolution. Understanding how genetic information is related to phenotypic variation of ecologically important traits presents a major challenge in evolutionary biology. It first requires the identification of those traits, their mode of expression, and finally, information on the key genes involved. Since it is hard to meet all these criteria, the relationship between genes, phenotypes and ecological function has only been achieved in a few cases in natural populations.

Distinct intra-specific variation in plant chemistry can occur in natural populations (Butcher *et al.* 1992) (Vernet *et al.* 1986), and can influence multi-species interactions (Linhart *et al.* 2005; Linhart *et al.* 1999). Such discontinuous chemical variations in a species are known as "chemotypes" and chemotypic variation in foliar terpenes is widespread in many Australian plants of the family Myrtaceae (Keszei *et al.* 2008). *Melaleuca alternifolia* is a Myrtaceous tree that shows chemotypic variation of foliar terpenes throughout its range (Butcher *et al.* 1994). The chemotype rich in terpinen-4-ol is widely used medicinally (Cox *et al.* 2001; Hammer *et al.* 2006).

Melaleuca alternifolia contains foliar oils that are a complex mixture of mono- and sesquiterpenes (Brophy *et al.* 1989). Terpene synthases are responsible for the direct production of the majority of the terpenes found in essential oils. They are capable of producing single (Novak *et al.* 2000) or multiple (Wise *et al.* 1998) compounds from the same prenyl-diphosphate substrate. Examining correlations between groups of terpenes and working under the assumption that strong positive correlation indicates a common biogenetic origin, (Zavarin (1970) predicted the presence of a single terpene synthase in *Picea* which catalyses the formation of sabinene, terpinolene and γ-terpinene in a set ratio. A gene was isolated later encoding a terpene synthase that showed the predicted catalytic profile (Gijzen *et al.* 1991). Qualitative variation in plant terpenes is likely due to variation in terpene synthase genes and may be deduced from identifying the correlations between compounds that potentially have the same biosynthetic origin.

In *M. alternifolia*, previous work has identified chemotypes based on variation in the concentration of 1,8-cineole, terpinen-4-ol and terpinolene. *M. alternifolia* is restricted to coastal flood plains of North-East New South Wales with neighbouring populations in the 'granite belt' in southeast Queensland. The geographic distribution of the three major monoterpenes reveals terpinen-4-ol in the majority in populations in the centre of the natural range, while terpinolene dominates in the northwest, and 1,8-cineole in the south. (Butcher *et al.* 1994) have shown that this corresponds to co-occurrence with the sister species *M. linariifolia* in the South and *M. trichostachya* in the North. *M. linariifolia* is also characterised by having both terpinen-4-ol and 1,8-cineole rich chemotypes, while *M. trichostachya* has terpinolene and 1,8-cineole rich chemotypes. It has been proposed that the respective *M. alternifolia* chemotypes may be a result of introgression from these sister species, however, convergent evolution of chemical traits due to similar evolutionary pressures cannot be ruled out.

The purpose of this part of the project is to identify correlations and chemical patterns that allow us to predict the biosynthetic origin of chemotypic variation in *M. alternifolia* terpenes. This in turn will facilitate a genomic search for the molecular components responsible for this variation.

Materials and Methods

Population Sampling

Mature leaf (ca. 20 g wet mass) was collected from 20 mature trees at eight sites and fifteen from a ninth site across the species' natural geographic range (Table 1), however, populations from the far south of the range (Port Macquarie) were excluded, because they have been proposed to be hybrids with southern lineages of *M. linariifolia* (Butcher *et al.* 1995). We chose trees that were at least 100m apart to avoid collecting from related trees (Rossetto *et al.* 1999) and the location of each tree was recorded. Samples were refrigerated at 4°C within 2 h of collection. The extraction of leaf oils was completed on the day of collection in order to minimise the possibility of evaporation and oxidative degradation.

Population	**Chemotype**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Flaggy Creek	-	-	-	2	18	-	-
Wooli Road	1	-	-	-	19	-	-
Chaffin Swamp	1	1	-	3	7	8	-
Dilkoon Creek	12	-	-	7	1	-	-
Devil's Pulpit	20	-	-	-	-	-	-
Casino Racecourse	15	-	-	-	-	-	-
Yellow Creek	19	-	-	1	-	-	-
Cannon Creek	-	14	5	-	-	1	-
Bald Rock Creek	-	14	2	-	1	2	1

Table 1. Frequency of chemotypes sampled at each site across the natural geographic range of *Melaleuca alternifolia*.

Extraction and Analysis of Terpenes

Mature leaf was separated from stems, weighed and placed in 10 ml of ethanol (Baker *et al.* 2000) containing 0.25 $g \cdot l^{-1}$ of *n*-tridecane as an internal standard. The samples were left in ethanol for seven days to complete extraction. A further 4 g of leaf from each tree was oven-dried at 60°C and weighed to calculate oil concentration on a dry weight basis.

Gas chromatography was carried out on an Agilent 6890 GC using an Alltech AT-35 (35% phenyl, 65% dimethylpolyoxylane) column (Alltech, Wilmington, DE). The column was 60 m long with an internal diameter of 0.25 mm with a stationary phase film thickness of 0.25 μm. Helium was used as a carrier gas. The ethanol extract was filtered through a 0.45 μm filter, and 1 μl was injected at 250°C at a 1:25 split ratio. The temperature program was as follows: 100°C for 5 min, ramping to 200°C at $20°C \cdot min^{-1}$ followed by a ramp to 250°C at $5°C \cdot min^{-1}$, and held at 250°C for 4 min. The total elution

time was 25 minutes. The components of the solvent extract were detected using an FID and an Agilent 5973 Mass Spectrometer dual setup through an SGE MS/FID splitter. Peaks were identified by comparisons of mass spectra to reference spectra in the National Institute of Standards and Technology library (Agilent Technologies, Deerfield, IL). Three individuals were reanalysed from each population on the final day of chromatography to confirm that retention times remained comparable. Peaks were quantitated based on the FID trace in MSD Chemstation Data Analysis (Agilent Technologies, Deerfield, IL) and converted to concentrations on a dry weight basis by comparison with the internal standard.

Statistical Analysis.

We separated monoterpene and sesquiterpene data, as the two chemical groups are biosynthetically distinct, and are influenced by separate genetic factors (Keszei *et al.* 2008). Initially we used principal components analysis to identify chemotypes as had been done previously (Butcher *et al.* 1994), however this method did not unambiguously separate some of the published chemotypes which were based on polymodal and disjunct distributions of major oil components (data not shown). Consequently, we calculated Pearson's correlation coefficients on the entire sample set based on the absolute concentrations of terpenes to eliminate false negative correlations potentially introduced when data on terpene proportions are used. The individual components were divided into groups based on the correlations, and the proportion of each of these groups was plotted on a ternary barymetric plot. To predict the effect of individual biosynthetic genes, we re-calculated Pearson's correlation coefficients within each of the resulting groups to obtain chemical groups characteristic of individual chemotypes.

Results

Gas chromatography resolved 47 compounds from the leaf oils comprising 20 monoterpenes and 27 sesquiterpenes. Significant variation in leaf oil concentration was found both between sites and among chemotypes. The lowest foliar oil concentration recorded was a tree of chemotype 3 from Bald Rock Creek, Queensland (10 $mg \cdot g^{-1}$ dry matter (DM)), whereas the highest yielding was an individual of chemotype 5 at Wooli Road, NSW with a foliar oil concentration of 141 $mg \cdot g^{-1}$ DM. The distribution of oil yield shows a normal distribution throughout the species, and this pattern persists when examined at the level of individual chemotypes (data not shown).

Monoterpenes

Monoterpenes are the dominant components of *M. alternifolia* leaf oil, and comprised between 9 and 128 $mg \cdot g^{-1}$ of dry leaf weight. Sesquiterpene content ranges between 0.8 $mg \cdot g^{-1}$ and 14 $mg \cdot g^{-1}$ or between 5 and 21% of the total leaf oils. Mono- and sesquiterpene concentrations are positively correlated with each other, the ratio of monoterpenes to sesquiterpenes is continuous and polymodal, with a major mode at 62%, one at 80%, and one at 86%. The three major monoterpenes: terpinen-4-ol, 1,8-cineole, and terpinolene show well defined, discontinuous concentration ranges from which it is possible to assign all but one of the sampled individuals to known chemotypes.

Pearson's correlation values between all of the monoterpenes within the full sample set established three groups showing strong intercorrelations ($r > 0.8$).

Group A: α-thujene, α-terpinene, γ-terpinene, terpinen-4-ol,

Group B: α-pinene, β-pinene, myrcene, limonene, 1,8-cineole, α-terpineol

Group C: α-phellandrene, terpinolene, linalool

Common biosynthetic intermediates (Figure 3) suggest that these three main groups may be the products of three distinct terpene synthases. *(Z)*-piperitol, *p*-cymene, β-phellandrene, sabinene and *(Z)*-sabinene hydrate did not show strong ($r > 0.8$) correlations to any other group, however, they showed the strongest correlation to group A. As each of these compounds readily undergoes post-biosynthetic conversion or is the product of such a process (Bohlmann *et al.* 1999; Keszei *et al.* 2008) their concentration may not directly reflect enzymatic activity. As such, the lack of a strong positive correlation need not rule out biosynthetic relatedness. Provided that there was either a strong statistical correlation or evidence of a common biosynthetic origin, monoterpenes were assigned to one of the three groups. The variation in the proportions of these three groups relative to the total monoterpene pool is discontinuous, and clearly resolves all chemotypes (Figure 2).

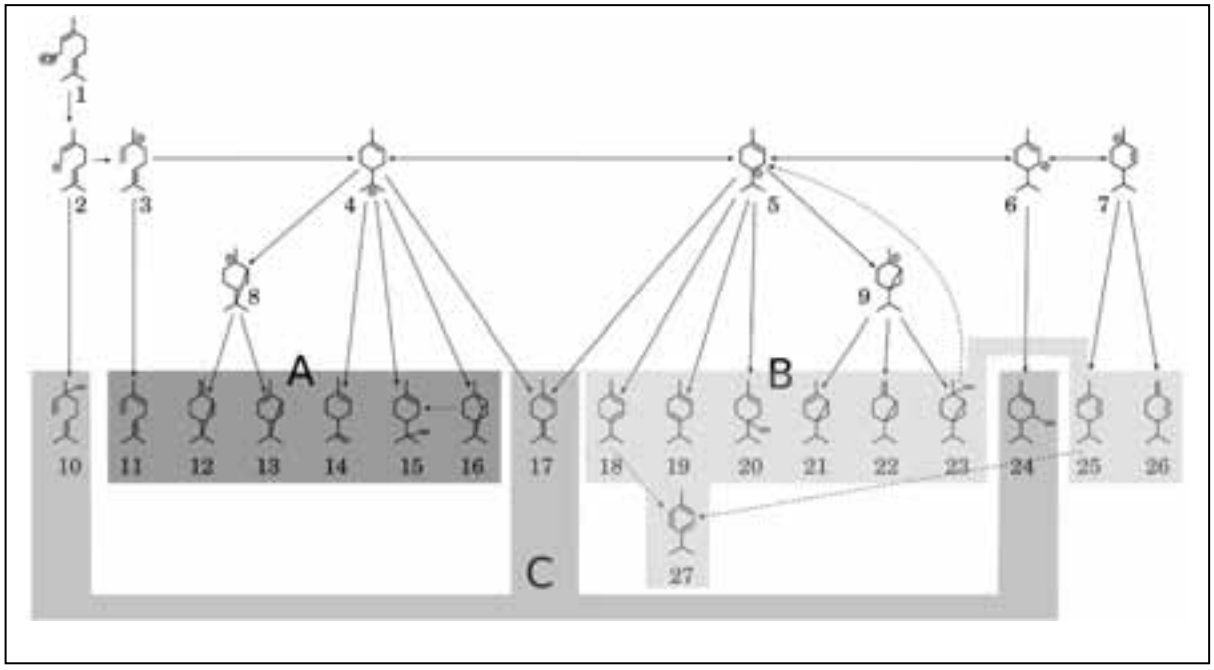

Figure 3. Biosynthetic relationships between monoterpenes found in *M. alternifolia*. The products of the three proposed terpene synthases are indicated by shaded boxes. 1: geranyl diphosphate, 2: linalyl carbocation, 3: geranyl carbocation, 4: α-terpinyl carbocation, 5: terpinenyl carbocation, 6: piperitenyl carbocation, 7: phellandryl carbocation, 8: pinyl carbocation, 9: sabinyl carbocation, 10: linalool, 11: myrcene, 12: β-pinene, 13: α-pinene, 14: limonene, 15: α-terpineol, 16: 1,8-cineole, 17: terpinolene, 18: α-terpinene, 19: γ-terpinene, 20: terpinen-4-ol, 21: α-thujene, 22: sabinene, 23: sabinene hydrate, 24: piperitol, 25: α-phellandrene, 26: β-phellandrene, 27: *p*-cymene.

The leaf oil of one individual from Bald Rock Creek, Queensland showed 1,8-cineole and terpinolene as the major components, however, the ratio between the two compounds did not fit any of the established chemotypes. Furthermore, the ternary plot resolved the individual as a distinct separate group.

Looking at the contributions of the three independent biochemical groups in Figure 2, it is evident that out of the seven chemotypes, chemotypes 1, 2 and 5 represent cardinal chemotypes dominated by a single compound, and chemotypes 3, 4 and 6 fall between these as intermediates. In the ternary plot, only chemotype 5 is plotted on an apex, while chemotype 1 shows significant contributions from both Groups B and C, and chemotype 2 shows significant contributions from Groups A and B. This can be explained by two processes: the major monoterpene synthases in chemotypes 1 and 2 may also synthesize compounds characteristic of the other major biosynthetic groups. In this case, compounds belonging to separate groups will show strong correlations in the cardinal chemotypes. It is also

possible that the monoterpene profile of these chemotypes is determined by more than one active terpene synthase. If this were the case we would expect the correlation matrices to separate monoterpenes of different biosynthetic origins into separate groups.

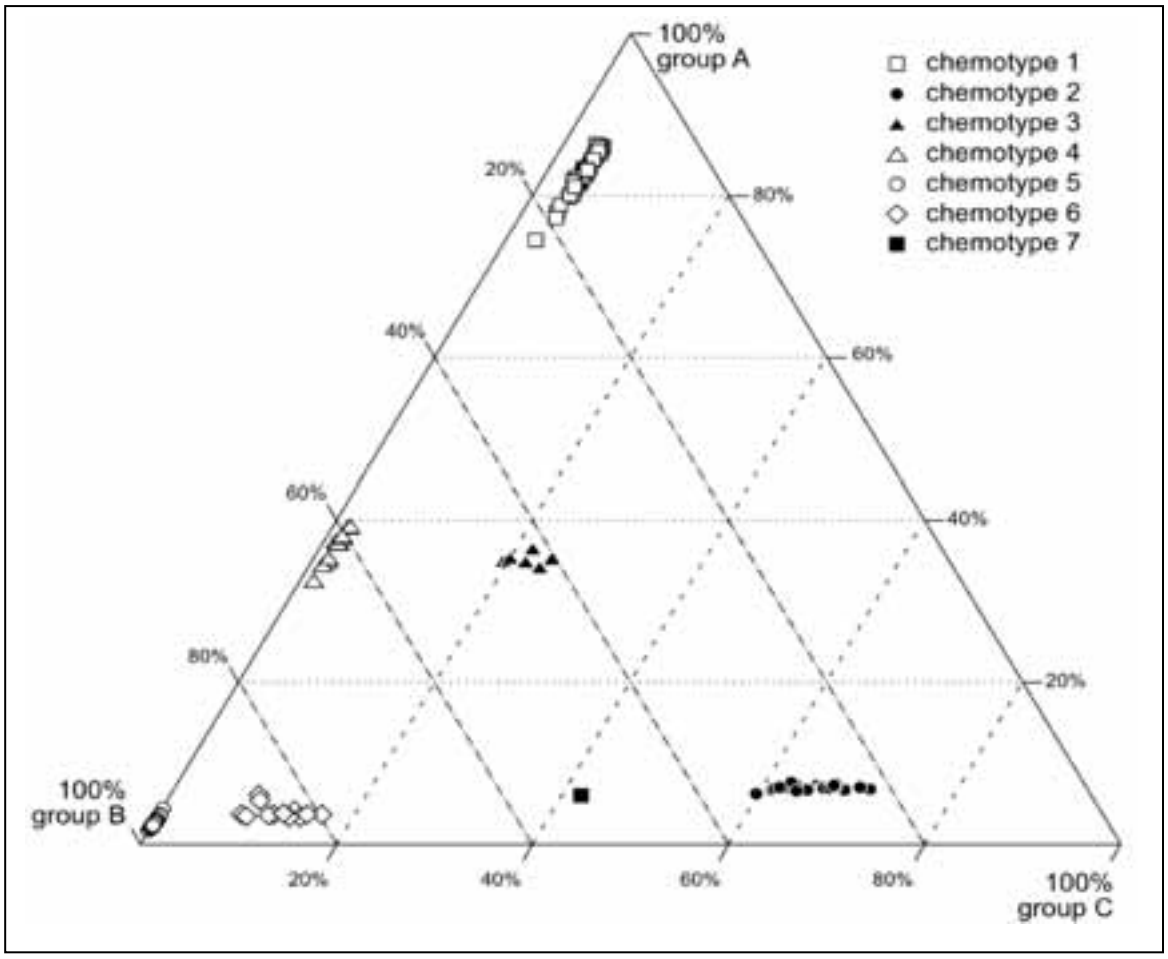

Figure 4. The distribution of the three biochemically based groups of Tea Tree monoterpenes expressed as a percentage of total monoterpenes and shown on a ternary barycentric graph. All six chemotypes are resolved and show discontinuous distributions.

To test these possibilities, Pearson's correlation coefficients were calculated between each of the monoterpenes in the cardinal chemotypes, and compared to the overall pattern. Table 2 shows the resulting correlation matrices. The correlations between monoterpenes in the individual chemotypes illustrate the likely products of the distinct enzymes (Table 3).

In chemotype 1, dominated by terpinen-4-ol, two groups emerge:
Group A1: α-thujene, a-terpinene, γ-terpinene, terpinolene, terpinen-4-ol, *(Z)*-piperitol.
Group B1: limonene, 1,8-cineole.

Myrcene, α-pinene, β-pinene, α-terpineol, α-phellandrene, and β-phellandrene show significant correlation to compounds in both groups, but they show somewhat stronger correlation to Group A1. Linalool, *p*-cymene, sabinene and sabinene hydrate show no significant correlations to other compounds.

In the terpinolene dominated chemotype 2, two strong groups can be distinguished:
Group C2: α-thujene, α-phellandrene, α-pinene, α-terpinene, γ-terpinene, terpinolene, linalool.

Group B2: α-pinene, myrcene, limonene, β-phellandrene, 1,8-cineole, α-terpineol, γ-terpinene, terpinen-4-ol.

β-pinene, *p*-cymene, sabinene and sabinene hydrate show no significant correlation to other compounds.

In chemotype 5, dominated by 1,8-cineole, only a single distinct group could be identified:
Group B5: α-pinene, β-pinene, myrcene, a-phellandrene, limonene, β-phellandrene, 1,8-cineole, α-terpineol, terpinen-4-ol.

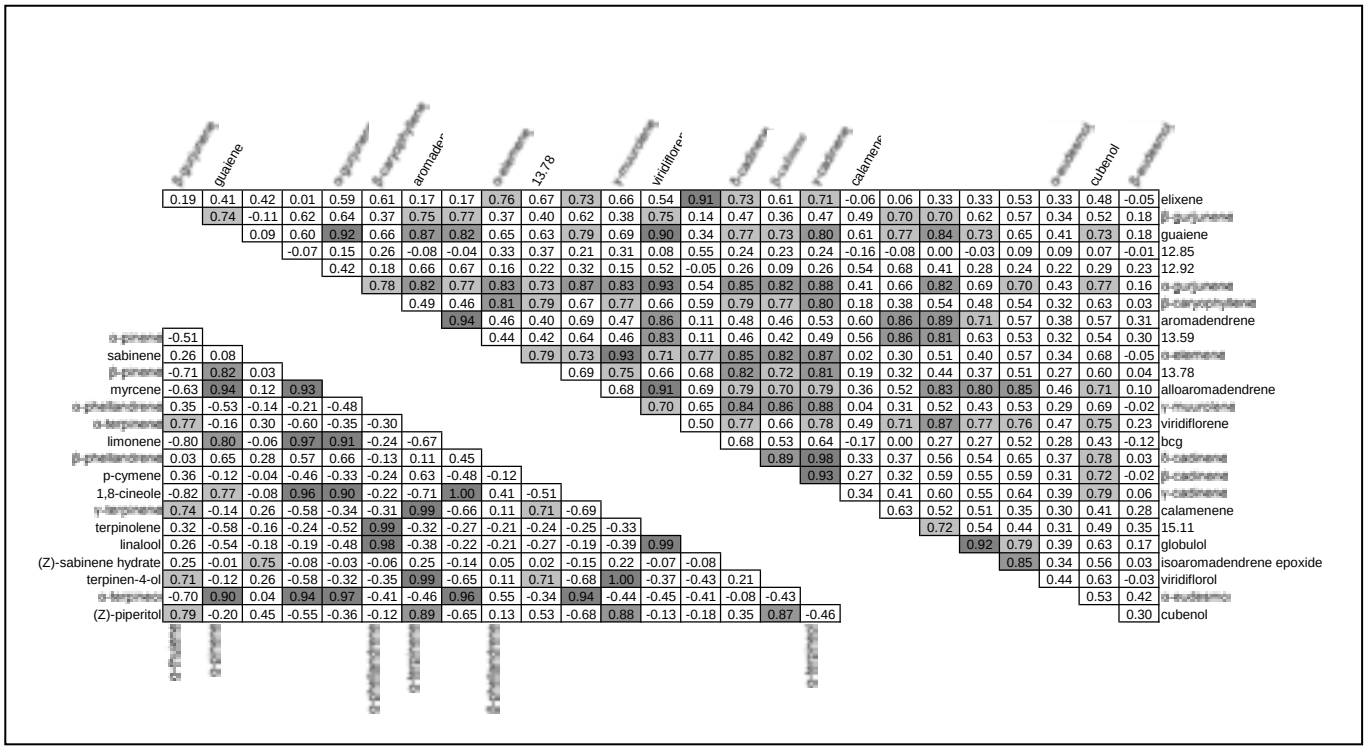

Table 2: Pearson's pairwise correlation coefficients between individual terpene concentrations, shown separately for monoterpenes (below) and sesquiterpenes (above). Values above 0.7, 0.8 and 0.9 are emphasized using incremental shading.

Sesquiterpenes

Sesquiterpenes have always been considered minor and therefore insignificant components of the leaf oils of all *M. alternifolia* samples. However, we found that their contribution to some of the oils was as high as 20% of the total oil that is higher than that of the main monoterpene groups in some chemotypes.

Calculation of the Pearson correlation coefficients between the concentrations of the sesquiterpenes (Table 3) showed strong correlations, and the sesquiterpenes β-gurjunene, unknown peaks at 12.85 min, 12.95 min, and 13.59 min, α-elemene, unknown peak at 13.76 min, calamenene, unknown peak at 15.11 min, β-eudesmol showed discontinuous distributions. Interestingly, none of the sesquiterpenes with discontinuous distributions correlated with each other. Based on Pearson's correlation coefficients of $r > 0.8$, we established the following groups;

Group D: elixene, bicyclogermacrene

Group E: guaiene, α-gurjunene, aromadendrene, unknown peak at 13.59 min, allo-aromadendrene, viridiflorene, globulol, viridiflorol, isoaromadendrene epoxide, unknown peak at 15.11 min, β-gurjunene

Group F: α-elemene, γ-muurolene

Group G: 13.79min, δ-cadinene, β-cadinene, γ-cadinene, cubenol, (also showed strong correlation to α-gurjunene and γ-muurolene)

None of the other sesquiterpenes showed significant correlation to any other compounds.

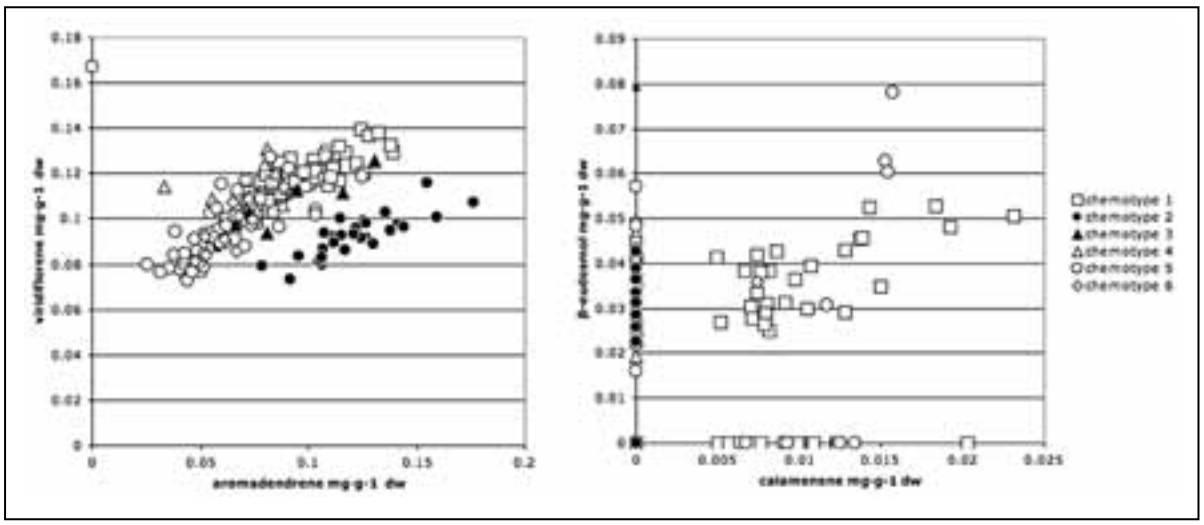

Figure 5: The relationships between aromadendrene vs. viridiflorene, and β-eudesmol vs. calamenene concentrations expressed as $mg \cdot g^{-1}$ dry leaf weight. Chemotype symbols used are consistent with those used in Figure 2.

The two strongest sesquiterpene patterns reflecting the monoterpene chemotype were the ratio between aromadendrene and viridiflorene (Figure 3), and the presence/absence of calamenene, shown together with another sesquiterpene showing discontinuous but independent distribution: β-eudesmol. Chemotype 2 shows a different viridiflorene to aromadendrene ratio to all other chemotypes, although correlation between the two compounds is high ($r > 0.8$) in all chemotypes. Calamenene content is also indicative of chemotype 2. Of the cardinal chemotypes, chemotype 2 is the only one that shows no calamenene content in any of the individuals sampled (Figure 3). None of the other chemotypes were correlated with any of the sesquiterpene patterns, and no more sesquiterpene patterns were found to be correlated with monoterpene chemotypes. The correlation among sesquiterpenes were generally lower than among monoterpenes, however this may be a result of their lower concentrations leading to lower measurement accuracy.

Geographic Distribution of Chemotypes

Variation of chemotypes was observed both within and among sites. Two populations contained only a single chemotype (chemotype 1), while two sites (Chaffin Swamp and Bald Rock Creek) contained 5 of the 7 chemotypes (Table 1). Interestingly, these two populations also contained some of the extremes of oil yield. Populations were dominated by either chemotype 1, 2 or 5, while other chemotypes occurred at lower frequencies. There was also a distinct chemical separation of trees across the geographic range with respect to these former three chemotypes. Approximately 70% of trees sampled at both Queensland sites were chemotype 2, while in the Richmond River valley in

northern NSW 95% of trees sampled chemotype 1, and more than 80% of trees along southern tributaries of the Clarence River system were chemotype 5 (Table 1).

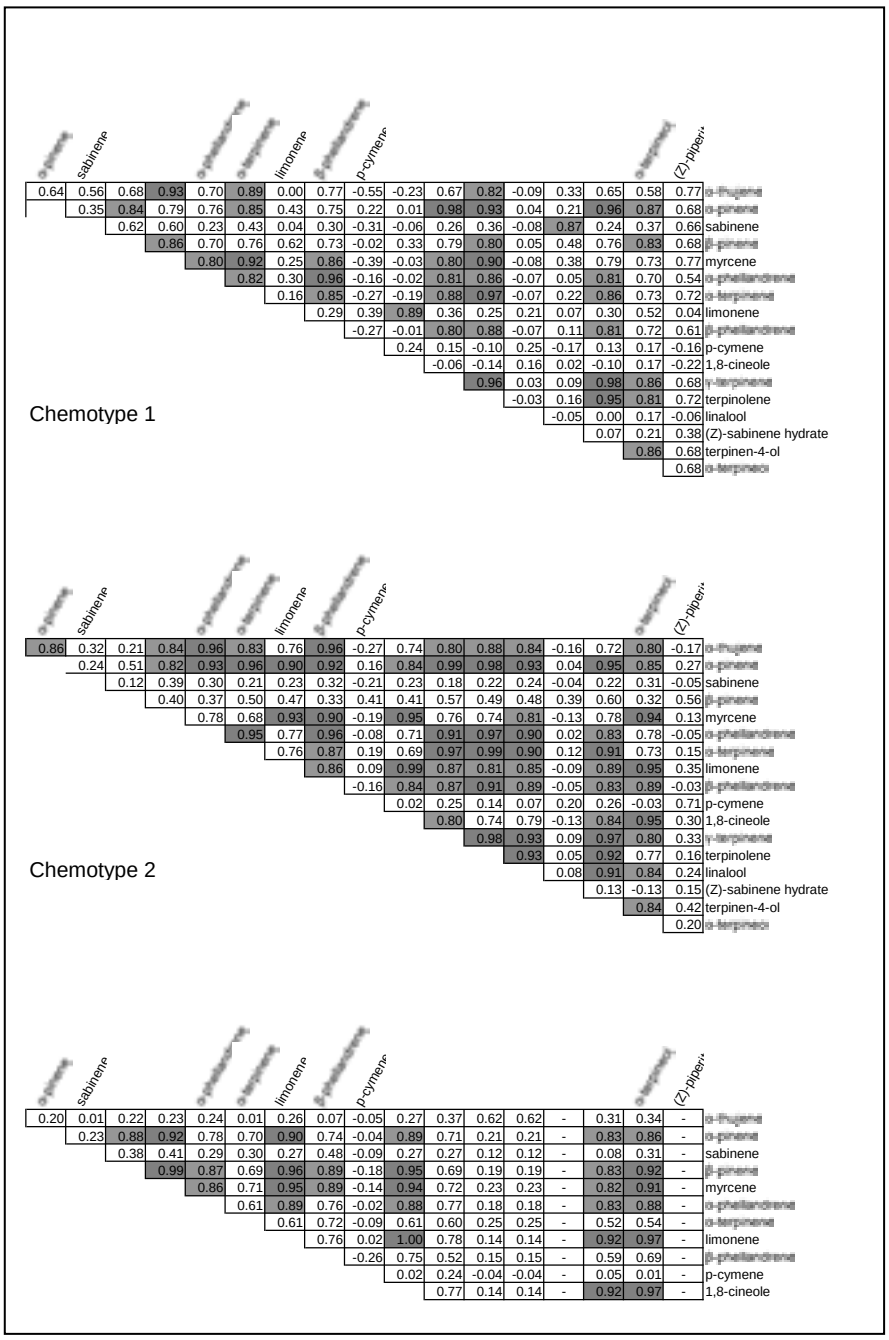

Table 3: Pearson's pairwise correlation coefficients between individual monoterpene concentrations shown separately in the three cardinal chemotypes. Values above 0.8 and 0.9 are emphasized using incremental shading.

Discussion

Monoterpene Chemotypes

We identified six chemotypes in this study, all of which had been previously described (Butcher *et al.* 1994) and an individual tree which did not group with any of the recognized chemotypes. We were able to group the monoterpenes on the basis of their likely biosynthetic origins (Figure 1) and calculate intercorrelations between them. Here we discuss the possible underlying biochemical factors that lead to these distinct patterns.

The two likely processes are based on different types of catalytic action. First, each terpene may be produced by a single, specific enzyme and certain terpenes may co-occur as a result of genetic linkage (Wilderman *et al.* 2004). Alternatively, terpenes in each group may be products of the same reaction by a single enzyme (Wise *et al.* 1998). It is more common for a single terpene synthase to produce multiple products than a single product exclusively, and therefore we expect that the latter explanation is the most likely. Support for this proposal comes from functional assays of terpene synthase enzymes in other species. Heterologous expression and functional characterisation of 1,8-cineole synthase from *Arabidopsis thaliana* (Chen *et al.* 2004) and *Salvia officinalis* (Wise *et al.* 1998) show that those enzymes catalyse not only the formation of 1,8-cineole but also limonene, myrcene, α-and β-pinene and α-terpineol as by-products, and that the ratios of products are constant characteristics of these enzymes. These compounds are also the major contributors to the Group B terpenes of *M. alternifolia*.

A single enzyme can also synthesise one product, which can subsequently be modified into multiple products by other processes. In particular, the Group A terpenes in *M. alternifolia* may result from re-arrangement of unstable *cis*-sabinene hydrate. In *Origanum majorana*, sabinene synthase also produces sabinene hydrate, which is rearranged after steam distillation to give terpinen-4-ol, α- and γ-terpinene and terpinolene (Fischer *et al.* 1987). Young leaves of chemotype 1 in *M. alternifolia* are rich in *cis*-sabinene hydrate (41%) (Russell *et al.* 2002) and these are also known to undergo temperature-catalysed transformation to the aforementioned products during hydrodistillation (Brophy *et al.* 1989). It has also been shown that *cis*-sabinene hydrate can undergo a non-enzymatic acid solvolysis in *vivo* as the leaf ages (Cornwell *et al.* 1995; Southwell *et al.* 1990), resulting in Group A monoterpenes in the mature leaf of *M. alternifolia*. This reaction would be expected to contain terpinolene as a by-product, and in chemotype 1 individuals, there is indeed a strong positive correlation between concentrations of terpinen-4-ol and terpinolene.

In chemotypes 2, 3, 6 and 7 however, terpinolene is most likely synthesised by a separate terpinolene synthase. In *M. alternifolia,* the concentration of terpinolene is positively correlated with those of linalool and α-phellandrene, and together they comprise Group C terpenes. However, when just focusing on chemotype 2 individuals, it can be seen that α-terpinene, γ-terpinene, α-phellandrene and β-phellandrene also show strong correlation with terpinolene. These compounds therefore may be synthesised by more than one of the enzymes that define the terpene composition of tea tree oil.

Three studies that examined the expression of terpinolene synthases show different products (Bohlmann *et al.* 1999; Huber *et al.* 2005; Wise *et al.* 1998). This suggests that terpinolene can be produced from different carbocation intermediates, and that it is not unusual that the monoterpene components positively correlated to terpinolene in *M. alternifolia* do not match the profile of any currently known terpinolene synthase.

Three groups of monoterpenes describe the majority of variation in Australian Tea Tree. Pairwise correlation of individual monoterpenes as a proportion of total monoterpenes shows that almost all monoterpenes show strong correlation to one of the major groups. The lowest correlation to any single group is shown by *p*-cymene, however it has been shown that this compound may be an oxidation

product that can be derived from multiple monoterpenes (Brophy *et al.* 1989), and its concentration may depend more on extrinsic factors that affect *in vivo* oxidation, rather than on individual enzymatic processes.

Sesquiterpene Chemotypes

This is the first study to examine patterns between sesquiterpenes in *M. alternifolia*, and it is clear from the analyses that there were strong correlations between individual compounds. Two significant patterns were observed in sesquiterpene variation, which corresponded to monoterpene chemotypes. First was within a group of correlated sesquiterpenes, most likely synthesized by the same enzyme. Viridiflorene and aromadendrene showed two distinctly different ratios, and the higher viridiflorene ratio was exclusive to chemotype 2. This would lead us to expect two allelic variants of aromadendrene synthase, which are responsible for synthesizing the same suite of compounds in slightly different, but nevertheless characteristic ratios. The second compound to show a pattern linked to a monoterpene chemotype was calamenene, which was consistently absent from all chemotype 2 individuals. Calamenene is an aromatic cadinane-type sesquiterpene, and as it has three unsaturated bonds, it is likely that its synthesis is linked to the presence of a reductase from the P450 superfamily of enzymes, as in mint monoterpene biosynthesis (Bertea *et al.* 2001). Compared to monoterpenes, there were many more compounds that were not correlated to other constituents and our sampling probably missed some combinations. It is interesting to note, however, that the only variation linked to monoterpene variation could be found in chemotype 2, and none of the other sesquiterpenes showing discontinuous distributions were correlated to each other or to other monoterpene chemotypes. This suggests that formation of sesquiterpenes is generally independent of the monoterpene chemotype in Tea Tree, but that in chemotype 2, there may be correlation at genomic level.

Biochemical Origin of the Chemotypes

We propose that three major enzymes are responsible for the three major chemotypes. These are (i) sabinene hydrate synthase that produces the Group A compounds, (ii) 1,8-cineole synthase that yields Group B compounds and (iii) a terpinolene synthase responsible for group C products. The barycentric plot showing the individual contributions suggests that in chemotypes 1 and 2, a separate 1,8-cineole synthase may also be contributing to the oil profile. Chemotype 3 demonstrates that all three enzymes may be present at once; therefore in a diploid organism such as *M. alternifolia*, at least two loci need to be involved in coding for the foliar monoterpene synthases that determine chemotype.

The effect of the proposed individual terpene synthases appears to be codominant. This is supported the intermediate position of chemotypes 3, and 4, and 6 compared to 1, 2 and 5 as shown in Figure 2. While chemotypes 3 and 4 fall exactly halfway between the contributing cardinal chemotypes indicating equal contribution from all sides, chemotypes 6 and 7 are not as straightforward. In chemotype 6, the contribution of the group B and group C compounds is 3:1 and 2:1. This may indicate two things: (i) our hypothetical 1,8-cineole synthase utilises GDP as a substrate more efficiently than does terpinolene synthase, or (ii) the two genes have a cumulative effect, where product proportions are determined by the number of alleles present of each synthase. In the case of multiple copy gene families such as terpene synthases, duplication is often the mechanism behind the increase in the number of genes contributing to the same process. Among duplicated genes, the process of gene conversion can occur whereby similar sequences recombine with each other regardless of actual chromosomal topology, and in effect a sharing of alleles will occur across loci. In such a scenario, the 3:1 contribution may be explained by two loci sharing alleles.

The distribution of chemotypes across the natural range further supports the idea that chemotypes 3, 4, 6 and 7 are likely intermediate chemotypes, particularly because, unlike chemotypes 1, 2 and 5, neither our or previous studies (Homer *et al.* 2000) have found them to occur in mono-chemotypic

populations. Furthermore, if they are intermediates, then they should be found in populations that contain the contributing cardinal chemotypes. In one population, Dilkoon Creek, which occurs between the two chemically distinct regions of NSW, 35% of trees recorded were chemotype 4. Chemotypes 1 and 5 are also present in this population. These are the two chemotypes that were used in a controlled cross experiment that reinforced the intermediate nature of chemotype 4 (Shelton *et al.* 2002b). Finally, chemotypes 3, 6 and 7 that contain group C compounds occur mostly in Queensland populations, where chemotype 2 is dominant.

Butcher et al. (Butcher *et al.* 1994) have raised the possibility of chemotypes 2 and 5 being the results of introgression from sister species *M. linariifolia* and *M. trichostachya*. In the case of *M. linariifolia*, chemical analyses have been devised which are able to separate the two species based on the ratio of *cis*- and *trans*-sabinene hydrate (Southwell *et al.* 1990). This shows that even though the two species apparently share similar chemotypes, the final chemistries are obtained via enzymatically different routes. This separation does, however depend only on differences between Group A compounds, and without at least as thorough analysis of the corresponding group B dominated chemotypes, it is hard to reach any conclusive answer. On the other hand, chemotype 2 of *M. alternifolia* is very similar to a corresponding chemotype of *M. trichostachya* (Brophy *et al.* 1996). Furthermore, not only does chemotype 2 differ from all others in its monoterpene composition, but it also displays further differences in two independent sesquiterpene traits. In other plants, genes for terpene biosynthesis often co-locate on the chromosomes and if this is also true in *Melaleuca*, the consistent co-variation of these traits indicate that chemotype 2 is not just the result of variation of a single gene, but of multiple linked loci.

Conclusion

The examination of purely chemical data from *M. alternifolia* from a biochemical point of view has enabled us to make much deeper conclusions about the origins of the different chemotypes. Using biochemical knowledge with the power of statistically analysing a large number of samples, it has been possible to develop a clear set of hypotheses about the nature of terpene synthases that are contributing to natural variation. Ultimately, this gives us a good framework for uncovering the individual molecular processes and identifying the actual genes contributing to the maintenance of such remarkable chemical variation of an industrially, medicinally and ecologically important species.

Chapter 2 - How did tea tree oil get its medicinal properties? Evolution of terpene chemotypes in *Melaleuca alternifolia*

Introduction

Intra-specific variation in plant phenotypes can have profound ecological consequences (Edwards *et al.* 1993). In particular, variation in defence compounds influences herbivores as selective agents on the survival of some individuals over others, and even dictates the success of biological control programmes for weeds (Wheeler 2006).

Understanding how intra-specific variation in plant chemical profiles arises at the molecular level would help explain how it is maintained in natural populations. The different genes responsible for chemical variation may arise in a species *de-novo* via gene duplication, they may be retained following speciation by lineage sorting, or they may be acquired following hybridisation. From an evolutionary perspective, a new chemotype may be the result of neo-functionalisation or a change in the regulatory control of duplicated genes. Although studies of model plants have been informative, considering these questions in long-lived woody plants in natural populations has the potential to provide the links between biochemistry and molecular evolution.

Although chemotypic variation has been intensely studied in multiple taxa, it is not known what types of molecular patterns contribute to the chemotypic variation in a species across its full geographic range. In thyme, we have a good understanding of the mode of inheritance and the epistatic effects resulting from the interactions in biochemical pathways that bring about known natural chemotypes (Vernet *et al.* 1986), but do not know the action or the ancestry of the individual genes responsible (Latta *et al.* 2003, Pluhar, 2008 #18). Studies in maize (Köllner *et al.* 2004) and sage (Kampranis *et al.* 2007) have provided insight into the way genetic drift can contribute to the mechanism bringing about phenotypic change in the chemical profile of plants, but it remains a challenge to link this to natural populations.

Chemotypic variation in foliar terpenes appears particularly widespread in the Myrtaceae. In a compilation of leaf oils of 111 species of *Eucalyptus* (Boland *et al.* 1991) 28 species show clear chemotypes,. Furthermore, in the compilation of essential oils of tropical *Asteromyrtus, Melaleuca* and *Callistemon*, (Brophy *et al.* 1996) show that 16 of the 37 species studied also exhibit different chemotypes. This means that close to half (42%) of the species and covered by these two studies show chemotypic variation.

In addition to having a large proportion of chemically variable species in Myrtaceae, the genes responsible for the variation, i.e. the terpene synthases also show large variability, with clades apparently segregating based on function containing members from both *Eucalyptus* and *Melaleuca* (Keszei *et al.* 2010b). This indicates that in Myrtaceae, genetic variation in the terpene synthases responsible has been present before the separation of present day taxa, and that chemotypic variability is an ancient trait in this Family.

Australian tea tree (*Melaleuca alternifolia:* Family Myrtaceae) is a long lived woody plant that occurs as six distinct foliar terpene chemotypes. Chemotypes 1, 2 and 5 are dominated by compounds from one biosynthetic group each ("cardinal" chemotypes); while chemotypes 3,4 and 6 are believed to be intermediate chemotypes. Only one of these chemotypes yields the medicinally valuable essential oil dominated by the monoterpene terpinen-4-ol (Butcher *et al.* 1994, Keszei, 2010a #28). Studies of the foliar chemistry of *M. alternifolia* led to the hypothesis that only three distinct terpene synthases are

responsible for the biosynthesis of over 80% of the leaf oil (Keszei *et al.* 2010). Some of the chemotypes found in *M. alternifolia* are also found in closely related species of *Melaleuca* (Southwell *et al.* 1992, Southwell, 1990 #27), thus providing the opportunity to understand the evolution of different chemotypes and whether particular chemotypes are more ancient than the species that contain them.

In this study we focus on the origins of chemotypes within a species to establish whether present day Myrtaceae developed from an already variable taxon from which only select traits had been retained, or if variability has been achieved independently as a response to environmental pressures.

To address these questions, we first describe the biosynthetic and molecular mechanisms behind individual chemotypes of *M. alternifolia*, secondly we establish the relatedness of the genes that determine the chemotypes within the species, and finally we examine the relationships between related genes in related species of Melaleuca with similar chemical traits.

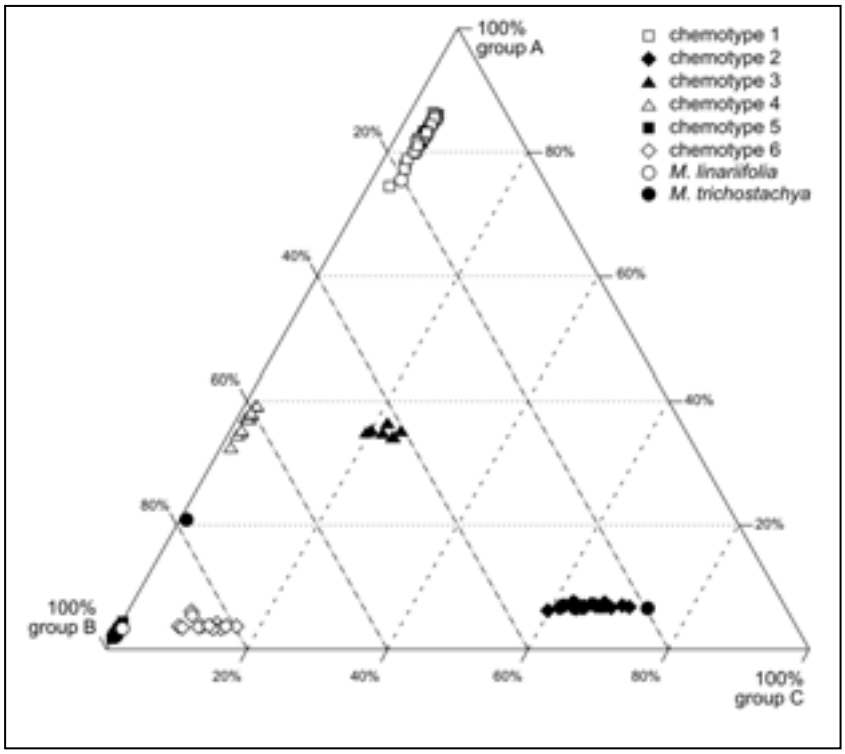

Figure 6: Natural chemotypes of *M. alternifolia*, *M. trichostachya* and *M. linariifolia*. *M. linariifolia* chemotypes 1 and 2 cluster together with *M. alternifolia* chemotypes 5 and 1 respectively, M. *trichostachya* chemotype 1 clusters with *M. alternifolia* chemotype 2, while its chemotype 2 falls between *M. alternifolia* chemotypes 4 and 5.

Materials and methods

Plant material and chemical analysis

Mature leaf (ca. 20 g wet mass) was collected from 20 mature trees at eight sites and fifteen from a ninth site across the natural geographic range of *M. alternifolia* (Table 1). We chose trees that were at least 100m apart to avoid collecting from related trees (Rossetto *et al.* 1999) and the location of each tree was recorded. Samples were refrigerated at 4°C within 2 h of collection. Young leaf was collected

in liquid nitrogen from 20 individuals from populations dominated by different chemotypes, in order to ensure that we had samples from trees belonging to each of the known chemotypes.

Extraction of nucleic acids

We extracted DNA and RNA from young leaf ground in liquid nitrogen using DNeasy and RNeasy kits (QIAGEN). For RNA extraction, we complemented the RLT extraction buffer with PVP and sodium isoascorbate. This combination gave unsatisfactory results in chemotype 2, where the addition of sodium isoascorbate inhibited RNA extraction.

TPS isolation, identification and characterisation

We used 3'RACE to obtain partial transcripts containing the DDxxD motif using the degenerate 'DDXYDfx' primer and T_{35}VN previously used to successfully isolate terpene synthases from 2x species of Myrtaceae (Keszei *et al.* 2010b). We ligated the amplification products into pGEM-T Easy or pCR2.1^{TOPO} cloning vectors, and sequenced the inserts from the M13 priming sites using BigDye v. 3.1 on an ABI 3130 capillary sequencer. Sequence information from the most abundant transcripts in each chemotype was used to design primers for upstream amplification. We used the SMART 5'RACE kit to amplify the 5' ends of the identified genes, and obtained sequence information as per the protocol described for 3'RACE. Following the assembly of the 3' and 5' contigs, we designed primers to obtain full-length cDNA clones. We used the FastPCR Professional software package to design primers, and used these to amplify clones encoding pseudo-mature proteins for characterisation.

MELal;TPS2 was amplified using the following primers: 1 and 2, and MELal;TPS4 using 3 and 4, MELal;TPS5 using 5 and 6. PCR was performed using Advantage 2 polymerase mix (BD Biosciences, Palo Alto, CA). The resulting PCR products were directly inserted as *Bsp*MI fragments into the expression vector pASK-IBA7 (IBA GmbH, Göttingen, Germany). Expression and partial purification of the recombinant protein followed the procedure described in Köllner et al. (2004). To determine the catalytic activity of the recombinant protein, enzyme assays containing 50 µl of the bacterial extract and 50 µl assay buffer (10 mM MOPSO [pH 7.0], 1 mM dithiothreitol, 10% [v/v] glycerol) with 10 µM substrate ((*E,E*)-GPP (Echelon Biosciences, Salt Lake City, UT, USA) and (*E,E*)-FPP, respectively), a divalent metal cofactor (10 mM $MgCl_2$), 0.2 mM Na_2WO_4 and 0.1 mM NaF in a Teflon-sealed, screw-capped 1 ml GC glass vial were performed. A solid phase microextraction (SPME) fibre consisting of 100 µm polydimethylsiloxane (SUPELCO, Belafonte, PA, USA) was placed into the headspace of the vial for 30 min incubation at 30 °C. For analysis of the adsorbed reaction products, the SPME fibre was directly inserted into the injector of the gas chromatograph.

A Shimadzu model 2010 gas chromatograph was employed with the carrier gas He at 1 ml min^{-1}, splitless injection (injector temperature: 220 °C, injection volume: 1 µl), a Chrompack CP-SIL-5 CB-MS column ((5%-phenyl)-methylpolysiloxane, 25 m x 0.25 mm i.d. x 0.25 µm film thickness, Varian, USA) and a temperature program from 50 °C (3-min hold) at 6 °C min^{-1} to 180 °C (1 min hold). The coupled mass spectrometer was a Shimadzu model QP2010Plus with a quadrupole mass selective detector, transfer line temperature: 230 °C, source temperature: 230 °C, quadrupole temperature: 150 °C, ionization potential: 70 eV and a scan range of 50-300 amu. Compounds produced by TPS2, TPS4 and TPS5 were identified by comparison of mass spectra and retention times to those of authentic standards or using the Wiley mass spectra library.

Semi-quantitative PCR and genomic diagnostics

We designed primers to differentiate between TPS4a and TPS4b, as well as between TPS2 and TPS4. PCR chain reaction was carried out using optimised conditions, with 50 µl of starting volume. 5 µl

aliquots were taken out at 24, 27, 30, 33, 36, 39, 42 and 45 cycles, and were visualised together by agarose-ethidium bromide gel electrophoresis. Ethidium bromide fluorescence was measured using BioRad software, at non-saturation exposures. The 500 bp band of 5 µl 100 bp ladder (Axygen Union City CA, USA) was used as a reference.

Results & Discussion

TPS Sequences from *M. alternifolia*

We isolated and sequenced the most abundant monoterpene synthase transcripts from the three cardinal *Melaleuca alternifolia* chemotypes. Besides obtaining full-length sequence from TPS2, partially described as the most abundant monoterpene synthase in chemotype 1 by a previous study (Keszei *et al.* 2010b), we identified two new *M. alternifolia* monoterpene synthase sequences. TPS4 is the characteristic TPS transcript from chemotype 5, and TPS5 from chemotype 2.

Functional characterisation of the full-length clones has confirmed the predicted origin of the major constituents of the monoterpene profile of *M. alternifolia*. Figure 9 shows the product profiles of TPS2, TPS4 and TPS5. The major product of TPS2 is sabinene hydrate, which readily converts to terpinen-4-ol, α-terpinene and γ-terpinene and terpinolene. TPS4 has 1,8-cineole as its major product, however it also contributes to the production of sabinene, limonene and α-terpineol. TPS5 produces terpinolene as its major product, and appears to be the most product-specific of the three terpene synthases.

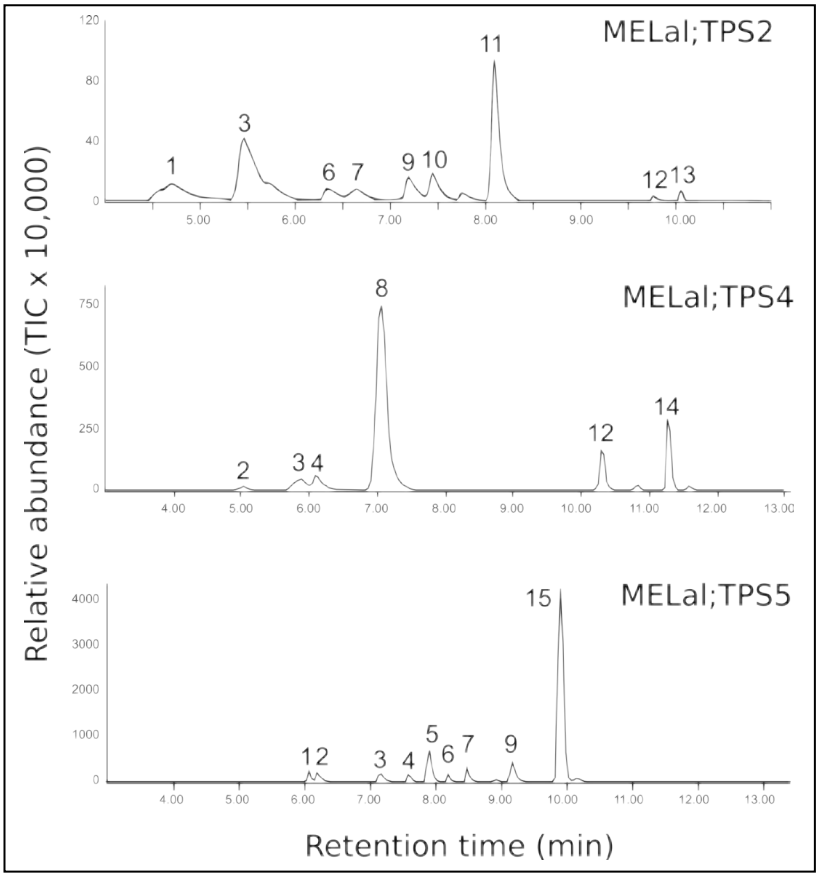

Figure 7 : Product profiles of terpene synthases expressed in the leaves of the major *M. alternifolia* chemotypes. 1: α-thujene, 2: α-pinene, 3: β-sabinene, 4: β- myrcene, 5: α-

phellandrene, 6: α-terpinene, 7: unknown monoterpene, 8: 1,8-cineole, 9: γ-terpinene, 10: E-sabinene hydrate, 11: Z-sabinene hydrate, 12: terpinen-4-ol, 13: α-terpineol, 14: geraniol (artefact), 15: terpinolene

Apart from the full-length monoterpene synthases, we have acquired numerous partial sequences, including an expressed putative sesquiterpene synthase (TPS1), and several additional monoterpene synthases (TPS5b and TPS6a-d), which we have not been able to isolate as cDNA fragments.

No significant differences could be found between sesquiterpene synthase fragment sequences originating from the different chemotypes, which is in agreement with the observation (Keszei *et al.* 2010) that sesquiterpene chemistry in *M. alternifolia* is less variable and not related to the monoterpene chemotypes.

Genomic and transcriptomic patterns of TPS genes

We searched for the presence of TPS2, TPS4 and TPS5 in genomic DNA. Whereas both TPS4 and TPS5 occur in all genomes (data not shown), TPS2 (sabinene hydrate synthase) only occurs in the genomes of individuals from chemotypes 1, 3 and 4, i.e. the chemotypes characterised as containing a significant proportion of terpinen-4-ol in their leaf oil (Figure 10)

Figure 8: PCR using selective primers for TPS2, a sabinene hydrate synthase from genomic DNA of 12 individuals representing all chemotypes using primers for exon 6 (left). PCR using selective primers for TPS4b from genomic DNA (right). Numbers indicate chemotypes. Amplification is only shown from individuals, which show high sabinene hydrate or terpinen-4-ol concentrations in their leaf oils.

We performed semi-quantitative RT-PCR to ascertain the expression levels of the three terpene synthases in leaves from different chemotypes (Figure 11). TPS2 (sabinene hydrate synthase) was only present in the transcriptome of individuals that contained terpinen-4-ol. TPS4 (1,8-cineole synthase) was highly expressed in all chemotypes, however TPS5 (terpinolene synthase) was most highly expressed in individuals from those chemotypes characterized by high foliar concentrations of terpinolene (Chemotypes 2, 3 and 6) and its expression was lower in all other chemotypes. Within individuals, the levels of expression in the early exponential phase of amplification are consistent with leaf chemistries, and provide characteristic patterns of individual chemotypes.

To confirm the identity of TPS5 in individuals from the low-terpinolene chemotypes, we sequenced the fragments obtained from genomic amplification from individuals of Chemotype 5 and Chemotype

1. The 20 clones examined represent six distinct sequences, with several amino acid differences, and even indel-type length differences between them. Phylogenetically, these sequences separated into two distinct groups within clade 1. Those most similar to the functionally characterised terpinolene synthase (TPS5) were named TPS5a and TPS5b, while the remaining four, not yet isolated as cDNA were named TPS6a-d.

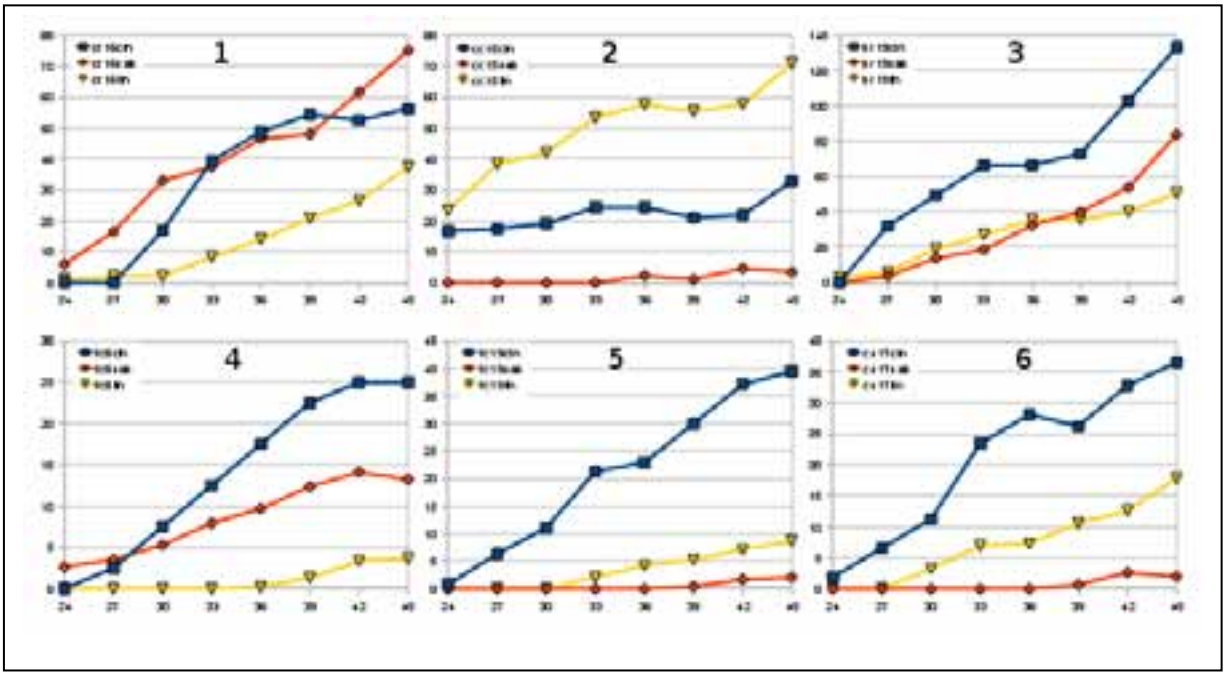

Figure 9: Semi-quantitative RT-PCR from all chemotypes. Cineole synthase (TPS4) is indicated by blue lines, sabinene hydrate synthase (TPS2) by orange lines, and terpinolene synthase (TPS5) by yellow. The x axis indicates the PCR cycle number, and the y axis indicates approximate product concentration (pmol/ul). As expected from genomic assays, TPS2 is only significantly expressed in individuals with sabinene hydrate in their leaf oils. TPS4 is highly expressed in all chemotypes, regardless of terpinolene concentration. Even though we have shown TPS5 to be present in the genome of all individuals, we did not expect to see significant levels of gene expression in individuals with little or no terpinolene in the oil profiles (chemotypes 1, 4 and 5).

Relationships between the terpene synthases determining the cardinal chemotypes.

Comparison of the three major monoterpene synthases shows that although terpinolene and sabinene hydrate synthases are more similar in their products of catalysis (with 5 products in common), it is in fact the cineole and the sabinene hydrate synthase genes which are similar to each other at the sequence level, with only 23 amino acid differences in the open reading frames. As TPS2 shows a simple pattern of genomic presence/absence, we expect it to be present in a single copy.

Cineole synthase is present in all chemotypes, and the ratio between cineole and the other major monoterpenes is variable (1:1 with terpinen-4-ol in chemotype 4, 1:1:1 with terpinen-4-ol and terpinolene in chemotype 3, and 1:3 with terpinolene in chemotype 6). This suggests that there is more than one cineole synthase contributing to the occurrence of cineole in the leaf oils of different chemotypes. To further pursue the question, we selected a 1kbp section of TPS4 from the C-terminal domain, and sequenced it directly after amplification from the genomic DNA of 152 individuals

(Figure 11). The obtained sequences show differences almost only in the introns. This indicates that both forms are still active and important for the plant, and neither of them has become a pseudogene. The fact that no exon variation, and only a little intron variation can be observed also suggests that the differentiation between the two variants is very recent. Many of the obtained sequences were heterozygous, which indicated the presence of more that one variant of the TPS4 gene. There is value in the heterozygosity of the sequence traces however, as we can infer that in the individuals from which clean reads were obtained, only a single copy of TPS4 is present, and based on these, the heterozygous traces could be clearly separated into the contributing haplotypes.

The phylogram of the TPS4 fragments (Fig 10) shows two variants of cineole synthase in *M. alternifolia*, one of which (TPS4b) is present in all terpinolene rich chemotypes (Chemotypes 2, 3, 6), while the other (TPS4a) is present in all chemotypes except for chemotype 2. The heterozygous traces also show us that a single TPS4 can be found in chemotypes 1,2,4 and 5, whereas only the terpinolene-rich intermediate chemotypes carry two copies of the gene.

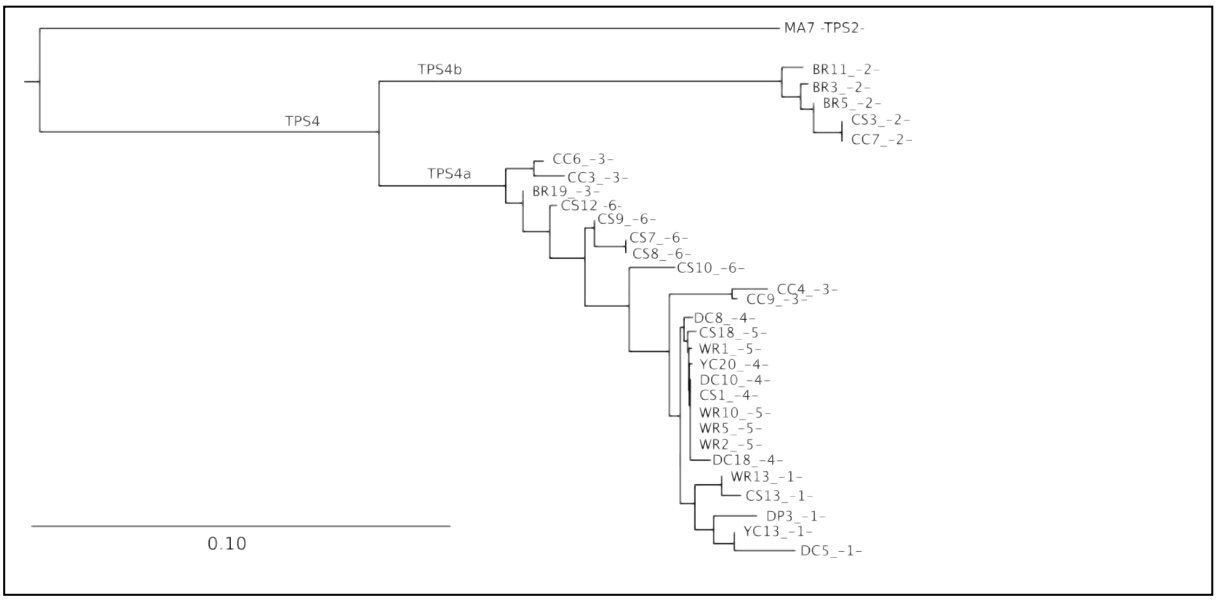

Figure 10: Maximum Likelihood tree of *M. alternifolia* TPS4 genomic fragments. Sabinene hydrate synthase, the most similar monoterpene synthase sequence was used as the outgroup (MA7 TPS2). The terminal node labels show the population abbreviations and individual numbers followed by the chemotype. The main phylogenetic split is between chemotype 2 sequences (TPS4b), and all others (TPS4a). Intermediate chemotypes are highly heterozygous, resulting in poor placement especially of sequences from chemotypes 3 and 6.

This pattern indicates that contrary to our expectations, TPS4 sequence variation does not correlate with cineole levels, but instead with the occurrence of terpinolene in the leaf oil profiles. Primers designed to differentiate between TPS4a and TPS4b confirm the result (TPS5 is responsible for the synthesis of terpinolene; and in probing for genomic presence we found a number of closely related sequences. In fact, TPS5 and TPS6 are the most diverse monoterpene synthases we found in tea tree.

Relationships between terpene synthases

The molecular basis and origins of tea-tree chemotypes

A biosynthetic interpretation of the patterns of terpenes in *M. alternifolia* leaves led to the hypothesis that as few as three genes may explain the qualitative variation in tea tree oil. The isolation and characterisation of cDNA clones TPS2 (sabinene hydrate synthase), TPS4 (1,8-cineole synthase) and

TPS5 (terpinolene synthase) confirms this. The products of the individual enzymes, each representing the most abundant monoterpene synthase transcript in the three cardinal chemotypes, match the biosynthetic groups predicted by analysing correlations in the foliar terpenes (Keszei *et al.* 2010b).

The most important factor that determined the contribution of TPS4 (1,8-cineole synthase) and TPS5 (terpinolene synthase) to the final oil profile of any chemotype was differences in regulation of expression. We hypothesise that the genomic changes consistent with the chemotypic differences in foliar terpene profiles will be found in either transcription factors or their recognition sites.

However, in contrast to this pattern, TPS2 (sabinene hydrate synthase) was only present in the genome of those chemotypes that contained significant concentrations of terpinen-4-ol in the mature leaves. As TPS2 is absent from the genome from individuals that show no or little terpinen-4-ol in their foliar oils, it is possible that the gene may be linked to an allelic indel spanning at least a part of the C-terminal domain, but perhaps even the whole operon.

1,8-cineole is the monoterpene most consistently present in all *M. alternifolia* leaf oils. All cardinal chemotypes contain around 10% cineole and related metabolites (e.g. α-terpineol, limonene). Chemotype 5 (dominated by 1,8-cineole) is the only chemotype to which only monoterpenes derived from a single biosynthetic source contribute. TPS4 (1,8-cineole synthase) is not only expressed in the genomes of all chemotypes, but is also expressed at consistently high levels.

Other closely related species of *Melaleuca* share some of the chemotypes found in *M. alternifolia*. Both *M. linariifolia*, and *M. trichostachya* have a cineole-rich chemotype (Brophy *et al.* 1996). A 1,8-cineole-dominated chemotype appears to be a universal, or perhaps an ancestral chemotype in these species. Indeed, the ubiquity of Myrtaceous plants in Australia with high concentrations of cineole, irrespective of taxonomic and geographical boundaries (Brophy *et al.* 1996; Brophy *et al.* 2002), indicates that a selective pressure preserving the trait acts almost on a continental scale. The amino acid phylogeny of terpene synthases from Australian Myrtaceae shows that sequences from 1,8-cineole-rich species belong to a common clade (Keszei *et al.* 2010b), which shows the strongest correlation between shared chemistries and sequence. We argue that this indicates that the ancestral monoterpene synthase function for *Melaleuca* (as well as for *Eucalyptus*) was that of cineole biosynthesis.

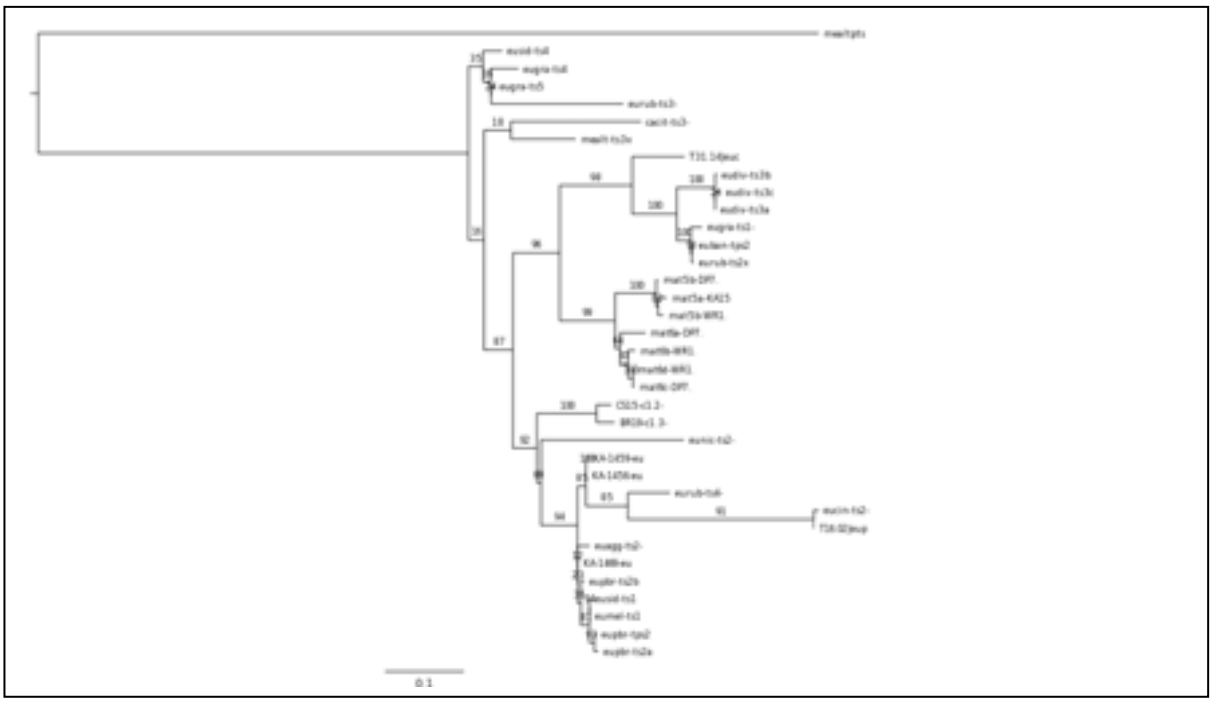

Figure 11: Maximum Likelihood phylogeny based on cDNA sequence of all monoterpene synthase fragments isolated. Branch labels are bootstrap values based on 1000 replicates. *M. alternifolia* putative isoprene synthase was used as an outgroup from within TPSb. TPS4a and TPS4b will need to be included in this phylogeny, as well as the *M. trichostachya* and *M. linariifolia* sequences.

The cineole synthase (TPS4) and sabinene hydrate synthase (TPS2) from *M. alternifolia* show much greater homology to each other than to terpinolene synthase (TPS5). Furthermore, while we have been able to identify multiple different sequences for terpinolene synthase, two sequence variants for cineole synthase, and four variants for an unknown monoterpene synthase, we have found only a single variant of sabinene hydrate synthase (TPS2). This suggests that TPS2 is the result neo-functionalisation of a duplicated cineole synthase gene, and represents a much more recent function than either cineole or terpinolene biosynthesis in *Melaleuca*.

Terpinolene biosynthesis is the dominant pathway in the leaves of seedlings of *M. alternifolia*, regardless of adult leaf oil chemotypes (Southwell *et al.* 2002). It is also significant during organ ontogeny; young leaf contains a higher proportion of terpinolene than older leaves on the same branch (Southwell *et al.* 1989). This together with the close phylogenetic relationship between TPS5 and Eucalyptus monoterpene synthases suggests that, the split coincides with a change of metabolism during ontogenetic development. There is insufficient data on the ontogenetic patterns of terpene secondary metabolites in other Myrtaceae to be able to say whether this particular pattern is unique or not to *M. alternifolia* but a similar chemical dichotomy between juvenile and mature leaves has been reported from the Myrtaceous genus *Leptospermum* (Brophy *et al.* 2000). The phylogenetic split between clades 1 and 3 of monoterpene synthase of Myrtaceae predates the split between *Melaleuca* and *Eucalyptus* (Keszei *et al.* 2010b). As TPS5 belongs to clade 1, we can assume that it is not associated with a recently developed function. Therefore, in *M. alternifolia*, it appears that the juvenile leaf chemistry has "resurfaced" as the adult terpinolene-rich chemotype 2. We propose that the terpinolene chemotype is a neoteny, resulting from a change in the regulation of an existing gene. The terpinolene chemotype is shared between the closely related species *M. alternifolia* and *M. trichostachya.* This suggests that that either the neoteny is not recent, or has occurred independently in these related species under similar environmental constraints. TPS5 and the phylogenetically similar TPS6 are the highest copy number monoterpene synthases (to our knowledge), with at least four different forms present in the genome of any individual irrespective of chemotype. This may indeed be the result of different copies persisting via temporal or spatial specialisation of function.

The Evolution of chemotypic variation in *Melaleuca*

Similarities between cineole synthase sequences from *M. alternifolia*, *M. trichostachya* and *M. linariifolia* can indicate one of three (or possibly more) scenarios

Scenario 1: Speciation between *M. alternifolia*, *M. trichostachya* and *M. linariifolia*. happened after the differentiation of TPS4 and TPS2. TPS4 from *M. linariifolia* is more similar to MaTPS4a (from chemotypes in the Southern distribution of *M. alternifolia*, towards the intergrade with *M. linariifolia*), and TPS4 from *M. trichostachya* is more similar to MaTPS4b (found in chemotypes from the NW of the distribution of *M. alternifolia*, where it occurs with *M. trichostachya*). This may indicate either hybridisation between the species, or very recent speciation compared to functional differentiation. TPS2 from *M. alternifolia* and *M. linariifolia* show common ancestry, indicating that the functional differentiation occurred before speciation.

Scenario 2: Cineole and sabinene hydrate synthases evolved from a common ancestor, but separately in the three species.

Scenario 3. Cineole and sabinene hydrate biosynthesis evolved separately after speciation, and within species variation has no parallels across species.

Conclusion

M. alternifolia shares cineole as a representative compound within its chemotypes, with closely related species, as well as with the majority of Australian Myrtaceae. An ancestral cineole synthase underwent several rounds of gene duplication, resulting in a diversity of monoterpene metabolites even before the present day genera of *Melaleuca* and *Eucalyptus* split from each other. A significant point of differentiation was the separation of metabolites based on ontogeny, with a separate suite of genes being expressed in young individuals and leaves than in mature plants. Genetic drift and subsequent gene duplications gave rise to new functions, however cineole biosynthesis persisted throughout Myrtaceae in the same clade, indicating that a locus corresponding to a set mode of expression remained advantageous. Cineole synthase has continued to contribute to new functions via duplication (such as for sabinene hydrate biosynthesis), but has also always been maintained as a copy retaining the original function. The biosynthesis of terpinolene and terpinen-4-ol therefore represent adaptations specific to *M. alternifolia* and its sister species, while cineole represents a broader scale adaptation to the Australian landscape.

Chapter 3 - The yield of essential oils in *Melaleuca alternifolia* is controlled by transcript abundance of genes in the MEP pathway and in *ippi*

Introduction

Three important considerations influence the profitability of essential oil crops. First, the profile of the oil must meet standards determined by customers and, where there is a medicinal use, regulatory authorities (Hammer *et al.* 2006). Secondly, the foliar concentration of oil in each plant must be high so that the marginal return on harvesting is also high. (Murtagh 1996). Finally, assuming that there is no tradeoff between growth of the plant and the foliar concentration of oil larger plants yield more oil per hectare. In *Melaleuca alternifolia* traditional breeding methods have led to significantly enhanced yields in a very short time and this is largely attributed to increasing foliar concentration (Doran *et al.* 2006).

There is significant naturally occurring variation in yield of terpenes in *M. alternifolia* (Butcher *et al.* 1994). Consistent with studies in other terpene-rich species, several lines of evidence argue that genetic factors are more important than environmental factors in determining oil yield per unit of tissue in *M. alternifolia*. For example, the narrow sense heritability of total oil concentration was estimated as 0.8 (Doran *et al.* 2006) and a breeding programme achieved an increase in foliar terpene concentration of between 32 and 54% in just three generations. Similarly, differences in oil yield between seed lots were consistent between years and between sites (Doran *et al.* 2002). This suggests that the differences in terpene yield between individuals are under strong genetic control and that environmental effects on yield are secondary to genetic effects.

Terpenes are synthesised from a series of enzymatic reactions in both the chloroplast and the cytosol. The last step of which involves the terpene synthase enzymes that were described in Chapter 2. However, it is the early steps in both the MEP and the MVA pathways that have the potential to limit the flow of substrates and hence the overall yield of terpenes. In other plant systems the early steps of the MEP pathway have been identified as bottlenecks for oil yield, e.g. peppermint and *Arabidopsis*. Over-expression of *dxr* in peppermint (*Mentha piperita*) led to plants accumulating 40% more oils in their glandular trichomes relative to the wild type (Wildung *et al.* 2005). Over-expression of *dxr* and *dxs* in *Arabidopsis thaliana* lead to increased accumulation of 350 and 650 % more taxadiene (a diterpene) than wild type plants, respectively (Carretero-Paulet *et al.* 2006) and over-expression of *dxs* in tomato, resulted in a 60% increase in isoprenoids (Enfissi *et al.* 2005). In grapes (*Vitis vinifera*) *dxs* co-localizes with a major QTL for the accumulation of three monoterpenes (linalool, nerol and geraniol) (Battilana *et al.* 2009). In glandular trichomes of basil (*Ocimum basilicum*) transcript and protein abundance as well as enzyme activity of *dxs* and *dxr* correlate with oil yield (Xie *et al.* 2008).

Identifying, the factors affecting yield is a substantially harder undertaking than identifying the basis of chemotypic variation. The work on M. *alternifolia* has been made possible by the identification of the genes involved in the terpene biosynthesis pathways in *E. globulus* and related eucalypts by (Kulheim *et al.* 2009). Interestingly, the highest yielding individuals of *M. alternifolia* occur in Chemotype 5, which is commercially undesirable. Therefore we restricted the study to commercially valuable chemotype 1 individuals to eliminate the need to account for qualitative variation as well as quantitative variation and to eliminate the need to compare yield between different chemotypes.

In the work described in this chapter, we have quantified both the transcripts from genes upstream of the terpene synthases in the terpenoid biosynthesis pathway, and the essential oils from the same leaves and found high degrees of correlation between transcript abundance and oil yield in genes from the MEP pathway. We further analysed single nucleotide polymorphisms and associated these with the oil phenotype. We conclude that much of the variation in oil yield can be explained by variation in transcript abundance and that the differences in gene expression are likely to be regulated by polymorphisms in the intron region of certain genes.

Methods and Materials

Plant Material

Samples from *Melaleuca alternifolia* plants for this study were collected from a New South Wales Department of Primary Industry (NSW DPI) experimental site at Ballina in Northern NSW (28.52. 00S; 153.34.00 W). The site contains plantings of more than 200 families from seed collected from 14 populations within the Clarence River catchment and one population from Port Macquarie. All populations contain predominantly chemotype 1 individuals in which the terpene profile is dominated by terpinen-4-ol (Butcher *et al.* 1994; Homer *et al.* 2000). One individual from each of 200 families was selected and samples of fully expanded foliage of ~1 year of age were removed for later extraction of DNA, RNA and terpenes. Samples for DNA and terpenes were refrigerated for the transport and then frozen in -80°C until further analysis. Samples for RNA extraction were immediately frozen in liquid nitrogen and also stored in -80°C.

Terpene extraction and analysis

Terpenes were extracted with ethanol containing an internal standard of tridecane according to the method described by Russell and Southwell (Russell *et al.* 2003) and analysed on GC-MS (Agilent, Santa Clara CA). For all individuals the major monoterpenes: 1,8-cineole, terpinen-4-ol, *cis*-sabinene hydrate, *trans*-sabinene hydrate and limonene were quantified, as well as the sesquiterpene bicyclogermacrene. For 32 samples, which were used for transcript quantification, all terpenes were quantified. A total of 47 terpene peaks were identified, 20 of which were monoterpenes and 27 sesquiterpenes. By comparing the mass spectra to reference mass spectral data in the National Institute of Standards and Technology mass spectral library, 29 terpenes could be identified.

RNA extraction and cDNA synthesis

Leaves were ground to a fine powder in liquid nitrogen. Total RNA was extracted from leaves with an RNeasy micro-extraction kit (QIAGEN, Valencia, CA) with the addition of sodium isoascorbate (Sigma-Aldrich, Sydney, Australia) to saturation and 20 µl of β-mercaptoethanol (Sigma-Aldrich, Sydney, Australia) to the extraction buffer to overcome the problems that arise with high concentrations of phenolics in the leaves. Samples for which this method was not successful had additionally 25 µl of polyvinylpyrrolidone (BDH chemicals, Westchester, PA) added to the extraction buffer. After the first wash step, DNase (Promega, Madison, WI) was added to the wash column and incubated for 30 minutes at room temperature. RNA concentrations were then quantified using a Nanodrop spectrometer (Thermo Scientific, Wilmington, DE). First strand cDNA synthesis was performed using the Moloney Murine Leukaemia Virus Reverse Transcriptase (Promega, Madison, WI), which was anchored with a $T_{30}VN$ primer according to the manufacturer. All samples were diluted to a final concentration of 15 $ng.\mu l^{-1}$ with MilliQ H_2O and used as template for real time PCR.

Real time PCR and analysis

Gene specific primers for the reference gene (*elongation factor 1α*), MEP pathway genes (*dxr, dxs* and, *dxs2*), mevalonate pathway (*mvk*) and terpenoid pathway (*ippi1, ippi2* and *gpps*) were designed. qRT-PCR was performed on a Roche Light Cycler 480 (Roche, Madison, WI) with the following PCR conditions; 96°C denaturing step for 6 min then 60 cycles of 94°C for15 s, 60°C for 15 s and 70°C for 12 s, followed by a melt curve. Three technical replicates were performed for each gene in each individual. Transcript abundance was calculated as $(1 + E)^{-\Delta Ct}$ and then normalized to *ef1α* ($\Delta C_t = \Delta Ct_{goi} / \Delta Ct_{ef1\alpha}$) according to the protocol by (Ramakers *et al.* 2003). The software package GenStat (GenStat version 12.1, VSN, International, Hemel Hempstead, UK) was used to perform simple linear regression between the relative expression of each gene and between gene expression and oil phenotype. To test the correlation of multiple transcript abundance to oil phenotype, a single trait was chosen (terpinen-4-ol) and two distinct methods of analysis were applied: first, model simplification (Hobbs *et al.* 2006) and second model averaging (Burnham *et al.* 2002).

DNA extraction, gene discovery, barcoding of individual samples and 454 sequencing on a GS-FLX platform

Genomic DNA was extracted from leaves using a Qiagen DNeasy 96 Plant kit (QIAGEN, Valencia, CA) using 100 mg of tissue from each individual. The DNA concentrations were measured using a Nanodrop spectrometer (Thermo Scientific, Wilmington, DE) and diluted to 25 ng / µl. Gene specific primers, which were designed for *Eucalyptus globulus* (Kulheim *et al.* 2009) were used to amplify gDNA from seven genes (*dxr, dxs1, dxs2, mvk, ippi1, ippi2* and *gpps*) from single random individuals. Amplicons were sequenced on an ABI 3130xl sequencer and gene specific primers for *M. alternifolia* were designed (Supplemental Table 2). Fragments for each of the seven genes were amplified for all of the 192 individuals, a sub-fraction was quantified on a gel and equimolar amounts from each fragment were pooled, resulting in 192 pools, each containing seven amplicons. Each pool was then quantified using a Qubit Quantification Platform (Invitrogen, Carlsbad CA), diluted to 10 ng / µl and then shredded by sonication using the Covaris S2 System (Covaris, Woburn, MA), using a duty cycle of 50%, intensity of 3,200 cycles per burst, run for 90 s for an average fragment length of 400 bp. Individual barcodes were ligated to the shredded DNA pools according to Meyer et al. (2008). The 192 individual barcodes are listed in Supplemental Table 3. Fifteen ng DNA per sample were pooled and were sequenced on a Roche Life Sciences GS-FLX according to standard procedures (454 Life Sciences, Branford, CT, USA). Four individuals were discarded because their terpene profile showed that they belonged to a different chemotype.

SNP identification and association

Genomic DNA was extracted from leaves using a Qiagen DNeasy 96 Plant kit (QIAGEN, Valencia, CA) using 100 mg of tissue from each individual. The DNA concentrations were measured using a Nanodrop spectrometer (Thermo Scientific, Wilmington, DE) and diluted to 25 ng / µl. Gene specific primers, which were designed for *Eucalyptus globulus* (Kulheim *et al.* 2009) were used to amplify gDNA from seven genes (*dxr, dxs1, dxs2, mvk, ippi1, ippi2* and *gpps*) from single random individuals. Amplicons were sequenced on an ABI 3130xl sequencer and gene specific primers for *M. alternifolia* were designed (Supplemental Table 1). Fragments for each of the seven genes were amplified for all of the 192 individuals, a sub-fraction was quantified on a gel and equimolar amounts from each fragment were pooled, resulting in 192 pools, each containing seven amplicons. Each pool was then quantified using a Qubit Quantification Platform (Invitrogen, Carlsbad CA), diluted to 10 ng / µl and then shredded by sonication using the Covaris S2 System (Covaris, Woburn, MA), using a duty cycle of 50%, intensity of 3,200 cycles per burst, run for 90 s for an average fragment length of 400 bp. Individual barcodes were ligated to the shredded DNA pools according to Meyer et al. (2008). The 192 individual barcodes are listed in Supplemental Table 2. Fifteen ng DNA per sample were pooled and were sequenced on a Roche Life Sciences GS-FLX according to standard procedures (454 Life

Sciences, Branford, CT, USA). Four individuals were discarded because their terpene profile showed that they belonged to a different chemotype.

SNP identification and association

We used CLC Genomics Workbench (CLC Bio, Aarhus, Denmark) to detect single nucleotide polymorphisms after reference assembly. The SNP discovery window was 7 bp, with central bp quality score of 40 and surrounding bp quality score of 25. In all 127 SNPs were identified of which 91 were selected to be genotyped in all individuals. The sequences were separated by barcode using Biopython version 1.52, not allowing for any mismatches. Of the 188 individuals 179 had sufficient reads for further analysis. The separated sequences were individually aligned to the reference sequence using CLC Genomics Workbench (CLC Bio, Denmark) and each SNP was scored when there were at least eight reads at a SNP position or five reads if all of them were of the same allele. A heterozygote allele was called when at least 30% of the reads were of the second allele.

TASSEL v2.1 Beta 2009 (Trait Analysis by Association Evolution and Linkage) (Bradbury *et al.* 2007) was used to perform analysis of linkage disequilibrium between the SNPs, as well as association to the oil phenotypes. A general linear model was used to test the association between individual SNPs and traits (total foliar oil, terpinen-4-ol, 1,8-cineole, limonene and bicyclogermacrene). The latter three were normalised due to non normal distribution of data. With the populations in our data set, adjustment for genetic structure was not possible, which was not necessary as all trees originate in the Clarence River catchment, except for the Port Macquarie population and are reported to have little genetic structure (Butcher *et al.* 1994; Rossetto *et al.* 1999). The inclusion or exclusion of the Port Macquarie population had no detectable effect on the associations.

Results

Quantitative and qualitative analysis of essential oils

For quantification of essential oil from chemotype 1 individuals of *M. alternifolia*, samples from 192 individuals from 14 populations were collected and their major terpenes quantified. Gas chromatography revealed 20 monoterpenes and 27 sesquiterpenes, of which 18 and 11, respectively, could be identified by comparison to reference mass spectral data in the National Institute of Standards and Technology mass spectral library. For the 32 individuals used in gene expression comparisons, all identified terpenes were quantified, while for 188 individuals only 1,8-cineole, limonene, bicyclogermacrene, terpinen-4-ol, *cis*-sabinene hydrate and *trans*-sabinene hydrate, as well as total oils were quantified. Four individuals were discarded, as they proved to be of a different chemotype.

Total terpene yield ranged from 39.4 -122.3 mg.g DM^{-1}, with an average of 73.2 mg.g DM^{-1}. Terpinen-4-ol, the major constituent, ranged from 14.5 – 68.5 mg.g DM^{-1} (average 31.6 mg.g DM^{-1}), 1,8-cineole ranged from 0.3 – 2 – 10.6 mg.g DM^{-1}, limonene from 0.3 – 0.6 – 1.2 mg.g DM^{-1} and the sesquiterpene bicyclogermacrene ranged from 0.7 – 2.7 – 6.7 mg.g DM^{-1}. Within the samples analysed, only the total oil and terpinen-4-ol were normal distributed (data not shown).

	dxr	*dxs1*	*dxs2*	*ippi1*	*ippi2*	*mvk*	*gpps*
dxr	-	0.755***	0.772***	0.58***	0.333***	0.334***	0
dxs1		-	0.468***	0.626***	0.332***	0.285**	0.01
dxs2			-	0.585***	0.409***	0.323***	0
ippi1				-	0.657***	0.528***	0.14*
ippi2					-	0.733***	0
mvk						-	0.121*
gpps							-

Table 4: Correlation between the expression of genes of terpene biosynthesis in *Melaleuca alternifolia*. * = P<0.05, ** = P<0.01, *** =P<0.001

Transcript abundance from genes in the terpenoid biosynthesis pathway

The relative abundance of transcripts for *dxr*, *dxs1* and *dxs2* from the MEP pathway, *mvk* from the MVA pathway, two copies of *ippi* as well as *gpps* was quantified. The comparison of transcript abundance between genes showed that there were high levels of correlation within as well as between pathways, except for *gpps*. The values for adjusted R^2 was highest between *dxr* and *dxs2* (0.772) followed by *dxr* and *dxs1* (0.755) and *ippi2* and *mvk* (0.733) (Table 4 and Figure 12). Two groups could be identified, the first with all genes from the MEP pathway plus *ippi1*, the second group with *mvk*, *ippi1* and *ippi2*. The R^2 values from a simple linear regression as well as their *P*-values are shown in Table 4.

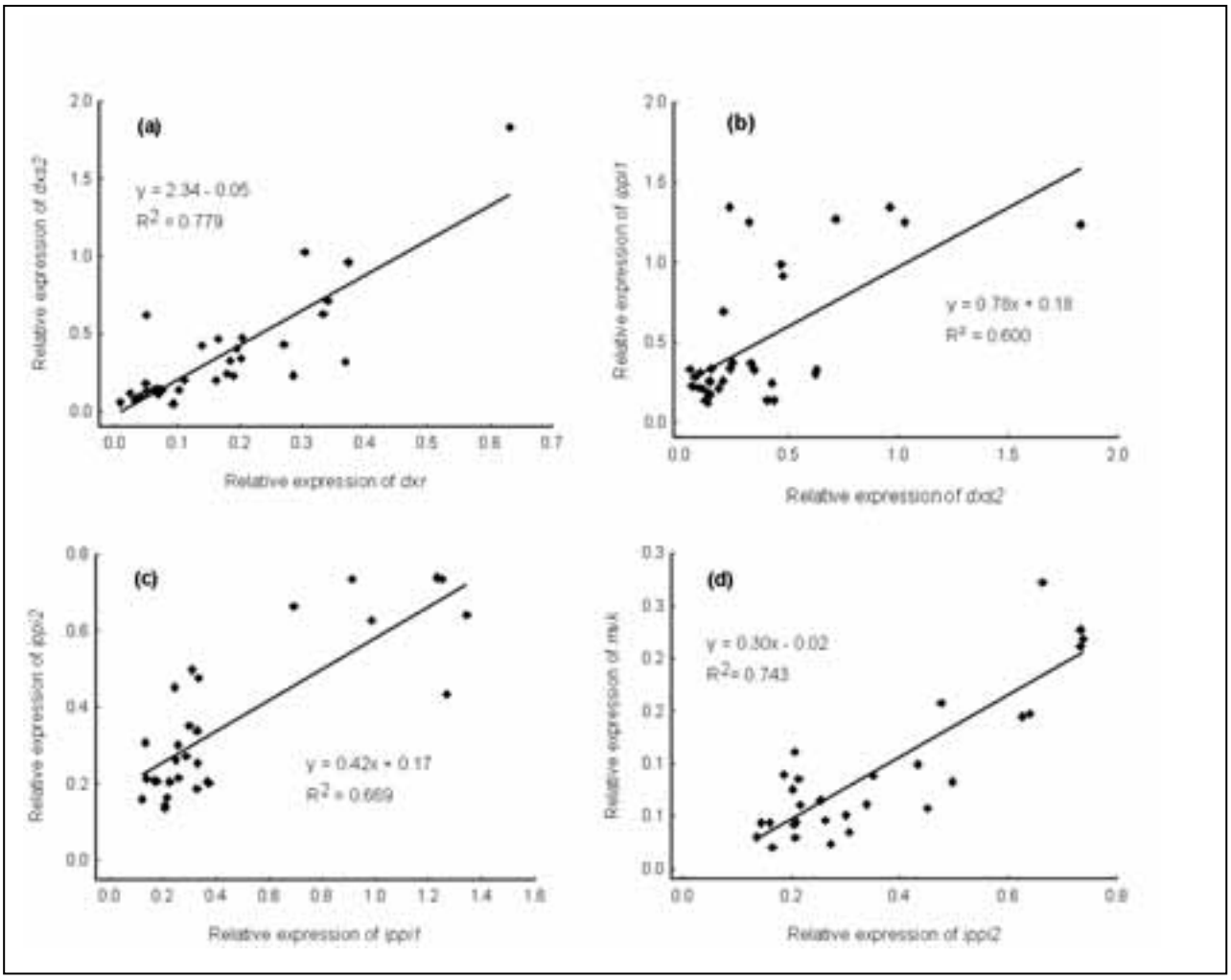

Figure 12. (a) correlation between the relative expression of *dxs2* and the relative expression of *dxr*, (b) correlation between the relative expression of *ippi1* and the relative expression of *dxs2*, (c) correlation between the relative expression of *ippi1* and the relative expression of *ippi2*, (d) correlation between the relative expression of *mvk* and the relative expression of *ippi2*.

Correlation of gene expression with quantitative variation of oil traits

The correlation between the relative expression of *dxr*, *dxs1*, *dxs2*, *ippi1*, *ippi2*, *mvk* and *gpps* was tested against a number of different oil traits using simple linear regression. Results from these regressions are shown in Table 5. Corresponding graphs from four of these correlations are shown in Figure 13. Individual CMA22 appears to be an outlier in all four graphs but in all cases the point fell

within the 95% confidence interval of normal quantile plots and was retained for analysis. Notably, the relative expression of *dxs1, dxs2* and *dxr* was correlated with the concentration of bicyclogermacrene, which is a sesquiterpene. There was no significant correlation between any of the genes studied and the foliar concentrations of 1,8-cineole, limonene, α-terpinolene, terpinolene, total sesquiterpene, aromadendrene, allo-aromadendrene, δ-cadinene, globulol, viridiflorol or any of the other terpenes quantified. There was no significant correlation between the expression of any of the genes and the ratio of total sesquiterpenes to total monoterpenes.

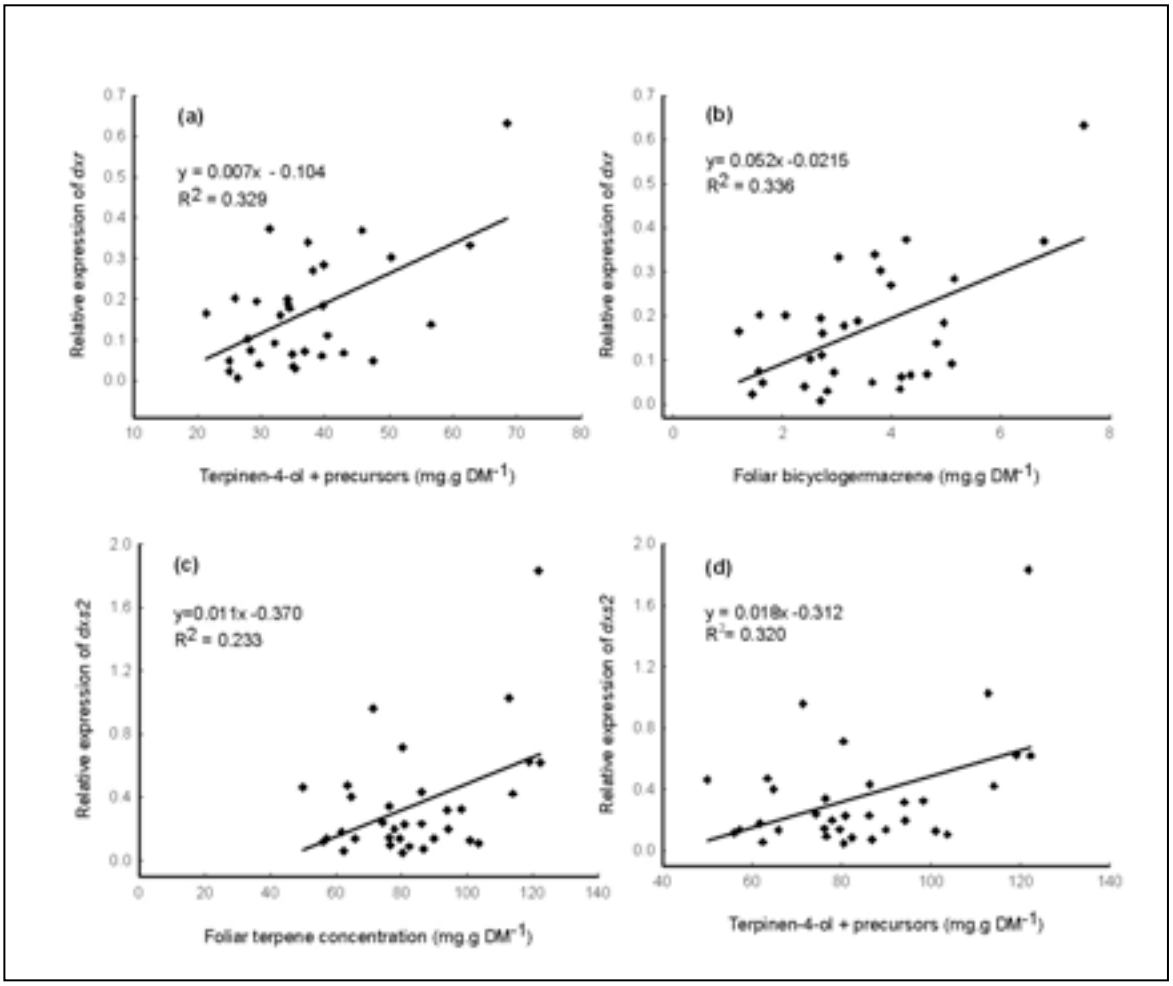

Figure 13: Correlation between (a) the relative expression of *dxr* and the concentration of terpinen-4-ol (plus precursor), b) the relative expression of *dxr* and the concentration of bicyclogermacrene (c) the relative expression of *dxs2* and the concentration of total terpenes, (d) the relative expression of *dxs2* and the concentration of terpinen-4-ol (plus precursors)

Since the transcript abundance of all investigated genes, except for *gpps* were strongly correlated, we aimed to test the relationship between multiple genes' transcript abundance and the phenotype, the quantitative variation in terpenes. Two different methods were applied, first the established method of model simplification and second model averaging (Burnham *et al.* 2002). For both approaches, individual CMA22 was excluded as an outlier. The expression of *ippi1* was significantly correlated to terpinen-4-ol as a quadratic and was included as $ippi1^2$. The relationship between *ippi1* and terpinen-4-ol (plus precursors) was complex, there appeared to be no relationship in low-yielding trees, but a strong quadratic relationship in high-yielding trees. With eight variables (the expression of the seven genes plus $ippi1^2$), there were 28 interaction terms, meaning there were 36 variables in total. This is greater than the number of data points in the study, which limits explanatory power since there are

effectively no degrees of freedom. Therefore, six different models were tested, with the original eight variables and four different interaction variables in each model. This analysis identified four significant interaction terms, *dxs1: ippi1, dxs1: ippi1*^{2}*, gpps: ippi1 and ippi1: ippi1*^{2}. The full model used the four interaction variables plus the eight original variables. Model simplification techniques were then used until the final model contained only significant variables. The resulting model explained 0.568 of variation in terpinen-4-ol (plus precursors) content at a significance level of $P < 0.0001$. This model consisted of *dxs1, dxs2, ippi1, ippi1*^{2} and *dxs1: ippi1*^{2}. Model averaging was performed using the four interaction terms described above and the seven original variables plus *ippi1*^{2} as variables. We tested 4095 models, of which 596 were significant at the $P < 0.05$ level. The resulting Akaike weights for each individual variable, which are the summed Akaike weights for each model the variable appears in, describes the relative importance of individual parameters, are shown in table 3. The gene with the highest Akaike weight was *mvk*, although it was not significant in any of the models, followed by *ippi1*^{2}, *dxs2* and *ippi1*.

Explanatory variable	Akaike weights
mvk	0.9999
dxs2	0.9982
ippi1^2	0.9907
ippi1	0.9834
dxr	0.9067
dxs1	0.4892
dxs1: ippi1^2	0.4347
gpps	0.3961
dxs1:ipp1	0.3298

Table 5: Akaike importance weights for individual variables, obtained through model averaging.

454 Sequencing and SNP discovery

For discovery and SNP genotyping in a single step, a method was used which was developed by Meyer and co-workers and includes the ligation of a short 10 nucleotide long 'barcode' to the DNA of each individual (Meyer *et al.* 2008). Seven loci were amplified from each of 188 individuals, mechanically shredded and the barcode ligated. After pooling equimolar amounts from each individual, we ran the DNA on one quarter of a GS-FLX sequencing plate. The sequencing run yielded 131,241 reads with an average length of 297.8 bp, totalling 39.1 Mbp. SNPs were first identified by aligning all reads to the reference sequence and then using the CLC Genomics Workbench software for the identification of SNPs. This could potentially lead to bias from

overrepresentation of individuals, but we felt confident with this method as the range of reads per position was typically 2-3 fold between individuals and our selection criteria were stringent. A polymorphism had to occur in 5% of the reads and be represented by a minimum of 500 reads. A few SNPs that occurred at <5% were selected by hand.

The reads were then divided into 188 pools according to their barcode using very stringent selection that did not allow for a single mismatch. Therefore, approximately one third of the reads was discarded (30,420). A less stringent selection could have been used, but the number of reads that had an intact barcode was sufficient for this study. The reads were then aligned to the reference sequences of the seven loci and the summary statistics of this can be found in Table 6.

Gene	Length (bp)	Reads	Total bases (kbp)	Average coverage per position	Average coverage per position & individual
dxr	1759	13840	4192.9	2383.7	12.7
dxs1	1576	25486	7721.2	4899.2	26.1
dxs2	684	32235	9765.9	14277.6	75.9
gpps	1303	12246	3710.0	2847.3	15.1
mvk	1573	17595	5330.5	3388.7	18.0
ippi1	1683	5146	1559.0	926.3	4.9
ippi2	2418	8933	2706.3	1119.2	5.9

Table 6: Summary statistics from reference alignment of sequence data for seven genes from terpene biosynthetic pathways in *Melaleuca alternifolia.*

The most important parameter shown is the average number of reads per position and individual. Of the 188 individuals, 5 had to be excluded as they had insufficient numbers of reads. In order to identify both alleles of a diploid DNA source, a certain threshold of coverage is required. We felt comfortable in identifying both alleles when eight or more reads were present at a given position, but also maintained several at six or seven reads (after individual examination). Interestingly, all data that could be acquired for *ippi1* was from the 3' end, therefore most SNPs at that region could be genotyped. The coverage of *ippi2* was too patchy and the locus had to be excluded for further analysis. 127 SNPs were identified, of which 91 were eventually used in the association study.

Linkage disequilibrium within *dxs1, dxs2, dxr, mvk, gpps* and *ippi1* was analysed. The linkage between SNPs in *dxs1, dxs2, dxr* and *ippi1* decayed quickly with r^2 decaying to less than 0.3 within 0.5 kb with a few exceptions. In *dxr,* SNPs at positions 325 and 326 were linked with SNPs at positions 1647 and 1684. The correlation between these SNPs was r^2 >0.6. Linkage between SNPs within *mvk* and *gpps* was slightly stronger, with r^2 > 0.4 at 1kb and 0.7 kb, respectively. There was

some evidence of linkage between two genes with several SNPs from *ippi1* and *gpps* being correlated at $r^2 > 0.4$.

Association study

Association analysis was performed using the general linear model (GLM) implemented in TASSEL (Bradbury *et al.* 2007). A total of 91 SNPs were tested against five different oil traits; total oil, terpinen-4-ol (plus precursors), normalised 1,8-cineole, normalized limonene and normalized bicyclogermacrene concentrations. This analysis produced 37 significant associations at the $P < 0.05$ level of significance. Of these eight were significant at the $P < 0.01$ level of significance. Of the SNPs identified as being significantly correlated with oil yield, up to 8% of the variation in oil yield can be explained by a single nucleotide polymorphism. The specific SNPs identified are being withheld from this publication and the results from this part of the project will only be discussed broadly.

Discussion

Quantitative variation in terpene yield in plants is widespread and of significant industrial importance. Terpene yield is under strong genetic control (e.g. Doran *et al.* 2006; Hanover 1966a) and although the genes involved in the two biosynthetic pathways leading to the formation of terpenes (MEP and MVA) are well-known in crops and model species such as *Arabidopsis*, very little is known about genetic basis of quantitative variation of terpene traits in wild plant populations. In this study we investigated the genetic control of terpene yield from the Australian Myrtaceae *Melaleuca alternifolia*. Fourteen populations of the economically important chemotype 1 were tested for correlation between terpene traits and transcript abundance of seven genes in the terpenoid biosynthesis and 99 SNPs were genotyped and tested for association with terpenoid traits. We successfully demonstrate that 'barcoding' (the ligation of a small, unique nucleic acid tag to an individuals' DNA), is a useful tool for large numbers of individuals and can be utilized in population genetics.

The first aim of the study was to identify correlations between the transcript abundance of candidate genes and terpene traits. When analysing and comparing the gene's transcript abundance to each other, we noticed that the level of co-regulation within the pathways is very strong. The highest levels of co-regulation could be found within the MEP pathway, as well as between *mvk* and both copies of *ippi*. Of the seven investigated genes only *gpps* did not correlate to any of the other genes.

Several studies have shown that the expression of *dxr* and *dxs* has a strong influence on oil yield. Our results show that a large proportion of the quantitative variation of terpenes in *M. alternifolia* can be explained by quantitative variation of transcript abundance of single genes in the early stages of terpenoid biosynthesis. Transcripts from two genes, *dxr* and *dxs2* showed the highest correlation to total oil yield with R^2 of 0.317 and 0.179, respectively. They also had the highest correlation with the most abundant monoterpene terpinen-4-ol with R^2 of 0.304 and 0.298, respectively. The correlation between monoterpenes and transcripts from the MEP pathway was expected as they both occur in the chloroplast, however we also found a significant correlation between *dxr, dxs1, dxs2* and bicyclogermacrene, a sesquiterpene, which is synthesized in the cytosol. This correlation strongly suggests that in *M. alternifolia* the substrate for farnesyl pyrophosphate synthase, isopentyl pyrophosphate, is transported out of the chloroplast and utilized for sesquiterpene synthases in addition to IPP produced through the MVA pathway in the cytosol. This transport has been shown before (Bick and Lange 2003), though the mechanism has not yet been described.

In an approach to combine the effect of single genes on terpene yield, we applied two forms of multiple linear regression: model simplification and model averaging. Both approaches lead to similar results, in model simplification *dxs1, dxs2, ippi1* and $ippi1^2$, were the variables that could best explain the quantitative variation in terpinen-4-ol, while with model averaging *dxs2, dxr, ippi1,* $ippi1^2$ and *mvk*

were the variables which occurred most in the significant models that were tested. The combined correlation of variables to terpene yield was approximately twice as large as that of single variables with R^2 reaching 0.568. This clearly shows that the terpene yield is determined by the combined regulation of multiple genes. The correlation between *ippi1* and terpinen-4-ol showed a strong quadratic relation in high-yielding trees but no correlation in low-yielding trees; this suggests that in low-yielding trees the terpene content is under a somewhat different control than in high-yielding trees.

We hypothesized that *ippi* would be correlated to the ratio of monoterpenes to sesquiterpenes, since the ratio of IPP to DMAPP is important in determining which type of terpene is predominantly produced. However, this hypothesis could not be accepted from our data and for this species. We suggest instead that the flux of IPP from the chloroplast to the cytosol contributes to the ratio of monoterpenes to sesquiterpenes. This is supported by our data, showing a strong correlation between the expression of MEP pathway genes and the most abundant sesquiterpene, bicyclogermacrene. Our data does further not support the idea of specialized chloroplast and cytosol forms of *ippi*, but rather that both copies function in both the cytosol and chloroplast. It appears that *ippi1* appears to be more strongly associated with the chloroplast, as its expression is most closely correlated with that of the MEP pathway genes, while *ippi2* may be closer associated with the cytosol as its expression is closest correlated to *mvk* from the MVA pathway. Further testing that includes molecular labelling and / or deletion mutants of each gene are needed to test these hypothesis.

Previously, it has been labour-intensive and costly to identify single nucleotide polymorphisms (SNPs) in non-model species and to genotype large numbers of individuals. Typically, to identify SNPs, a few candidate genes were sequenced by traditional Sanger sequencing from a number of individuals that cover the expected range of genetic variation. This was then followed by expensive SNP genotyping. In this study we combined these two steps into one following the approach of Meyer and co-workers (Meyer *et al.* 2008), making it faster, easier, cheaper and more reliable than previous techniques.

It is widely believed that linkage disequilibrium (LD) is very low in forest trees, with strong linkage decaying to an $R^2 < 0.3$ within 1 to 2 kb in most species (Neale 2007; Neale *et al.* 2004; Savolainen *et al.* 2007; Thornsberry *et al.* 2001; Thumma *et al.* 2005). The low levels of LD within forest trees is thought to be due to high levels of out-crossing within these populations, large effective population sizes and a long history of recombination (Brown *et al.* 2004; Ingvarsson 2005; Nordborg 2000). The LD was tested in the seven candidate genes and mostly the LD decayed below $R^2 < 0.3$ within 1 kb, we found however a number of exceptions where linkage was larger than $R^2 > 0.6$ over more than 1.3 kb (*dxr)*, and $R^2 > 0.4$ over 1 kb and 0.8 kb in *mvk* and *gpps*, respectively. Despite these instances, LD is still low when compared to inbreeding species such as *Arabidopsis* where linkage commonly extends to about 10 kb and can extend to over 250 kb (Kim *et al.* 2007; Nordborg 2000). However, the findings from this study suggest that LD may not decay as fast or uniformly in Myrtaceae as is suggested by the published literature (Neale 2007; Savolainen *et al.* 2007; Thumma *et al.* 2005). Studies that extend over several kb could shine light on this question and the tools explored in this study can easily be used for further studies on variation and absolute LD in forest trees.

It is important to be aware of the levels of LD when attempting association studies. High levels of LD make association of functional variants difficult. Generally a reverse genetic approach, such as looking at knockout mutants and over-expression of genes in transgenic plants is used to determine the role of specific genes in plants. The low levels of LD in forest trees mean fine scale mapping of functional variants relatively easy, opening up the possibility of being able to detect more subtle effects of particular alleles in wild populations. We tested the association between SNPs and terpene traits to test two hypothesis; Firstly, whether a non-synonymous SNP could change the catalytic properties of the enzyme such as in the fast-slow mutation from the *Drosophila melanogaster* alcohol dehydrogenase (*adh)* (Choudhary *et al.* 1991) or second whether SNPs in the gene of interest can have an affect on transcriptional regulation, such as several SNPs in the first intron and the 3’

untranslated region from the *D. melanogaster adh.* In our study we could not find any non-synonymous SNP that significantly associated with terpene yield and therefore have to discard hypothesis 1 for this study. Several SNPs, located in known regulatory regions such as the 3' untranslated region and the first intron were found that associated with either total oil yield or terpinen-4-ol yield. SNPs in the 3' UTR have been shown to affect the stability of mRNAs, to be involved in successful mRNA maturation, as well as affecting gene expression by influencing post-transcriptional controls. The first intron has been shown to regulate gene expression in a number of other genes in other plant species and nucleotide variation in this region has been shown to have effects on gene expression. It is a feasible assumption that the SNPs identified play a regulatory role in the control of oil yield in *M. alternifolia*.

Regulation of gene expression functions on many different levels, such as enhancement or repression by transcription factors, typically in the 5' region of the gene or the first intron (e.g. review by Wray *et al.* 2003), degradation of transcripts, post-transcriptional modifications and translational modifications. It is therefore often difficult to discern which factor(s) is the most important one in a system of interest. Prior to the publication and analysis of the *Eucalyptus* genome, it was very difficult to include steps such as transcription factors and the choice at hand is too look at expressed genes for which ESTs have been sequenced. While the most important factor(s) may still be elusive for our genes of interest, we have discovered some polymorphisms that play a role in the regulation of these genes.

While the analytical power of an association study with less than 200 individuals is not as large as would be desirable and so false positives can occur, we are confident that most of the SNPs identified to be associated with terpene traits are real. Many of the SNPs are associated with more than one trait and the chance of several false positive occurrences of the same SNP are thus much lower than a single occurrence. Clearly it is important to validate these SNPs in a greater number of families and this should be the next step in this work.

Conclusions

Melaleuca alternifolia is an ideal study object for the control of terpene yield. Our collection of 190 individuals showed a three fold variation in total terpene yield and more than four fold variation of the major constituent terpinen-4-ol. In order to identify the regulatory mechanisms of this huge variation, we quantified transcripts from genes which are known to be involved in the regulation of terpene yield from other plants. Our data shows that the expression of *dxr*, *dxs1*, *dxs2*, and *gpps* is positively correlated to terpinen-4-ol yield and that the expression of multiple genes from the terpenoid biosynthetic pathway explain about twice as much of the variation of traits than single genes do. Approximately 100 SNPs were genotyped in the 188 individuals and a small number of SNPs appear to play a regulatory role in these genes.

Implications

The essential oil of *Melaleuca alternifolia* has been studied many times previously but limited statistical analysis and a poor understanding of the biosynthesis of terpenes has limited the utility of these studies. In the work described here a more thorough examination of the correlations between the different chemical constituents has revealed patterns that had not been previously appreciated. This showed that as few as three genes were responsible for most of the monoterpenes that give tea tree oil its character.

Once these three genes were isolated and characterized it was clear that the gene products explained the occurrence of the vast majority of the foliar monoterpenes. However, there was one monoterpene, d-limonene, that we expected to see as a gene product but was not observed in the products of any of the three genes discovered. In particular we expected limonene to be a product of the cineole synthase gene. Limonene has been a problematic for the tea tree industry and it is widely seen as undesirable product that can attract attention of regulators and so inhibit market access when it occurs in refined oil. The fact that limonene was not a product of the dominant terpene synthases in *M. alternifolia* suggests that it is produced by another as yet unknown but minor terpene synthase gene. This is exciting because it means that there is potentially the opportunity to identify its source and then select against it and so largely eliminate limonene from Australian tea tree oils. We believe that identifying the source of limonene would allow immediate steps to be taken to reduce its occurrence in Australia tea tree oils and provide an immediate benefit to industry.

Recommendation: That future work identify the gene producing limonene in tea tree oil with a view to screening the breeding population to select plants in which this gene is poorly expressed.

The oil component that is most valuable is terpinen-4-ol which is formed by the re-arrangement of the product of a specific gene discovered in this work called sabinene hydrate synthase. Sabinene hydrate re-arrangesspontaneously to form terpinen-4-ol. Sabinene hydrate synthase is a recent addition to the terpene synthases of *M. alternifolia* and has arisen as a result of a recent gene duplication and subsequent mutation of an existing cineole synthase. We found only 21 amino acid differences between these two genes. Thus the genetic basis of the whole tea tree industry resides in these 21 amino acid differences! No other primary industry is so reliant on such a small genetic change to produce its product. The difference between profit and loss in the industry is thus very small. The industry needs a plan to conserve its genetic resources since the sabinene hydrate synthase gene is absent from the genome of *M. alternifolia* over most of its range and only present in the genome of Chemotype 1 individuals (and the less valuable intermediate chemotypes) from a restricted range in NE NSW. We do not know whether there is significant differences amongst sabinene hydrate synthase genes in *M. alternifolia* but such variation would not be unexpected (e.g. the variation in cineole synthase genes is extensive and allowed us to identify TPS4b). Variation in sabinene hydrate synthase could involve genes with optimal product profiles. At present the breeding programme is focused mainly on selecting for high yield and there is a risk that other important sources of genetic variation will be lost.

Recommendation: The tea tree industry should ensure that it has a plan to conserve its genetic resources to ensure that variation in Chemotype 1 (other than variation solely in yield) is captured.

Different variants of sabinene hydrate synthase may produce different product profiles (e.g. greater amounts of sabinene hydrates and smaller amounts of terpinene than the example that we report in Chapter 3). These different product profiles may allow growers to make finer adjustments to their

products for particular markets or to breed a better quality oil overall. Sabinene hydrate synthases may also differ in the ratio of different stereochemical forms. S and R, as well as E and Z stereochemistry may affect the medicinal efficiency of tea tree oil, and would also have an effect on the properties of polymers used for electronics obtained from the oil – an application currently under testing. In addition there are likely to be differences in the efficiency of different forms of the enzyme in converting the substrates into the desired commercial product. These differences would be likely to be reflected in differences in yield. We did not consider this possibility in the current project because the characterization of enzyme kinetics is difficult and time consuming. However we believe that this possibility needs to be considered and collaboration with the laboratories in Germany on terpene synthase characterization is essential to continue this part of the work.

Recommendation: Variation in sabinene hydrate synthase should be investigated as a means of optimising the profile of oil.

The genes identified in this study, together with the variations in genomic occurrence and expression profiles allowed the development of a set of diagnostic tools that permit the different chemotypes to be separated. This can be done from very small amounts of material without the need to grow the material beyond the cotyledon stage. These diagnostic tests are performed by a PCR reaction on DNA extracted from the tissue. They are fast to conduct and would cost about $1.20-1.50 (including DNA extraction) each in an adequately equipped laboratory (excluding labour costs). Being certain about the chemotype that is planted means that selection and production of seed orchards can be done faster and more assuredly than waiting until the plants grow sufficiently for gas chromatographic analysis of oils in the leaves. The test is robust and can be applied immediately.

Recommendation: Tools that allow the different chemotypes to be determined at an early stage have been developed and can be applied now at a low cost

The identification of gene variants associated with high yield of terpenes was very successful. The approach taken was a candidate gene approach in which we used prior knowledge from other plants to identify a smaller subset of genes that could be examined in detail. The positive relationships between the level of expression of these genes and yield of oils indicated the success of this approach. Nonetheless, it is likely that other elements of other genes (e.g. transcription factors) that were not studied here may also be significant influences on yield. A number of Single Nucleotide Polymorphisms (SNPs) were identified which associated significantly with yield. However, before these can be used to screen for high-yielding trees, they need to be validated in a larger number of families. This will increase the statistical power used to detect significant associations as well as ensuring that these polymorphisms are widespread.

Recommendation: The Single Nucleotide Polymorphisms identified, as being associated with high yield must be validated in a larger number of families before they can be used as a diagnostic tool.

We believe that this project provided exceptional value for money for RIRDC and the tea tree oil industry. As well as direct development of diagnostic tools for enhancing quality and yield, the work developed a series of new methods that can be used to improve other non-traditional crops as well as supporting and training students who have the skills and enthusiasm to contribute to the industry in the future.

Whether these benefits can be realized more widely depends on greater investment in genomics in Australia. There seems to be little interest or understanding of the benefits that genomics brings to

Australian primary industries outside of medicine. For example, the *Eucalyptus* genome (specifically *Eucalyptus grandis*) has been sequenced by a group in the United States Department of Energy. It has been provided to Australia entirely free of charge even though Australia made no contribution towards doing the work.

Nonetheless there appears to be little awareness of the resource in Australia and little interest or within private or public agencies in Australia to take advantage of this opportunity. The benefits are immense and range from insights into improving wood quality, improving responsiveness of forestry to climate change and enhancing conservation of rare native plant communities. The large degree of synteny between the genomes of *Melaleuca* and *Eucalyptus* means that much of the data on *Eucalyptus* can be used directly in *M. alternifolia*. The direct benefits for a small industry like tea tree oil, is apparent from the work described here. It is somewhat ironic that the tea tree industry (which is one of the smallest sectors relying on *Eucalyptus* and its relatives) is now a leader in molecular breeding in Australia while the much larger forestry sector lags well behind! The impetus that the first tree genome, that of *Populus,* gave to bioenergy crops in the USA and Europe was immense but it appears that the benefits of the *Eucalyptus* genome will be captured mostly outside Australia, in particular by Brazil, Chile and South Africa.

Recommendation: RIRDC as one of the few agencies that has shown the vision to engage with native plant genomics could play a significant role by sponsoring a workshop in which the basics of genomics and the benefits of the* Eucalyptus *genome are explained for other Australian government agencies and private industry.

One issue that remains unresolved is whether the findings described here should or can be protected from exploitation by tea tree growers outside Australia. From a researcher's perspective, funding in Australian science generally is linked to publications and without those, it is difficult to maintain the skilled personnel necessary to maintain a prolonged presence in research. However, the most likely way that these findings could be protected is via non-disclosure rather than patenting and so publication becomes problematic.

References

Allen BE, Banthorpe DV (1981) Partial purification and properties of prenyltransferase from *Pisum sativum*. *Phytochemistry* **20**, 35-40.

Andrew RL, Peakall R, Wallis IR, Wood JT, Knight EJ, Foley WJ (2005) Marker-based quantitative genetics in the wild?: The heritability and genetic correlation of chemical defenses in *Eucalyptus*. *Genetics* **171**, 1989-1998.

Baker GR, Lowe RF, Southwell IA (2000) Comparison of oil recovered from tea tree leaf by ethanol extraction and steam distillation. *Journal of Agricultural and Food Chemistry* **48**, 4041-4043.

Barnes D, Franklin S, Schultz J, Henry R, Brown E, Coragliotti A, Mayfield SP (2005) Contribution of 5 '- and 3 '-untranslated regions of plastid mRNAs to the expression of *Chlamydomonas reinhardtii* chloroplast genes. *Molecular Genetics and Genomics* **274**, 625-636.

Barton AFM, Cotterill PP, Brooker MIH (1991) Short Note - Heritability of cineole yield in *Eucalyptus kochii*. *Silvae Genetica* **40**, 37-38.

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11-21.

Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando MS (2009) The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theoretical and Applied Genetics* **118**, 653-669.

Bertea CM, Schalk M, Karp F, Maffei M, Croteau R (2001) Demonstration that menthofuran synthase of mint (*Mentha*) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Archives of Biochemistry and Biophysics* **390**, 279-286.

Blanc VM, Pichersky E (1995) Nucleotide sequence of a *Clarkia breweri* cDNA clone of ipi1, a gene encoding isopentenyl pyrophosphate isomerase. *Plant Physiology* **108**, 855-856.

Bohlmann J, Phillips M, Ramachandiran V, Katoh S, Croteau R (1999) cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the Tpsd gene family from grand fir (*Abies grandis*). *Archives of Biochemistry and Biophysics* **368**, 232-243.

Boland DJ, Brophy JJ, House APN (1991) '*Eucalyptus* leaf oils: use, chemistry, distillation and marketing.' (Inkata Press: Melbourne)

Bouvier F, Suire C, dHarlingue A, Backhaus RA, Camara B (2000) Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *The Plant Journal* **24**, 241-252.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Brophy JJ, Davies NW, Southwell IA, Stiff IA, Williams LR (1989) Gas chromatographic quality control for oil of *Melaleuca* terpinen-4-ol type (Australian tea tree). *Journal of Agricultural and Food Chemistry* **37**, 1330-1335.

Brophy JJ, Doran J (1996) 'Essential oils of tropical *Asteromyrtus*, *Callistemon* and *Melaleuca* species.' (Australian Centre for International Agricultural Research: Canberra)

Brophy JJ, Goldsack RJ, Bean AR, Forster PI, Lepschi BJ (2000) Leaf essential oils of the genus *Leptospermum* (Myrtaceae) in eastern Australia. Part 6. *Leptospermum polygalifolium* and allies. *Flavour and Fragrance Journal* **15**, 271-277.

Brophy JJ, Southwell IA (2002) *Eucalyptus* chemistry. In '*Eucalyptus*: The Genus *Eucalyptus*'. (Ed. J Coppen) pp. 102–160. (Taylor & Francis: London and New York.)

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences* **101**, 15255-15260.

Burnham KP, Anderson DR (2002) 'Model selection and multimodel inference: a practical information-theoretic approach. 2nd Edition. .' (Springer-Verlag, : New York)

Butcher PA, Bell JC, Moran GF (1992) Patterns of genetic diversity and nature of the breeding system in *Melaleuca alternifolia* (Myrtaceae). *Australian Journal of Botany* **40**, 365-375.

Butcher PA, Byrne M, Moran GF (1995) Variation within and among the chloroplast genomes of *Melaleuca alternifolia* and *Melaleuca linariifolia* (Myrtaceae). *Plant Systematics and Evolution* **194**, 69-81.

Butcher PA, Doran JC, Slee MU (1994) Intraspecific variation in leaf oils of *Melaleuca alternifolia* (Myrtaceae). *Biochemical Systematics and Ecology* **22**, 419-430.

Carretero-Paulet L, Ahumada I, Cunillera N, Rodriguez-Concepcion M, Ferrer A, Boronat A, Campos N (2002) Expression and molecular analysis of the *Arabidopsis* DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Plant Physiology* **129**, 1581-1591.

Carretero-Paulet L, Cairó A, Botella-Pavía P, Besumbes O, Campos N, Boronat A, Rodríguez-Concepción M (2006) Enhanced flux through the methylerythritol 4-phosphate pathway in *Arabidopsis* plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Molecular Biology* **62**, 683-695.

Chan MT, Yu SM (1998) The 3 ' untranslated region of a rice alpha-amylase gene mediates sugar-dependent abundance of mRNA. *Plant Journal* **15**, 685-695.

Chen F, Ro D-K, Petri J, Gershenzon J, Bohlmann J, Pichersky E, Tholl D (2004) Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant Physiology* **135**, 1956-1966.

Choudhary M, Laurie CC (1991) Use of In vitro mutagenesis to analyze the molecular basis of the difference in ADH expression associated with the allozyme polymorphism in *Drosophila melanogaster*. *Genetics* **129**, 481-488.

Cornwell CP, Leach DN, Wyllie SG (1995) Incorporation of oxygen-18 into terpinen-4-ol from the H218O steam distillates of *Melaleuca alternifolia* (tea tree). *Journal of Essential Oil Research* **7**, 613-620.

Costa JH, de Melo DF, Gouveia Z, Cardoso HG, Peixe A, Arnholdt-Schmitt B (2009) The alternative oxidase family of *Vitis vinifera* reveals an attractive model to study the importance of genomic design. *Physiologia Plantarum* **137**, 553-565.

Cox SD, Mann CM, Markham JL (2001) Interactions between components of the essential oil of *Melaleuca alternifolia*. *Journal of Applied Microbiology* **91**, 492-497.

Croteau RB, Davis EM, Ringer KL, Wildung MR (2005) (-)-Menthol biosynthesis and molecular genetics. *Naturwissenschaften* **92**, 562-577.

Cunningham FX, Gantt E (2000) Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in *Escherichia coli*. *Plant and Cell Physiology* **41**, 119-123.

Decker CJ, Parker P (1995) Diversity of cytoplasmic functions for the 3' untranslated region of Eukaryotic transcripts. *Current Opinion in Cell Biology* **7**, 386-392.

Delfine S, Loreto F, Pinelli P, Tognetti R, Alvino A (2005) Isoprenoids content and photosynthetic limitations in rosemary and spearmint plants under water stress. *Agriculture Ecosystems & Environment* **106**, 243-252.

Deshpande A, Pastore A, Deshpande AJ, Zimmermann Y, Hutter G, Weinkauf M, Buske C, Hiddemann W, Dreyling M (2009) 3 ' UTR mediated regulation of the cyclin D1 proto-oncogene. *Cell Cycle* **8**, 3592-3600.

Doran JC, Baker GR, Williams ER, Southwell IA (2002) Improving Australian tea tree through selection and breeding (1996-2001). *RIRDC Publication 02/017, Canberra*.

Doran JC, Baker GR, Williams ER, Southwell IA (2006) Genetic gains in oil yields after nine years of breeding *Melaleuca alternifolia* (Myrtaceae). *Australian Journal of Experimental Agriculture* **46**, 1521-1527.

Doran JC, Matheson AC (1994) Genetic parameters and expected gains from selection for monoterpene yields in Petford *Eucalyptus camaldulensis. New Forests* **8**, 155-167.

Edwards PB, Wanjura WJ, Brown WV (1993) Selective herbivory by Christmas Beetles in response to intraspecific variation in *Eucalyptus* terpenoids. *Oecologia* **95**, 551-557.

Eibl C, Zou ZR, Beck A, Kim M, Mullet J, Koop HU (1999) In vivo analysis of plastid psbA, rbcL and rpl32 UTR elements by chloroplast transformation: tobacco plastid gene expression is controlled by modulation of transcript levels and translation efficiency. *Plant Journal* **19**, 333-345.

Eisenreich W, Schwarz M, Cartayrade A, Arigoni D, Zenk MH, Bacher A (1998) The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chemistry & Biology* **5**, R221-R233.

Enfissi EMA, Fraser PD, Lois L-M, Boronat A, Schuch W, Bramley PM (2005) Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. *Plant Biotechnology Journal* **3**, 17-27.

Fischer N, Nitz S, Drawert F (1987) Original flavour compounds and the essential oil composition of marjoram (*Marjorana hortensis* Moench). *Flavour and Fragrance Journal* **2**, 55-61.

Franklin EC, Snyder EB (1971) Variation and Inheritance of Monoterpene Composition in Longleaf Pine. *Forest Science* **17**, 178-&.

Gijzen M, Lewinsohn E, Croteau R (1991) Characterization of the constitutive and wound-inducible monoterpene cyclases of Grand Fir (*Abies grandis*). *Archives of Biochemistry and Biophysics* **289**, 267-273.

Goodger JQD, Connelly CA, Woodrow IE (2007) Examination of the consistency of plant traits driving oil yield and quality in short-rotation coppice cultivation of *Eucalyptus polybractea*. *Forest Ecology and Management* **250**, 196-205.

Guevara-Garcia A, San Roman C, Arroyo A, Cortes ME, Gutierrez-Nava MD, Leon P (2005) Characterization of the *Arabidopsis* clb6 mutant illustrates the importance of posttranscriptional regulation of the methyl-D-erythritol 4-phosphate pathway. *Plant Cell* **17**, 628-643.

Hammer KA, Carson CF, Riley TV, Nielsen JB (2006) A review of the toxicity of *Melaleuca alternifolia* (tea tree) oil. *Food and Chemical Toxicology* **44**, 616-625.

Han KP, Lincoln DE (1994) The evolution of carbon allocation to plant secondary metabolites - a genetic analysis of cost in *Diplacus aurantiacus*. *Evolution* **48**, 1550-1563.

Hanover JW (1966a) Genetics of terpenes .I. Gene control of monoterpene levels in *Pinus monticola* Dougl. *Heredity* **21**, 73-81.

Hanover JW (1966b) Inheritance of 3-carene concentration in *Pinus monticola*. *Forest Science* **12**, 44-53.

Hassan Y (2007) Making scents: Molecular biology of terpene production in medicinal tea tree (*Melaleuca alternifolia*). Australian National University.

Henery ML, Wallis IR, Stone C, Foley WJ (2008) Methyl jasmonate does not induce changes in *Eucalyptus grandis* leaves that alter the effect of constitutive defences on larvae of a specialist herbivore. *Oecologia* **156**, 847-859.

Hobbs, NT, Hilborn R (2006) Alternatives to statistical hypothesis testing in ecology: A guide to self teaching. *Ecological Applications*, **16**, 5–19.

Homer LE, Leach DN, Lea D, Lee LS, Henry RJ, Baverstock PR (2000) Natural variation in the essential oil content of *Melaleuca alternifolia* Cheel (Myrtaceae). *Biochemical Systematics and Ecology* **28**, 367-382.

Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195-204.

Huber DPW, Philippe RN, Godard KA, Sturrock RN, Bohlmann J (2005) Characterization of four terpene synthase cDNAs from methyl jasmonate-induced Douglas-fir, *Pseudotsuga menziesii*. *Phytochemistry* **66**, 1427-1439.

Hugueney P, Camara B (1990) Purification and characterization of farnesyl pyrophosphate synthase from *Capsicum annuum*. *FEBS Letters* **273**, 235-238.

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**, 945-953.

Jeon JS, Lee S, Jung KH, Jun SH, Kim C, An G (2000) Tissue-preferential expression of a rice alpha-tubulin gene, OsTubA1, mediated by the first intron. *Plant Physiology* **123**, 1005-1014.

Jeong YM, Jung EJ, Hwang HJ, Kim H, Lee SY, Kim SG (2009) Roles of the first intron on the expression of *Arabidopsis* (*Arabidopsis thaliana*) genes for actin and actin-binding proteins. *Plant Science* **176**, 58-65.

Kampranis SC, Ioannidis D, *et al.* (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: Structural insights into the evolution of terpene synthase function. *Plant Cell* **19**, 1994-2005.

Katz YS, Danon A (2002) The 3 '-untranslated region of chloroplast psbA mRNA stabilizes binding of regulatory proteins to the leader of the message. *Journal of Biological Chemistry* **277**, 18665-18669.

Keeling CI, Bohlmann J (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytologist* **170**, 657-675.

Keszei A, Brubaker C, Köllner T, Degenhardt J, Foley WJ (2010b) Functional and evolutionary relationships between
terpene synthases from Australian Myrtaceae. *Phytochemistry*.

Keszei A, Brubaker CL, Foley WJ (2008) A molecular perspective on terpene variation in Australian Myrtaceae. *Australian Journal of Botany* **56**, 197-213.

Keszei A, Hassan Y, J. FW (2010) A biosynthetic interpretation of terpene chemotypes in *Melaleuca alternifolia*. *Journal of Chemical Ecology* **(accepted)**.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* **39**, 1151-1155.

Kim SM, Kim YB, Kuzuyama T, Kim SU (2008a) Two copies of 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase (CMK) gene in *Ginkgo biloba*: molecular cloning and functional characterization. *Planta* **228**, 941-950.

Kim SM, Kuzuyama T, Kobayashi A, Sando T, Chang YJ, Kim SU (2008b) 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms *Ginkgo biloba* and *Pinus taeda*. *Planta* **227**, 287-298.

King DJ, Gleadow RM, Woodrow IE (2006) The accumulation of terpenoid oils does not incur a growth cost in *Eucalyptus polybractea* seedlings. *Functional Plant Biology* **33**, 497-505.

Köllner TG, Schnee C, Gershenzon J, Degenhardt J (2004) The variability of sesquiterpenes cultivars is controlled by allelic emitted from two *Zea mays* variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* **16**, 1115-1131.

Kulheim C, Yeoh SH, Maintz J, Foley WJ, Moran GF (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *Bmc Genomics* **10**.

Latta RG, Linhart YB, Snyder MA, Lundquist L (2003) Patterns of variation and correlation in the monoterpene composition of xylem oleoresin within populations of ponderosa pine. *Biochemical Systematics and Ecology* **31**, 451-465.

Linhart YB, Keefover-Ring K, Mooney KA, Breland B, Thompson JD (2005) A chemical polymorphism in a multitrophic setting: Thyme monoterpene composition and food web structure. *American Naturalist* **166**, 517-529.

Linhart YB, Thompson JD (1999) Thyme is of the essence: Biochemical polymorphism and multi-species deterrence. *Evolutionary Ecology Research* **1**, 151-171.

Bick JA, Lange BM (2003) Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of Biochemistry and Biophysics* **415,** 146–154.

Martin DM, Gershenzon J, Bohlmann J (2003) Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce. *Plant Physiology* **132**, 1586-1599.

McGarvey DJ, Croteau R (1995) Terpenoid Metabolism. *Plant Cell* **7**, 1015-1026.

Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**, 267-278.

Murtagh GJ (1996) Month of harvest and yield components of tea tree .1. Biomass. *Australian Journal of Agricultural Research* **47**, 801-815.

Murtagh GJ, Smith GR (1996) Month of harvest and yield components of tea tree .2. Oil concentration, composition, and yield. *Australian Journal of Agricultural Research* **47**, 817-827.

Muzika RM (1993) Terepnes and phenolics in response to nitrogen fertilization: A test of the carbon/nutrient balance hypothesis. *Chemoecology* **4**, 3-7.

Nakamura A, Shimada H, Masuda T, Ohta H, Takamiya K (2001) Two distinct isopentenyl diphosphate isomerases in cytosol and plastid are differentially induced by environmental stresses in tobacco. *Febs Letters* **506**, 61-64.

Neale DB (2007) Genomics to tree breeding and forest health. *Current Opinion in Genetics & Development* **17**, 539-544.

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* **9**, 325-330.

Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics* **154**, 923-929.

Novak J, Bitsch C, Langbehn J, Pank F, Skoula M, Gotsiou Y, Franz CM (2000) Ratios of cis- and trans-Sabinene Hydrate in *Origanum majorana* L. and *Origanum microphyllum* (Bentham) Vogel. *Biochemical Systematics and Ecology* **28**, 697-704.

Ohnuma S, Koyama T, Ogura K (1989) Kinetic studies on the prenyl chain elongation by undecaprenyl diphosphate synthase with artificial substrate homologues. *FEBS Letters* **257**, 71-74.

Pare PW, Tumlinson JH (1999) Plant volatiles as a defense against insect herbivores. *Plant Physiology* **121**, 325-331.

Penuelas J, Estiarte M, Llusia J (1997) Carbon-based secondary compounds at elevated CO2. *Photosynthetica* **33**, 313-316.

Phillips MA, D'Auria JC, Gershenzon J, Pichersky E (2008) The *Arabidopsis thaliana* type I isopentenyl diphosphate isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. *Plant Cell* **20**, 677-696.

Pogson BJ, Woo NS, Forster B, Small ID (2008) Plastid signalling to the nucleus and beyond. *Trends in Plant Science* **13**, 602-609.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**, 62-66.

Rocchini LA, Lindgren BS, Bennett RG (2000) Effects of resin flow and monoterpene composition on susceptibility of lodgepole pine to attack by the Douglas-fir pitch moth, *Synanthedon novaroensis* (Lep., Sesiidae). *Journal of Applied Entomology* **124**, 87-92.

Rockwood DL (1973) Variation in monoterpene composition of 2 oleoresin systems of Loblolly Pine. *Forest Science* **19**, 147-153.

Roff DA (1997) 'Evolutionary and Quantitative Genetics.' (Chapman and Hall: New York)

Rohdich F, Zepeck F, *et al.* (2003) The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: Studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proceedings of the National Academy of Sciences* **100**, 1586-1591.

Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* **16**, 565-574.

Rossetto M, Slade RW, Baverstock PR, Henry RJ, Lee LS (1999) Microsatellite variation and assessment of genetic structure in tea tree (*Melaleuca alternifolia* Myrtaceae). *Molecular Ecology* **8**, 633-643.

Russell M, Southwell I (2002) Monoterpenoid accumulation in *Melaleuca alternifolia* seedlings. *Phytochemistry* **59**, 709-716.

Russell MF, Southwell I (2003) Preferred age for assessment of qualitative and quantitative characteristics of the essential oil of tea tree (*Melaleuca alternifolia*) seedlings prior to plantation establishment. *Journal of Agricultural and Food Chemistry* **51**, 4254-4257.

Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. *Current Opinion in Plant Biology* **10**, 162-167.

Schnee C, Köllner TG, Gershenzon J, Degenhardt J (2002) The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (E)-beta-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. *Plant Physiology* **130**, 2049-2060.

Shalgi R, Lapidot M, Shamir R, Pilpel Y (2005) A catalog of stability-associated sequence elements in 3 ' UTRs of yeast mRNAs. *Genome Biology* **6**.

Shelton D, Aitken K, Doimo L, Leach D, Baverstock P, Henry R (2002a) Genetic control of monoterpene composition in the essential oil of *Melaleuca alternifolia* (Cheel). *Theoretical and Applied Genetics* **105**, 377-383.

Shelton D, Leach D, Baverstock P, Henry R (2002b) Isolation of genes involved in secondary metabolism from *Melaleuca alternifolia* (Cheel) using expressed sequence tags (ESTs). *Plant Science* **162**, 9-15.

Southwell IA, Russell MF (2002) Volatile oil comparison of cotyledon leaves of chemotypes of *Melaleuca alternifolia*. *Phytochemistry* **59**, 391-393.

Southwell IA, Stiff IA (1989) Ontogenetical changes in monoterpenoids of *Melaleuca alternifolia* leaf. *Phytochemistry* **28**, 1047-1051.

Southwell IA, Stiff IA (1990) Differentiation between *Melaleuca allternifolia* and *Melaleuca linariifolia* by monoterpenoid comparison. *Phytochemistry* **29**, 3529-3533.

Southwell IA, Stiff IA, Brophy JJ (1992) Terpinolene varieties of *Melaleuca*. *Journal of Essential Oil Research* **4**, 363-367.

Squillace AE (1971) Inheritance of monoterpene composition in cortical oleoresin of Slash Pine. *Forest Science* **17**, 381-&.

Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559-1564.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf polymorphisms associate with variation in flowering time. *Nature Genetics* **28**, 286-289.

Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257-1265.

Vernet P, Gouyon PH, Valdeyron G (1986) Genetic control of the oil Content in *Thymus vulgaris* L - a case of polymorphism in a biosynthetic chain. *Genetica* **69**, 227-231.

Walter MH, Hans J, Strack D (2002) Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant Journal* **31**, 243-254.

Wheeler GS (2006) Chemotype variation of the weed *Melaleuca quinquenervia* influences the biomass and fecundity of the biological control agent *Oxyops vitiosa*. *Biological Control* **36**, 121-128.

Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ (2004) Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiology* **135**, 2098-2105.

Wildung MR, Croteau RB (2005) Genetic engineering of peppermint for improved essential oil composition and yield. *Transgenic Research* **14**, 365-372.

Wise ML, Savage TJ, Katahira E, Croteau R (1998) Monoterpene synthases from common sage (*Salvia officinalis*) - cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *The Journal of Biological Chemistry* **273**, 14891-14899.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20**, 1377-1419.

Xie Z, Kapteyn J, Gang DR (2008) A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *The Plant Journal* **54**, 349-361.

Zavarin E (1970) Qualitative and quantitative co-occurrence of terpenoids as a tool for elucidation of their biosyntheses. *Phytochemistry* **9**, 1049-1063.

Photo: Wild Tea Trees by Yasmin Hassan

Genetic Tools for Improving Tea Tree Oils

by A. Keszei, H. Webb, C. Kulheim, W. Foley

RIRDC Pub. No. 10/189

This research has applied new advances in the understanding of the biosynthesis of terpenes, the major constituents of essential oils, in particular those from Tea Tree (*Melaleuca alternifolia*).

For the first time, this work has identified the genes that produce terpenes in *Melaleuca alternifolia* that provide a direct diagnostic test to ensure that only the commercially valuable chemotype is planted. This will enable the tea tree industry to optimise oil profiles at a much-reduced cost for breeding programmes. The work has also identified the genes and gene variants that are associated with higher yields of essential oils in different trees.

This is a significant achievement and makes possible direct selection of high yielding plants without the need for extensive traditional breeding programmes. In addition, it means that Australia can control the genetic resources for an important industry.

This publication is aimed primarily at breeders and individual producers in the tea tree industry. It emphasises to the industry that its profitability is dependent on small random changes in a few genes, and so conserving the genetic variability in the species is vital.

This project was funded from industry revenue which is matched by funds provided by the Australian Government.

This report, an addition to RIRDC's diverse range of over 2,000 research publications, forms part of our Tea Tree Oil Program, which aims to to support the continued development of an environmentally sustainable and profitable Australian tea tree oil industry that has established international leadership in marketing, in value-adding, and in product reliability and production.

Most of RIRDC's publications are available for viewing, free downloading or purchasing online at www.rirdc.gov.au.

Purchases can also be made by phoning 1300 634 313.

www.rirdc.gov.au

Contact RIRDC:
Level 2
15 National Circuit
Barton ACT 2600

PO Box 4776
Kingston ACT 2604

Ph: 02 6271 4100
Fax: 02 6271 4199
Email: rirdc@rirdc.gov.au
web: www.rirdc.gov.au
Bookshop: 1300 634 313