Genetic and environmental control of essential oil biosynthesis in West Australian Sandalwood

(Santalum spicatum)

Jessie Moniodis (BSc, Hons)

This thesis is presented for the Degree of

Doctor of Philosophy (PhD)

The University of Western Australia

School of Plant Biology and

School of Chemistry and Biochemistry

2014
Summary

West Australian sandalwood, *Santalum spicatum* R. Br. DC, has long been exploited for its aromatic heartwood which contains a rich mixture of sesquiterpenes. Essential oil composition can be highly variable and while this variation is of interest from a chemical ecology perspective, it also offers insight into the importance of sesquiterpene biosynthesis in sandalwood. Western Australia contains the largest plantations of the two commercially dominant species of sandalwood; *S. album* in the Tropical North and *S. spicatum* in the semi-arid south-west (wheatbelt) region. Plantations were developed to address conservation concerns and ensure continuation of the growing industry as *S. spicatum* is still harvested from natural stands. Improved management of plantations and natural stands will be enhanced by better understanding:

i) the natural variation in heartwood-oil composition in wild *S. spicatum* trees

ii) the genetic and environmental influences on oil composition and

iii) pathway genes involved in total oil production.

This information could be used to guide future tree improvement programs.

The aims of this study are two fold. One is to advance the current knowledge of the genetic and environmental factors which control essential oil biosynthesis in order to improve and exploit the commercial aspects of the oil, and the second is to aid in the conservation of this species which is currently harvested from natural stands.
Mature heartwood of *Santalum spicatum* contains a complex mixture of sesquiterpene olefins and alcohols. In an effort to better understand the basis of compositional variation in *S. spicatum* heartwood, approximately 200 wild trees from the arid (northern and goldfields) regions and the semi-arid southern (wheatbelt) area of Western Australia, were investigated for patterns in essential oil composition. The sesquiterpenes *E,E*-farnesol, *α*-, *β*, *epi*-β-santalene, *α*-exo-bergamotene and their hydroxlated analogues (the santalols) and dendrolasin were always present in heartwood extracts. Cluster analysis indicated at least four separate enzyme classes synthesize the major components in *S. spicatum* oil. There was high variation in total *α*- and *β*-santalol content, the components which impart value to sandalwood oil, and ranged from 1–56% of the total mixture, with some trees approaching the quality expected for the more valuable *S. album*, particularly trees from the north and south-west of the state. Most individuals contained considerable levels of the allergenic *E,E*-farnesol which ranged from 2 to 46% of the total oil content. Several provenances contained trees with good quality heartwood consisting of high santalol and low farnesol amounts.

Terpene synthase and cytochrome P450 genes involved in sesquiterpene production in sandalwood were sequenced and biochemically characterised in vitro. Using next-generation sequencing, a cDNA library derived from the xylem tissue of *S. spicatum*, was generated and screened for candidate genes using homologous sequences. From the 12, 537 contiguous gene sequences, several terpene synthases (TPS), cytochromes P450 and allylic phosphatases, and genes in the mevalonic acid (MEV) and methylerythritol phosphate pathway (MEP) were mined. The genes and encoded enzymes responsible for santalene biosynthesis in *S. spicatum* were characterised, as well as three additional terpene synthases (TPS). The santalene synthase produced *α*-, *β*-, *epi*-β-santalene, and *α*-exo-
bergamotene, precursors to the commercially valuable santalols. Additional cDNAs that encoded a bisabolol synthase, sesquisabinene B synthase and a guaiol synthase were cloned and functionally characterized. The santalene and bisabolol synthases were highly conserved across divergent sandalwood species (including *S. album* and *S. austrocaledonicum*) suggesting these terpenes have played an important role in the evolution of the *Santalum* genus.

Unlike *S. spicatum*, the four hydroxylated analogues of the santalenes and bergamotene in *S. album* (α-, β-, and epi-β-santalol and α-exo-bergamotol), comprise approximately 90% of the total oil. By mining a transcriptome library of *S. album*, nine cDNAs encoding a small family of cytochrome P450-dependent monooxygenases in the *SaCYP76* family, were cloned and functionally characterized using yeast based on their similarity to known terpene modifying monooxygenases. Members of the *SaCYP76* group were able to produce the santalols and bergamotol *in vitro* and *in vivo*. Results presented in this thesis provide a foundation for production of sandalwood oil by means of metabolic engineering as demonstrated by the engineered yeast cells which produced santalols and bergamotol. In addition, this biotechnology opportunity can address conservation concerns by reducing pressure on supply of sandalwood from natural stands.

The results of the oil analysis from *S. spicatum* was combined with data from neutral genetic markers (microsatellites) and the local habitats, to help source the origin of chemical variability. Oil composition was strongly associated to provenance, geographic location and soil types (*P* < 0.01), indicating that factors within the natural habitat are likely to be influencing total oil production and contributing to the observed
variability. Individuals from the arid regions (north and goldfields) were genetically
distinct from trees found in the semi-arid regions of WA (wheatbelt). Most trees were
sourced from the semi-arid south (wheatbelt) region where *S. spicatum* is currently
harvested, and chemotypes within this region were more influenced by provenance
(geographic location, soil, host species) than by genetics. Results show that the
heartwood oil composition in *S. spicatum* is likely to be influenced by selection
pressures arising from abiotic and biotic factors in the local environment and to a lesser
extent, the genetic composition of individuals. Further work is needed to ascertain
specific elicitors of sesquiterpene production, including a study on the differential
expression patterns of sesquiterpenes which will enable a better understanding of causes
underlying chemical variation in this species, and across the genus. Knowledge of the
 genetic and environmental factors controlling oil biosynthesis will form the basis for
managing plantations into the future to improve their productivity, as well as provide
the possibility for developing biomarkers to assess the initiation of oil production in
order to enhance the value of the sandalwood industry and aid future conservation
efforts.
Table of Contents

Summary............................................................................................................................................. ii
Table of Figures....................................................................................................................................... ix
Table of Tables ....................................................................................................................................... xvii
Acknowledgements .................................................................................................................................. xix
Statement of Candidate contribution ........................................................................................................ xxii

Chapter One: An Introduction and Literature Review on West Australian Sandalwood (*Santalum spicatum* R. Br. DC) ......................................................................................................................... 1

References ................................................................................................................................................ 20

Chapter Two: Essential Oil Variation in West Australian Sandalwood (*Santalum spicatum* R. Br. DC) ......................................................................................................................................................... 31

Abstract................................................................................................................................................. 33
Introduction ............................................................................................................................................. 34
Experimental.......................................................................................................................................... 37
Results and Discussion .......................................................................................................................... 40
Conclusions ........................................................................................................................................... 53
Supplementary Materials ....................................................................................................................... 53
Acknowledgements ............................................................................................................................... 54
References ................................................................................................................................................ 55

Chapter Three: Sandalwood Fragrance Biosynthesis Involves Sesquiterpene Synthases of Both the Terpene synthase (TPS)-a and TPS-b Subfamilies, Including Santalene Synthases ........................................................................................................ 59

Abstract............................................................................................................................................... 60
Introduction ........................................................................................................................................... 61
Experimental......................................................................................................................................... 63
Results ................................................................................................................................................... 70
Discussion ............................................................................................................................................. 77
Conclusions ........................................................................................................................................... 82
Acknowledgements ............................................................................................................................... 82
Chapter Seven: Genetic and environmental influences on essential oil composition in West Australian sandalwood (*Santalum spicatum*) ............................................................... 201

Abstract .................................................................................................................. 202
Introduction ........................................................................................................... 203
Experimental .......................................................................................................... 207
Results ................................................................................................................... 211
Discussion .............................................................................................................. 222
Acknowledgements ............................................................................................... 231
References ............................................................................................................. 232

Chapter Eight: Final Conclusions .......................................................................... 239
References ............................................................................................................. 244
Table of Figures

1.1. Distribution of Santalum in the Pacific Region .................................................... 4

2.1. Biosynthesis of major sesquiterpene alcohols in S. spicatum. Proposed formation of farnesol from farnesyl diphosphate (FPP) via a terpene synthase (TPS) or phosphatase activity, and the known pathway for santalol biosynthesis in sandalwood via TPS and cytochrome P450 (P450) activities ......................... 36

2.2. Map of Western Australia showing locations for collection of 194 S. spicatum heartwood-cores for chemical analysis ............................................................... 38

2.3. Frequency of S. spicatum individuals with different relative amounts of (A) α- and β- santalol concentration and (B) E,E-farnesol in heartwood samples of 194 trees growing in natural stands in Western Australia: 152 trees were from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the North regions (Carnarvon and Shark Bay) ........................................... 45

2.4. Two dimensional principal component analysis (PCA) ordination scores of S. spicatum sesquiterpene oil samples from the heartwood of 194 trees. Each point represents an individual tree, and points close together are similar in terms of oil composition. The first three components represent 95% of the total variance A) represents components 1 and 2 and B) represents components 1 and 3. Lines indicate ordination scores of the variables used .................................................. 49

2.5. Dendrogram showing co-occurrence patterns of 13 major components using hierarchical cluster analysis from the heartwood-oil samples of 194 S. spicatum individuals distributed in Western Australia: 152 trees from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the north regions (Carnarvon and Shark Bay) .......................................................... 51

S2.1. The linear correlation between E,E-farnesol, dendrolasin, E-nerolidol and E-β-farnesene based on GC-MS analysis 194 heartwood cores of S. spicatum trees in natural stands of Western Australia: 152 trees from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the north regions (Carnarvon and Shark Bay) using percent composition data ............................................. 53

3.1. Biosynthesis of sesquiterpenes in sandalwood commences with the TPS-catalyzed rearrangements of farnesyl diphosphate. Specific oxidation at C12 is proposed to occur via a cytochrome P450 enzyme ................................................. 62
3.2. GC-MS chromatogram of in vitro assays with recombinant santalene synthases; SaSSy (solid trace), SauSSy (dashed trace), and SspiSSy ........................................... 71

3.3. GC-MS chromatogram of incubations of (Z,Z)-FPP with SaSSy (A), (E,E)-FPP with SaSSy (B), and the olefin fraction of authentic sandalwood oil for comparison (C). Peaks: 1) α-endo-bergamotene, 2) α-santalene, 3) α-exo-bergamotene, 4) (Z)-β-farnesene, 5) epi-β-santalene, 6) β-santalene, 7) E-β-farnesene ............................................................................................................................. 72

3.4. In vivo production of santalenes in overnight E. coli cultures. Peak 1, α-santalene; peak 2, α-exo-bergamotene; peak 3, epi-β-santalene; peak 4, β-santalene .............................................................................................................. 73

3.5. Detailed mechanism explaining the formation of products resulting from incubations of SaSSy with (E,E)-farnesyl diphosphate in the presence of Mg$^{2+}$ ............................................................................................................................. 76

3.6. Detailed mechanism explaining the formation of products resulting from incubations of SaSSy with (Z,Z)-farnesyl diphosphate in the presence of Mg$^{2+}$ ............................................................................................................................. 78

3.7. Neighbor-joining phylogenetic tree of the TPSs compared in this study. The santalene synthase and bisabolene/bisabolol synthase orthologs group with the TPS-b clade. Abbreviations and GenBankTM accession numbers are listed under “Experimental Procedures”. The asterisk indicates a bootstrap value greater than 95% ................................................................................................. 80

3.S3. SDS-PAGE of Ni$^{2+}$ affinity purified recombinant proteins. A; S. spicatum TPS enzymes. B; SaSSy and S. austrocaledonicum TPS enzymes. Expression of SauSesquiTPS and SspiSesquiTPS (approximate size 68 kDa) was particularly poor ........................................................................................................................................ 85

3.S5. GC-MS chromatogram and mass spectra of sesquiterpenes produced by SauBS and SspiBS when incubated with (E,E)-FPP and Mg$^{2+}$ ...................................................................................................................... 88

3.S7. GC-MS trace and mass spectra of sesquiterpenes produced by SauSesquiTPS when incubated with (E,E)-FPP and Mg$^{2+}$ ...................................................................................................................... 90

3.S8. GC-MS trace and mass spectra of sesquiterpenes produced by SspiSesquiTPS when incubated with (E,E)-FPP and Mg$^{2+}$ ....................................................................................................................... 91

3.S10. ClustalX protein alignment of TPS genes in this study. Germacrene D synthase from Populus trochoxocarpa x deltoides, 5-epi-aristolochene synthase from Capsicum annuum var. annuum and cineole synthase from Salvia fruticosa are included for comparison. R(P/R)X₈W and DDXXD motifs are indicated with a black overhead line. Red overhead lines indicate positions most likely to influence product profiles and substrate utilisation, as identified by Kampranis et al. (1). Gray histogram indicates degree of conservation of residues ......................... 93
3. S11. Genomic structures of deduced open reading frames of TPS genes amplified from 
*S. album, S. austrocaledonicum, S. spicatum and S. murrayanum*. Exons are in 
red. R(P/R)X₈,W and DDXXD motifs are highlighted. Double dagger indicates a 
10 nucleotide deletion in the *SaSSy* gDNA sequence and asterisk indicates a stop 
codon in exon 3 of *SmBS*; an *S. murrayanum* bisabolene synthase-like gene .....

4.1. Sesquiterpenes found in *S. spicatum* essential oil................................. 103

4.2. Collection and extraction of *S. spicatum* oil; A. Sandalwood tree drilled at 30 
cm; B. Grind wood shavings; C. Ground wood shavings extracted in 25 mL 
volumetric flasks using hexane for 7 days followed by D. Analysis by GC-MS 
........................................................................................................................... 109

4.3. The environmental differences in several collection sites across the wheatbelt in 
south Western Australia: A. clay-type soils; B. rocky ground cover; C. 
extensive ground covering with shrubs............................................................. 111

4.4. a) GC-MS oil profile of *S. spicatum*. Sesquiterpenoids known to be produced by 
characterised TPS enzymes are highlighted with arrows according to their 
retention time. The percent composition of each compound is based on extracts 
of 156 *S. spicatum* individuals from the semi-arid wheatbelt region of Western 
Australia. b) Summary of TPS enzymes from *S. spicatum* (Jones et al. 2011) and 
the products resulting from incubations with FPP (abbreviations: FPP, farnesyl 
pyrophosphate; *SspiSSy, S. spicatum* santalene synthase; *SspiBS, S. spicatum* 
bisabolol synthase; *SspiSesqui, S. spicatum* sesquiterpene synthase; TPS, terpene 
synthase)............................................................................................................ 113

5.1. Schematic biosynthetic pathway for santalols and bergamotol in sandalwood. 
Compounds identified with numbers are: α-santalene (1), α-exo-bergamotene 
(2), epi-β-santalene (3), β-santalene (4), (Z)-α-santalol (5), (E)-α-santalol (7), 
(Z)-α-exo-bergamotol (6), (E)-α-exo-bergamotol (8), (Z)-epi-β-santalol (9), 
(E)-epi-β-santalol (11), (Z)-β-santalol (10), (E)-β-santalol (12). Numbers match 
the numbers in Table DMADP, dimethylallyl diphosphate; IPP, 
isopentenyl diphosphate; FPP, farnesyl diphosphate; FPPS, farnesyl 
diphosphate synthase; *SaSSy, S. album* santalene synthase ..................... 124

5.2. Phylogenetic tree of *S. album* CYP76F proteins and related terpene-modifying 
P450s. The neighbor-joining tree was constructed with members of the CYP71 
clan, using *Picea sitchensis* PsCYP720B4 (ADR78276) as an outgroup. *S. 
album* CYP76F proteins fell into two clades, clade I santalene/bergamotene 
oxidases and clade II bergamotene oxidases. *CaCYP76B4, Camptotheca 
acuminata* putative geraniol-10-hydroxylase(AES93118); *CrCYP76B6, 
*Catharanthus roseus* geraniol-10-hydroxylaseb(Q8VWZ7); *SmCYP76B4, 
*Swertia mussotii* geraniol-10-hydroxylase (D1MI46); *OsCYP76M7 Oryza 
sativa* ent-cassadiene C11-α-hydroxylase (NP_001047185); *MpCYP71A32, 
*Mentha x piperita* menthofuran synthase (Q947B7); *PaCYP71A1, Persea
americana (P24465); CiCYP71AV8, Cichorium intybus valencene oxidase (ADM86719); MpCYP71D13, Mentha x piperita; (-)-limonene-3-hydroxylase (AY281027); NiCYP71D20, Nicotiana tabacum, 5-epi-aristolochene-1,3-dihydroxylase (AF368376); GaCYP706B1, Gossypium arboresum (+)-delta-cadinene-8-hydroxylase (AAD86719); MpCYP71D13, Mentha x piperita; (-)-limonene-3-hydroxylase (AY281027); NiCYP71D20, Nicotiana tabacum, 5-epi-aristolochene-1,3-dihydroxylase (AF368376); GaCYP706B1, Gossypium arboresum (+)-delta-cadinene-8-hydroxylase (AAD86719). This work: SaCYP76F37v1 (KC533717); SaCYP76F37v2 (KC698966); SaCYP76F38v1 (KC53371S); SaCYP76F38v2 (KC53371S); SaCYP76F39v1 (KC533716); SaCYP76F39v2 (KC698967); SaCYP76F40 (KC698968); SaCYP76F41 (KC698969); SaCYP76F42 (KC698965); SaCYP76F43 (KC533719) ................................. 134

5.3. GCMS analysis of products formed in vitro with SaCYP76F39v1. A sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Supplementary Figure 5.S3) was incubated with microsomes containing SaCYP76F39v1 and SaCPR produced in yeast. (A) Product profile (extracted ion chromatogram, EIC) of assays with SaCYP76F39v1. (B) Authentic S. album oil. (C) Control assays were performed with microsomes isolated from yeast cells transformed with the empty vector. Mass spectra of compounds corresponding to peaks 5-12 identified in assays with SaCYP76F39v1 (left panel) and S. album oil (right panel) are shown in Figure 5.5S4. Peak numbers match the numbers in Table 5.1 and Figure 5.1 ........................................................................................................... 137

5.4. GCMS analysis of products formed in vitro with clade I SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of products formed in vitro with (A) SaCYP76F39v2; (B) SaCYP76F40; (C) SaCYP76F41; (D) SaCYP76F42. Assays were performed with a sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Figure 5.S3) as substrate and microsomes prepared from yeast transformed with SaCPR, individual clade I candidate SaCYP76F cDNAs, or (E) empty vector as control. Peak numbers match the numbers in Table 5.1 and Figure 5.1 ........................................................................................................... 138

5.5. GCMS analysis of products formed in vitro with clade II SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of products formed in vitro with (A) SaCYP76F38v1; (B) SaCYP76F38v2; (C) SaCYP76F37v1; (D) SaCYP76F37v2. Assays were performed with a sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Figure 5.S3) as substrate and microsomes prepared from yeast transformed with SaCPR, individual clade II candidate SaCYP76F cDNAs, or (E) empty vector as control. Peak numbers match the numbers in Table 5.1 and Figure 5.1 ........................................................................................................... 140

5.6. GCMS analysis of products formed in vivo with SaCYP76F39v1. GCMS analysis (extracted ion chromatograms) of compounds formed in vivo in yeast cells expressing SaSSY, SaCPR2 and (A) SaCYP76F39v1 or (B) an empty vector. (C) Mass spectra of compounds corresponding to peaks 5-12 identified in (A). Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks in (A) and (B) marked with symbol (*) correspond to farnesol also produced in yeast cells.
without SaCYP76F. Peaks in (A) marked with symbol (#) represent yeast in vivo modifications of santalols (see Figure 5.56) ...................................................... 146

5.S1. Amino acid sequence alignment of S. album CYP76F genes. SaCYP76F37v1, SaCYP76F37v2, SaCYP76F38v1, SaCYP76F38v2, SaCYP76F39v1, SaCYP76F39v2, SaCYP76F40, SaCYP76F41, SaCYP76F42 and SaCYP76F43. Red, dark grey and light grey shading denote 100% and 80% and 50% conserved residues, respectively. Horizontal arrows denote the proline region (a), O₂ binding motif (b) and heme binding motif (c). Boxes indicate the putative substrate recognition sites (SRS) regions originally described by Gotoh [31]. Multiple sequence alignment was performed with the software CLUSTALW [32] and visualized with Gendoc v2.7 ......................... 153

5.S2. Reduced CO-difference spectra of isolated microsomes containing S. album CYP76F proteins. CO difference spectra of microsomal fractions from S. cerevisiae harboring a cytochrome P450 or an empty vector are shown. Concentration of SaCYP76F proteins are given based on an extinction coefficient of 91,000 M⁻¹ cm⁻¹ .................................................................................................................. 154

5.S3. GCMS analysis (extracted ion chromatogram) of a sesquiterpene mixture and fractions separated by TLC. The sesquiterpene mixture was produced with a recombinant yeast strain expressing SaSSy (10) and was provided to us by Allylix Inc. It contained (A) α-santalene (1), α-exo-bergamotene (2), epi-β-santalene (3), β-santalene (4). The mixture was separated by TLC into three fractions containing mainly (B) α-santalene (1); (C) α-exo-bergamotene (2); or (D) β-santalene (4). Mass spectra of peaks 1 to 4 are provided. Peak numbers match the numbers in Table 5.1 and Figure 5.1 ..................................................................................... 155

5.S4. Mass spectra of products formed in vitro with SaCYP76F39v1. Mass spectra of compounds corresponding to peaks 5-12 shown in Figure 5.3 and identified in assays with CYP76F39v1 (left panel) and S. album oil (right panel). Peak numbers match the numbers in Table 5.1, Figure 5.1, and Figure 5.3............. 156

5.S5. GCMS analysis (extracted ion chromatogram) of products formed in vitro with SaCYP76F39v1 or SaCYP76F37v1 using partially purified substrates. Product profile in assays with SaCYP76F39v1 using (A) α-santalene, (B) α-exo-bergamotene, or (C) epi-β-santalene and β-santalene as substrate. Product profile in assays with SaCYP76F37v1 using (D) α-santalene, (E) α-exo-bergamotene, or (F) epi-β-santalene and β-santalene as substrate. (G) Products were identified by comparison to authentic standards. Peak numbers match the numbers in Table 5.1 and Figure 5.1................................................................. 157

5.S6. GCMS analysis (extracted ion chromatogram) and mass spectra of sesquiterpenes produced in yeast expressing SaSSy. (A) GCMS analysis of sesquiterpenes extracted from pelleted yeast cells expressing SaSSy. (B) Mass spectra of peaks 1-4: α-santalene (1), α-exo-bergamotene (2), epi-β-santalene (3),
and β-santalene (4). Compounds were identified by comparison to an authentic standard and retention indices. Peak numbers match the numbers in Table 5.1 and Figure 5.1.

5.S7. Modification of sandalwood oil compounds in yeast cell culture. GCMS analysis of sesquiterpenols of natural sandalwood oil sample before (A) and after (B) overnight incubation with yeast cells, which do not contain a SaCYP76F gene. Peaks in (B) marked with symbol (#) represent yeast in vivo modifications of santalols independent of SaCYP76F. Peak numbers match the numbers in Table 5.1 and Figure 5.1.

5.S8. GCMS analysis of products formed in vivo with clade I SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of compounds formed in vivo in yeast cells expressing SaSSy, SaCPR2 and (A) SaCYP76F39v2, (B) SaCYP76F40, (C) SaCYP76F41, or (D) SaCYP76F42. Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks marked with symbol (*) correspond to farnesol produced also in yeast cells without SaCYP76F. Peaks in marked with symbol (#) represent yeast in vivo modifications of santalols (see Figure 5.S7).

5.S9. GCMS analysis of products formed in vivo with clade II SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of compounds formed in yeast cells expressing SaSSy, SaCPR2 and (A) SaCYP76F38v1, (B) SaCYP76F38v2, (C) SaCYP76F37v1, (D) SaCYP76F37v2, or (E) SaCYP76F43. Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks marked with symbol (*) correspond to farnesol produced also in yeast cells without SaCYP76F. Peaks in marked with symbol (#) represent yeast in vivo modifications of santalols (see Figure 5.S7).

6.1. Figure 6.1 - Neighbor-joining phylogenetic tree of TPS from S. spicatum (TPS-a and b), and select TPS of other species (see Supplemental Table 1). Sequences were aligned using Clustal W and the phylogenetic tree constructed with MEGA5.2 (Tamura et al., 2011). Bootstrap confidence values over 50% are shown at branch points based on 500 replications.

6.2. GC-MS chromatogram of in vitro assay products of recombinant SspiTPS4 using FPP as a substrate and Mg$^{2+}$. Peaks: 1 unknown, 2 sesquisabinene B, 3 α-acoradiene, 4 γ-curcumene, 5 β-bisabolene, 6 β-sesquiphellandrene.

6.3. GC chromatogram of in vitro assay products of recombinant SspiTPS4 using GPP as a substrate and Mg$^{2+}$ or Mn$^{2+}$ (dotted). Peaks: 1 α-pinene, 2 β-pinene, 3 sabinene, 4 myrcene, 5 linalool, 6 α-terpineol.

6.4. Proposed biosynthetic mechanism for the formation of sesquiterpene products by SspiTPS4. The reaction starts with the ionization of E,E-FPP, isomerization of the farnesyl cation and proceeds via cyclisation, hydride shifts and
deprotonations. The numbering refers to that of E,E-FPP beginning the reaction

S6.1. Metabolic pathway for terpene biosynthesis. Abbreviations: AACT, Acetyl-CoA thiolase; HMGS, 3-Hydroxy-3-methylglutaryl-CoA Synthase; MK, mevalonate Kinase; PMK, phosphomevalonate kinase; MDC, pyrophosphomevalonate decarboxylase; IPPI, isopentenylpyrophosphate isomerase; DMAPP, dimethylallyl pyrophosphate FPPS, farnesyl pyrophosphate synthase; DXS, 1-Deoxy-D-xylulose 5-phosphate synthase; DXR, 2-C-methylerythritol 4-phosphate reductase; MCT, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 1-hydroxy-2-methyl-butenyl 4-diphosphate reductase; GPS, geranyl pyrophosphate synthase; GGPP, geranylgeranyl pyrophosphate synthase; TPS, terpene synthase; P450, cytochrome P450.

S6.2. Functional classification of S. spicatum xylem transcriptome

S6.3. Comparison of the deduced amino acid sequence of SspiTPS4 from eight individuals of S. spicatum. Amino acid differences are marked by black boxes. The aspartate-rich DDxxD region is shown. The N terminal domain is labelled with a bar.

S6.4. Genomic structure of SspiTPS4 showing the intron/exon boundaries. R(R/P)X₈ motif in the N-terminal region implicated in prenyl diphosphate ionization and the aspartate rich metal ion binding domain (DDxxD) are highlighted (Aubourg et al., 2002, Lesburg et al., 1998, Starks et al., 1997, Williams et al., 1998).

S6.5. Mass spectral data of in vitro assay products of recombinant SspiTPS4 using FPP as a substrate and Mg²⁺. Numbering according to Figure 6.2: 1 unknown, 2 sesquisabinene B, 3 α-acoradiene, 4 γ-curcumene, 5 β-bisabolene, 6 β-sesquiphellandrene.

S6.6. Mass spectral data of in vitro assay products of recombinant SspiTPS4 using GPP as a substrate and Mg²⁺. Numbering according to Figure 6.3: 1 α-pinene, 2 β-pinene, 3 sabinene, 4 myrcene, 5 linalool, 6 α-terpineol.

7.1. Map of Western Australia showing the distribution of 186 individuals sampled for genetic and chemical comparisons in the arid regions (Northern and Goldfields) and the semi-arid (Wheatbelt) region in the south-west.
7.2. Assignment probability of individuals from the arid (north and goldfields) and semi-arid south (wheatbelt) region using STRUCTURE ................................. 214

7.3. Assignment probability of individuals to each genetic cluster from (a) arid region including the north and goldfields (two distinct clusters) and (b) the semi-arid south west wheatbelt (two distinct clusters) using STRUCTURE ... 215

7.4. MDS (multidimensional scale) plot of genetic distance between regions of *S. spicatum*. North trees are in red, goldfields trees are in black and wheatbelt trees represented by green circles .............................................................................. 216

7.5. a) Population based phylogeny consensus tree (maximum likelihood) and b) dendogram which clusters oil profiles together ................................................................. 217
Table of Tables

2.1. Sesquiterpene composition of S. spicatum heartwood oil from three different regions of West Australia (Wheatbelt, Goldfields and North) ......................... 42

3.S1. Primers used in this study. Restriction sites are underlined ............................. 83
3.S2. Sequences and abbreviated names used for phylogenetic tree ......................... 84
3.S4. Product profiles of SaSSy, SauSSy and SspiSSy. Positive identification of products was based on comparison to authentic standards and mass spectral data, along with comparison to Kovats retention indices quoted in the literature. * = run on a different DB-WAX column under slightly different oven conditions: 1 min at 40°C then 10°C min⁻¹ to 230°C held 25 min. nd = none detected .......... 86
3.S6. Product profiles of SauBS and SspiBS. tr = trace ................................................. 89

4.1. Summary of total oil content of 147 S. spicatum heartwood samples extracted from wild trees in the wheatbelt and nine individuals from an 11 yr old plantation in the wheatbelt ................................................................................ 108
4.2. Total santalols and E-E-farnesol as percentage of total oil composition of S. spicatum heartwood samples extracted from wild trees (N = 147) and individuals from an 11 year-old plantation (N = 9) ............................................................................................................. 109

5.1. Retention indices of sesquiterpenes and sesquiterpenols identified in the enzyme assays with cytochromes P450 of the S. album CYP76F subfamily and of sesquiterpene alcohols of S. album oil ................................................................. 133
5.2. Relative activities of SaCYP76F39v1 and SaCYP76F37v1 with different sesquiterpenes ........................................................................................................ 143
5.S1. Primers designed for amplification of cDNAs from S. album ......................... 151
5.S2. Summary of transcriptome mining for CYP76 family members in the S. album Sanger and 454 sequence data ................................................................................. 152
6.1. Summary of genes identified in the xylem transcriptome encoding enzymes of the MEV and MEP pathway, TPS, cytochrome P450, and allylic phosphatase..... 178

6.2. Product profile $Ssp\text{TPS}4$ with FPP and M$g^{2+}$ identified by GC-MS analysis using a DB-Wax column, NIST08 library and Kovats retention indices. Peak numbering matches peaks shown in Figure 6.2 ................................................ 181

6.3. TPS used to construct the neighbor-joining phylogenetic tree (Fig. 6.1) .......... 187

7.1. Summary of genetic diversity estimates obtained with eleven nuclear microsatellites for Santalum spicatum. Results include combined individuals.213

7.2. Mantel test results for oil composition and genetic and geographic distances for (a) all 186 individuals of S. spicatum from semi-arid wheatbelt and arid north and goldfields and (b) south wheatbelt ................................................................. 218

7.3. Adonis test results showing P-values for oil composition and environmental and morphological features for (a) all 186 individuals of S. spicatum from semi-arid wheatbelt and arid north and goldfield and (b) south wheatbelt......................... 221
Acknowledgements

First and foremost, I wish to thank the list of supervisors which have provided me with support, instruction and guidance throughout my PhD: Dr. Chris Jones, Dr. Liz Barbour, Prof. Julie Plummer, Assoc. Prof. Michael Renton and Prof. Emilio Ghisalberti. I have been lucky to work with such a wonderful group of people and I deeply appreciate all their friendship, advice, ideas and funding opportunities which have made my PhD research possible. I am indebted to Prof. Joerg Bohlmann who allowed me to visit his fantastic lab in Vancouver, and for sharing his extensive knowledge and wisdom, providing frequent advice and support throughout. I am extremely grateful to my supervisor at DPaW, Dr. Margaret Byrne for her immense body of knowledge on genetics, the opportunity to work in her laboratory and continued assistance throughout my PhD.

I appreciate all the funding provided to this linkage grant through the Australian Research Council. I thank the Forest Products Commission for funding and the opportunity to visit parts of Western Australia I never thought I would see. On these trips I thank Mr. Len Norris for sharing his knowledge and experience with sandalwood and taking the time to navigate and kindly assist with all field trips. I am thankful to Mr. Jon Brand for sharing his vast experience on S. spicatum and allowing me to continue accessing his plantation trees. I am grateful to Mr Dave Evans who assisted on my Northern WA field trip, without which I would have spent a long time in navigation. In addition, I am grateful to other travel scholarships provided by the school of Plant Biology and Canada's (Banff) travel award which made attending the conference in Calgary, Canada, much easier.
I am deeply grateful to all the people in the laboratory of Prof. Joerg Bohlmann in Vancouver, Canada, (the Michael Smiths Laboratory) who not only always made me feel welcome when I visited, but shared their extensive scientific knowledge and provided much support for my research. I especially thank Lina for being so wonderful and generous with her time and knowledge, Maria D for her experience with enzyme expression and for teaching me a suite of useful skills, Christopher, Karen, Harpreet, Britta and Bjoern, Hannah, Andreas, Phillip, Melissa, Sharon, Angela, Alfonso, Dawn, Katherine and Maria Li for providing invaluable advice and excellent suggestions all throughout.

I would like to acknowledge all the people in the Science Division at the Department of Parks and Wildlife, especially Bronwyn, Shelley and Melissa for their technical assistance, advice, support and friendship over the years.

I am deeply grateful to all my family, friends and fellow PhD students which have been a wonderful source of friendship and help over the years, without which I would have surely been lost. I am indebted to my sisters, Zena and Laura, and Matt who put up with my practice talks for the first year. I appreciate all the support from my fellow PhD students, Xiangling, Diane, Xing, Xin Li, Meta and I thank my cousin Joe for his support and guidance over the years. I thank Anna, Nicole, Maria, Anna, George for letting me take breaks to visit Romania, Cale, Niyaaz, Bella, Trinity, Joycey and Florence, who all helped and supported in some capacity when I was feeling under pressure from work - I really appreciate those times. I thank my parents for their continued love and support and I thank my 'adopted parents' in Canada, Agi and Frank, who made my first experience so welcoming and memorable. I am grateful to the many
other students who enriched my experiences, Arpiwi, Xue, Xin, Christine, Eduardo and other volunteers in the Rottnest/Ecophysiology committee. I am thankful to Mashid and little Qianah in Vancouver who during my second visit, provided me with much care and welcomed me into their family. I am deeply grateful to my family at Kanyana Wildlife who have been supportive and understanding throughout my journey, in particular Jade, Wendy, Nat, Johanna, Gwyn and my other shift members who have been wonderful friends, mentors, and lots of fun. Finally, thank-you to Andrew, who supported and put up with me while I was writing.
DECLARATION FOR THESIS CONTAINING PUBLISHED WORK
AND/OR WORK PREPARED FOR PUBLICATION

This thesis contains published work and/or work prepared for publication, some of which have been co-authored. The chapters in this thesis are presented below with the candidate contribution. The contribution for Chapters One, Two, Four, Six, Seven and Eight is > 90 % by the author of this thesis and <10 % for additional authors.

Chapter One: An Introduction and Literature Review on West Australian Sandalwood (Santalum spicatum R. Br. DC)
Authors: Jessie Moniodis and Liz Barbour

Chapter Two. Essential Oil Variation in West Australian Sandalwood (Santalum spicatum R. Br. DC)
Authors: Jessie Moniodis, Christopher G Jones, Michael Renton, E Liz Barbour, Julie A Plummer, Emilio L Ghisalberti and Joerg Bohlmann

This chapter has been submitted to the journal Phytochemistry.
Chapter Three: Sandalwood Fragrance Biosynthesis Involves Sesquiterpene Synthases of Both the Terpene Synthase (TPS)-a and TPS-b Subfamilies, including Santalene Synthases

Authors: Christopher G. Jones, Jessie Moniodis, Katherine G. Zulak, Adrian Scaffidi, Julie A. Plummer, Emilio L. Ghisalberti, Elizabeth L. Barbour, and Joerg Bohlmann

Accepted for publication in the Journal of Biological Chemistry, 2011; 286(20), pp 17445-17454

The author contributed one third of the paper pertaining to the sandalwood species *Santalum spicatum*. The remaining two species were investigated by additional authors Dr. Chris Jones for *Santalum album*, who initiated the research, and Dr. Katherine Zulak for *Santalum austrocaledonicum*. Although results for *S. spicatum* could have been published as a single paper, results were combined in order to submit to a journal with a higher impact factor.

Chapter Four: Essential oil components and their biosynthesis in West Australian sandalwood oil (*Santalum spicatum*)

Authors: Jessie Moniodis, Christopher G Jones, Julie A Plummer, Liz Barbour, Emilio Ghisalberti, Joerg Bohlmann
Chapter Five: Biosynthesis of Sandalwood Oil: *Santalum album* CYP76F

Cytochromes P450 Produce Santalols and Bergamotol

Authors: Maria L. Diaz-Chavez, Jessie Moniodis, Lufiani L. Madilao, Sharon Jancsik, Christopher I. Keeling, Elizabeth L. Barbour, Emilio L. Ghisalberti, Julie A. Plummer, Christopher G. Jones, and Joerg Bohlmann

Accepted for publication in PLoS ONE 8(9), 2013, e75053.

The contribution is 40%. This work was conducted in the Michael Smith laboratory of Professor Joerg Bohlmann at the University of British Columbia, Vancouver, Canada during 2010 (preliminary work) and again during 2013 (final work). During the first visit in 2010, the PhD candidate was responsible for the initial mining of genes involved in cytochrome P450 expression using homologous sequences from other species which may oxidise similar substrates, including candidate oxidases of the CYP76, CYP91 and CYP71 families, cytochrome P450 reductases (CPR1 and CPR2). All sequences were used for subsequent expression work. Several candidate P450 genes were selected, reverse transcribed to obtain full length sequences and cloned in high copy storage vectors for later use. In addition, both CPRs' were test *in vitro* for their electron donor capacity using various truncated and full length forms which were all found to function as adequate electron donors. The PhD candidate provided full length clones and
sequences for candidate P450s to test, the cytochrome P450 reductases, and other MEV pathway genes which may help improve expression (for e.g. HMGCoA reductase) for metabolic engineering opportunities. Initial P450 expression was trialed in E. coli using the CYP71 but was unsuccessful, thus other microorganisms were sought to test. The CYP76 sequence was then used by co-author M. Diaz to mine an additional 8 CYP76 sequences involved in santalene/bergamotene oxidation, which were then employed in a yeast in vivo expression system and found to be santalene/bergamotene hydroxylases. The sequence provided by the PhD student was a bergamotene oxidase. The PhD candidate performed initial kinetic assays, finalised experiments for expression, assisted with separation of enzyme products, identification of compounds, tested the differential activity of the cytochrome P450 reductases and trialed other terpene synthases (bisabolene synthase from S. album and S. spicatum) for their activity with the cytochrome P450s. A large amount of this work was excluded from the manuscript but was essential for the processes involved in identifying P450s from Santalum. A short description of the preliminary work is provided here and the manuscript which includes the final published results are presented as a thesis chapter as the final publication addresses the initial research hypothesis.

Chapter Six: The transcriptome of sequiterpenoid biosynthesis in heartwood xylem of Western Australian sandalwood (Santalum spicatum)

Authors: Jessie Moniodis, Christopher G Jones, Julie A Plummer, Liz Barbour, Emilio Ghisalberti and Joerg Bohlmann

This chapter has been accepted for publication in Phytochemistry
Chapter Seven: Genetic and environmental influences on essential-oil composition in West Australian sandalwood (*Santalum spicatum*)

Authors: **Jessie Moniodis**, Michael Renton, Liz Barbour and Margaret Byrne

This chapter has been prepared for (future) publication into *The Australian Journal of Botany*

Candidate, Jessie Moniodis

Signature

Co-ordinating Supervisor, A/prof. Michael Renton

Signature
Statement of original contribution

This thesis represents a summary of my PhD course at the University of Western Australia. The results presented offer an original contribution to sandalwood chemistry, biochemistry, genetics and secondary metabolite biosynthesis. The work represents my ideas, with guidance and critique from supervisors. All co-authors are dutifully acknowledged and I have highlighted my contribution to co-authored papers.
CHAPTER ONE

An Introduction and Literature Review on West Australian Sandalwood (*Santalum spicatum* R. Br. DC)

Jessie Moniodis\textsuperscript{1,2} and Liz Barbour\textsuperscript{1}

1. School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley WA 6009 Australia

2. School of Chemistry and Biochemistry (M310), University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia
1.1. Distribution and importance of *Santalum*

The sandalwoods (*Santalum* L., Santalaceae) are a group of hemi-parasitic plants widely distributed throughout Australia, India and the Pacific Islands (Fig. 1.1). Sandalwood is well known for its fragrant heartwood-oil which is used in perfumery, medicine and aromatherapy (Brennan and Merlin, 1993; Jones, 2001; Rai, 1990). Almost all members of sandalwood produce sesquiterpenes, including the related African sandalwood sandalwood *Osyrius tenuifolia* (Santalaceae). There is phylogenetic evidence to support the origin of *Santalum* in Australia (Harbaugh and Baldwin, 2007). Several long distance dispersal events out of Australia, probably mediated by birds, led to the subsequent colonisation and diversification of sandalwood in Melanesia, Polynesia, Hawaii and stretching to the far south-east off the Coast of Chile to the Juan Fernandez Islands (Fig 1.1) (Harbaugh and Baldwin, 2007). There are 16 known species of *Santalum* (~ 14 varieties), with seven native to the Australian continent, including *S. lanceolatum* and *S. leptocladum* whose taxonomy has recently been revised (Applegate and McKinnell, 1993; Harbaugh, 2007). More than a quarter of sandalwoods are endemic to the Hawaii Islands where a rich history is shared (Harbaugh and Baldwin, 2007). Several species of *Santalum* have been exploited for use in natural healing, fine artwork and have held importance in Hindu and Buddhist religious ceremonies - particularly in Asian cultures (Brennan and Merlin, 1993; Loneragan, 1990; Tonts and Selwood, 2003). Nowadays, the sesquiterpene-rich oil which resides in mature heartwood is highly sought after by the flavour and fragrance industries and also contains bioactive components which are showing potential to treat medical conditions (such as herpes) in humans (Benencia and Courrèges, 1999; Dwivedi and Zhang, 1999; Hammer et al., 1998; Jirovetz et al., 2006; Tinson, 2001). Indian sandalwood (*S. album*)
is the most valuable species due to its heartwood properties (composition and to a lesser extent quantity); followed by *S. yasi*, *S. austrocaledonicum*, *S. macgregorii*, *S. spicatum* and *S. lanceolatum* (distribution is shown in Fig. 1.1). High demand and unsustainable exploitation of *Santalum* has led to the depletion of many natural resources (Brennan and Merlin, 1993; Butaud et al., 2005; Cherrier, 1993; Clarke, 2006; Jiko, 1993) and in the Juan Fernandez Islands, the once endemic *S. ferdandezianum* has been exploited to extinction (Darwin, 1839; Stuessy et al., 1992). Thus in the past 15 years there has been a movement to establish plantations of sandalwood to ensure a continued supply into the future, especially as worldwide demands grow. The largest sandalwood plantations are in Western Australia, with *S. spicatum* established in the semi-arid south-west and *S. album* in the tropical North of the state (Clarke, 2006). Plantations will alleviate the need to harvest from natural populations where most sandalwood is currently sourced. Since sandalwood shares a rich history with many cultures, this conservation effort is needed as natural resources continue to decline. If variability in the heartwood oil is heritable across *Santalum*, causes underlying this variation can be identified and the opportunity to select superior trees with desired heartwood properties for seed orchards, breeding programs and plantations becomes possible. Thus there is a considerable amount of research needed to improve current plantation management to meet market demand.
1.2. Sandalwood oil

Several species of *Santalum* produce highly-prized aromatic oil in their heartwood. Since the 1950’s, studies into the chemical composition of heartwood extracts have attributed the pleasant odour to a complex sesquiterpene-blend, with over 100 constituents now identified (Adams, 1995; Baldovini et al., 2011; Birch et al., 1953; Braun et al., 2007; Braun et al., 2003; Brophy et al., 1991; Brophy et al., 2009; Valder et al., 2003). The most important components are α- and β-santalol which impart the desired odour properties and value, particularly for the fragrance industry (Clarke, 2006). A number of species including *S. album*, *S. yasi*, *S. austrocaledonicum*, *S. macgregorii*, *S. insulare*, *S. spicatum* and *S. lanceolatum* (Fig. 1.1), contain a high
enough proportion of these two sesquiterpenes to be commercially harvested. No detailed investigations have examined whether all *Santalum* species produce oil in their heartwood. There is a possibility that species which have not been extensively studied, for e.g. *S. murrayanum* and *S. obtusifolium* - two ancestral species from Australia (Harbaugh and Baldwin, 2007), also accumulate oil in their heartwood but this needs further investigation. One known exception is *S. acuminatum*, a native to Western Australia, which does not contain detectable levels of sesquiterpenes in its heartwood but is still commercially pursued for its fruit (Clarke, 2006). *Santalum album* produces the highest proportion of α- and β-santalol which consistently makes up greater than 80% of the final distilled oil. However there have been reports of variations in total extractable oil which can range from 0 – 9% dry weight (Jones et al., 2006; Verghese et al., 1990). Other oil-bearing sandalwoods display variations in both α- and β-santalol content and total oil yields across their natural range (Brand et al., 2007; Brophy et al., 2009; Moretta, 2001; Page et al., 2010).

The two commercially dominant species of sandalwood in Australia are *S. album* and *S. spicatum* (Clarke, 2006). Most tree improvement research has focused on *S. album*, a tropical species which competes for highly sought-after agricultural land in Northern Australia. Australian sandalwood, on the other hand, has been used as a necessary deep-rooted perennial to assist with water-balance and salinity control on the poorer soils in the south-west agricultural landscape. The rapid expansion of *S. spicatum* plantations (more than 15 000 Ha) in Western Australia over the past 15 years (Australian Sandalwood Network) and initial observations have shown that tree improvement and market development for this species is also needed (Brand et al., 2007). *Santalum spicatum* is native to WA and, when harvested from wild stands, contains commercially
significant amounts of essential oil which is considered a suitable substitute for *S. album* (Adams et al., 1975). Unlike *S. album* however, the chemical profile of the aromatic oil is highly variable across its natural range (Loneragan, 1990; Moretta, 2001b), where there are known genetic and environmental differences (Byrne et al., 2003a; Byrne et al., 2003b; Byrne et al., 2008).

The oil of *S. spicatum* is a complex mixture of sesquiterpenoids with most of the same compounds as *S. album*, but with some additional ones, especially *E,E*-farnesol (Adams et al., 1975; Howes et al., 2004; Piggot et al., 1997; Valder et al., 2003). Farnesol is more concentrated in the heartwood of this species and has been labelled a potential allergen (Schnuch et al., 2004), which reduces its market value particularly for skin-based perfume and cosmetic products. Total oil contents have also been reported to vary from 2 to 4.6% (Moretta, 2001). To place this in context, other more valuable sandalwood species, *S. yasi*, and *S. austrocaledonicum*, produce approximately 5% and 3-5% oil respectively (Applegate et al., 1990). Understanding the natural variability of heartwood oils in *S. spicatum* and the origin is important for future tree improvement programs to align plantations with market demand.

1.3. **West Australian Sandalwood (Santalum spicatum)**

1.3.1. **History, Importance and Economic Value**

West Australian sandalwood (*Santalum spicatum*) is a commercial species that contributes significantly to the market for sandalwood products in the form of joss
(incense) sticks, extracted oil, carvings and religious ornaments (Tonts and Selwood, 2003). Historically, the fragrant timber was harvested from the southern parts of the West Australian ‘wheatbelt’ region and exported to China, India, and the Dutch East Indies for religious and artistic use (Loneragan, 1990; Tonts and Selwood, 2003). By 1848, the trade of *S. spicatum* had become the primary export earner for Western Australia (Loneragan, 1990). While this industry was largely overtaken by mining (the gold rush) and farming in the twentieth century, a small industry persisted which has re-emerged in the past 10 years to fill niche markets on the global scale, for exotic and specialised commodities ($11.6 million in 1999-2000) (Anon., 2000; Tonts and Selwood, 2003; Woodall and Levinson, 2006; Woodall and Robinson, 2003). Research into sandalwood has also focused on the ability of the oils to treat skin and urinary tract diseases, inflammation, bronchitis, cystitis, the *Herpes simplex* virus, *Candida spp* and the resistant *Staphylococcus aureus* strain (Benencia and Courrèges, 1999; Dwivedi and Zhang, 1999; Hammer et al., 1998; Rai, 1990; Tinson, 2001). The development of plantations of *S. spicatum* in the West Australian wheatbelt (and *S. album* in the North West of the state) is beneficial as an economic strategy as it provides an extra source of income for rural businesses, farmers, local communities and allows diversification of local income (Tonts and Selwood, 2003).

On-farm production of *S. spicatum* has added environmental benefits. Clearing for agriculture in the South-West Australian wheatbelt has significantly contributed to problems of soil salinisation, erosion and loss of biodiversity, which has impacted local flora and fauna communities (Tonts and Selwood, 2003; Woodall and Robinson, 2003). Plantations, which include the hosts, offer additional habitat for wildlife, and regulates soil salinity by providing a deep-rooted perennial species to help contribute to the
ecology of this vegetatively-fractured region (Woodall and Robinson, 2003). Farmers currently benefit from on-farm conservation goals from these plantations but, to fully benefit from economic diversification, further research is still required to ensure product quality and quantity during the plantation rotation of 25 years. Greater knowledge of the mechanisms involved in heartwood profile development and total oil production will insure greater plantation value to encourage wider conservation plantings.

_Santalum spicatum_ wood is currently sourced from natural populations. Regeneration of these natural stands has not been successful due to low rates of natural germination and this has been attributed to climate change over the last 100 years, grazing by feral goats and susceptibility to fires (Loneragan, 1990). To harvest a sandalwood tree, the entire sandalwood tree is pulled from the ground since the roots, butt, stem and branches all contain extractable oil which also varies in sesquiterpene composition across these sections (Loneragan, 1990; Piggot et al., 1997). Although plantations will be the preferred source for sustainable sandalwood in the future, there are economic advantages for sourcing wild trees.

Heartwood-oil production in _S. spicatum_ initiates late in the development of sandalwood trees, with detectable yields at approximately 10 years (Brand et al., 2003) and accumulates as the tree ages (Brand and Pronk, 2011). Plantations may require at least 25–30 years before producing enough heartwood oil to be commercially viable. Wild-sourced trees are thought to be older than 80 years as their heartwood fills the bole of the tree producing adequate amounts of extractable oil. The long term aim for the sandalwood industry is to improve plantation profitability through the introduction of
superior plant material which will ensure high quality, economically viable oil-producing plantations.

1.3.2. **The Sandalwood Tree**

*Santalum spicatum* is a slow-growing, hemi-parasitic tree or shrub which reaches up to 8 m in height (Loneragan, 1990). This native species from Western Australian naturally grows in a climate which ranges from the semi-arid south of WA (~ 300–600 mm rainfall) to the arid northern areas (~150–300 mm rainfall) and South Australia (Fox and Brand, 1993, Hewson and George, 1984, Loneragan, 1990). In Western Australia, it is absent from high rainfall regions (rainfall > 650 mm) (Woodall and Robinson, 2003). Due to irregular rainfall and temperature variations across its natural distribution, flowering and seed germination is sporadic (Loneragan, 1990). Like other *Santalum* species, *S. spicatum* is hemi-parasitic in nature, obtaining some growth nourishment from nitrogen-fixing host species, such as *Acacia* or *Allocasuarina* (Herbert, 1925; Loneragan, 1990; Woodall and Robinson, 2003). *Santalum spicatum* mostly grows in low shrub-lands of *Acacia* species and on the borders of *Eucalypt* woodlands, avoiding woodlands of almost pure species. Distribution through other regions is scattered and irregular (Loneragan, 1990; Woodall and Robinson, 2003) and has been reduced in the south-west and goldfields regions largely due to land clearing for agriculture (Woodall and Robinson, 2003). Sandalwood trees display considerable variation in their height, form, leaf shape, colour and fruiting patterns across their geographic range. Such variations are a likely result of both genetic and environmental factors which can
include temperature, rainfall, soil composition and biotic factors such as tree age and interactions with the host. This may all contribute to total heartwood oil production and influence the observed variation in oil extracts within the species. Trees in the dry arid regions of Western Australia have been shown to contain greater amounts of oil in the wood than in semi-arid areas (Loneragan, 1990; Brand et al., 2003; Piggot et al., 1997). Within *S. spicatum*, the tree-to-tree variation in heartwood composition offers the possibility to improve oil quality through selection and domestication.

1.4. Knowledge Gaps

1.4.1. Oil Production

Many plants synthesise a large and structurally diverse suite of isoprenoid natural-products which are involved in both primary and secondary metabolism (Chappell, 1995; Kirby and Keasling, 2009). The primary metabolites are essential for plant growth and development and include sterols, carotenoids, quinones, strigolactones, cytokinins, brassinosteroids, gibberellins and abscisic acid (Kirby and Keasling, 2009; Wu and Chappell, 2008). Isoprenoid secondary natural products are more structurally diverse and often species-specific, contributing to the chemical arsenal of plants for roles in defence against competitors and communication including signalling to other plants, attraction of pollinators or herbivore predators (Bohlmann et al., 1998a; Engelberth et al., 2004; Erbilgin et al., 2006; Gershenzon and Dudareva, 2007; Lange and Croteau, 1999; Messchendorp et al., 2000; Schnee et al., 2006). In sandalwood, the
accumulation of sandalwood oil in the heartwood likely constitutes a defence mechanism given its innermost location and anti-microbial properties (Hammer et al., 1998; Jirovetz et al., 2006). This suggests a long term defence strategy and could serve several purposes throughout the life of the plant. Parallels may be drawn to for example the well studied conifer models which produce complex oleoresin mixtures with multi-function defence against insect herbivores and fungal pathogens (Bohlmann et al., 1998).

All isoprenoids, or terpenoids, are classified based on the number of carbon atoms they contain; with monoterpenes (C$_{10}$), sesquiterpenes (C$_{15}$), diterpenes (C$_{20}$), and triterpenes (C$_{30}$); and all are synthesised in specific cell compartments (Kirby and Keasling, 2009). Terpenoids are derived from the two C$_{5}$ building blocks, isopentyl diphosphate (IPP) and its’ isomer dimethyl allyl diphosphate (DMAPP) which are produced in either of two pathways: the cytosolic mevalonate (MEV) pathway or the 1-deoxy-D-xylulose-5-phosphate (MEP) pathway in plastids (Rodríguez-Concepción and Borona, 2002; Kirby and Keasling, 2009). Condensation of DMAPP with one, two or three units of IPP via the action of prenyltransferases generates the longer isoprenoid precursors, geranyl diphosphate (GPP) for monoterpenes, farnesyl diphosphate (FPP) for sesquiterpenes, and geranylgeranyl diphosphate (GGPP) for diterpenes. Production of terpenes is mostly compartmentalised with mono- and diterpene synthesis occurring in the plastids via the MEP pathway, and sesqui- and triterpenes being generated in the cytosol through the MEV pathway. Cross talk between the two pathways has been reported, however this is not significant enough to rescue a block in either pathway (Bick and Lange, 2003). Many genes within the MEV and MEP pathways regulate the flux to terpene compounds (Rodríguez-Concepción and Borona, 2002) and are therefore
important to consider for future studies into oil production and its regulation in *Santalum*.

IPP and DMAPP are polymerised into their respective branch point intermediates GPP, FPP and GGPP, substrates for a large family of enzymes, the terpene synthases (TPS). TPSs are classified into subfamilies, TPSa-g within the angiosperm or gymnosperm clades (Bohlmann et al., 1998b; Bohlmann et al., 1999; Martin et al., 2004). This class of enzyme is largely responsible for the structural variety of terpene metabolites in the plant kingdom (Tholl, 2006). The reaction mechanism proceeds via a divalent cation dependent ionisation of the prenyl diphosphate to a cationic enzyme bound intermediate. Depending on the enzyme, the electrophilic intermediate may undergo a variety of rearrangements including recapture of the diphosphate, internal cyclisations, Wagner-Meerwein methyl shifts, and hydride shifts. TPS reaction mechanisms ordinarily terminate with a final deprotonation of the bound cation from one or more positions to yield olefinic hydrocarbons or water capture to produce alcohols. This variability in the terminal reaction step allow multiple product generation from individual active sites. Variations on these mechanisms among TPSs are responsible for the wide variety of terpenoid carbon skeletons derived from the universal acyclic substrates GPP, FPP and GGPP (Davis and Croteau, 2000; Green et al., 2011; Kollner et al., 2004). The active site shape largely determines product profiles, which has been confirmed by site directed mutagenesis studies (Green et al., 2011; Kollner et al., 2004). Sequence differences across sandalwood species may provide DNA markers for selection of superior genotypes.
Another class of enzymes which can modify terpenoid structures are the heme-containing cytochrome P450 monooxygenases (CYP or P450) (Nelson et al., 1996). This superfamily of catalysts is likely to help generate the chemical complexity and diversity of sesquiterpenols in sandalwood oil. For example, the C-12 hydroxylation of the santalenes to the commercially-significant α- and β-santalol, is mediated by a cytochrome P450 (Diaz et al., 2013, Chapter Five, this thesis). Identification of the relevant genes for sesquiterpene biosynthesis in *S. spicatum* would enable strategies for improving production.

### 1.4.2. Future Possibilities for Tree Improvement

There are several options available for sourcing natural products which have applications in agriculture, medicine and aromatherapy, as well as in flavour and fragrance industries. These include extracting compounds, oils, resins from natural resources (Adams, 1995), total or partial chemical synthesis of target molecules (Castro et al., 2004), improvement of existing plant material through selection and breeding (Croteau et al., 2005), genetic engineering (Kim and Keasling, 2001; Mahmoud and Croteau, 2001) and *in vitro* production using microbial hosts (Ajikumar et al., 2008). The essential oil of sandalwood is currently extracted from natural resources using steam distillation and solvent extraction techniques (Brand et al., 2007). Depletion of natural resources however, supports the need to explore additional methods for tree improvement to both supplement the sandalwood industry and aid conservation efforts. Chemical synthesis of the santalols has been reported (Brocke et al., 2008; Christenson and...
Willis; 1980, Muratore et al., 2010) however, several low-recovery steps makes this synthetic approach inefficient for large-scale production.

Improvement of natural products in plants through selection and breeding of desired chemotypes has been successful in species such as basil (*Ocimum, Lamaniaceae*), which have used key chemical and morphological characteristics to select superior chemotypes (Simon et al., 1990). To increase the abundance of santalol in oil profiles and reduce the undesirable farnesol content via future breeding programs a similar approach could be adopted for *S. spicatum* in future breeding programs, seed orchards and plantations, once a better understanding of the natural chemical-diversity across its range is achieved.

Alternatively, the opportunity to engineer higher levels of oil in the sandalwood tree or use microbial hosts to produce desired components becomes possible if the pathway genes involved in sesquiterpene biogenesis are known. In sandalwood, a number of terpene synthase and cytochrome P450s genes have been characterised (Diaz-Chavez et al., 2013; Jones et al., 2008; Jones et al., 2011; Chapters Three, Five and Six, this thesis) and form the basis for studies into alternative methods to improve heartwood-oil quality and total content. For example, in *E. coli*, transgenic manipulation of the DXP pathway successfully increased carotenoid and ubiquinone levels by 40% (Harker and Bramley, 1999) and in peppermint (*Mentha x piperita* L.), transgenic plants accumulated approximately 50% greater yields of essential oils (Mahmoud and Croteau, 2001). Although these provide examples of MEP pathway genes, both studies demonstrate the possibility of regulating essential oil quality and quantity. In *S. spicatum* the chance to
influence quality may lead to reduced levels of the allergenic farnesol, thus improving market value.

Finally, microbial engineering of essential oils could be a means for industrial-scale production of desirable sesquiterpene-blends in sandalwood (Kirby and Keasling, 2009). For example, metabolic engineering of *E. coli* has enabled the production of high levels of diterpenes (in the range of 10 to 100 mg L\(^{-1}\)) using bioreactor-grown cultures. This was achieved through incorporation of a heterologous MEV pathway. High yields were also obtained when the MEP pathway was enhanced by overexpressing key pathway genes (Morrone et al., 2010). In *Santalum*, further developing the sandalwood industry using new tools will be possible once there is a better understanding of the pathway genes involved in heartwood oil production and their regulation.

1.4.3. Contributions to total oil production

*Santalum spicatum* is an obligate hemi-parasite which takes up water and nutrients from a host species by means of specialized structures known as haustoria (Herbert, 1925; Loneragan, 1990). Plant health, growth and heartwood oil production may all be influenced in part by interactions between a parasite and its host, however, limited information about the ecology of plant/host interaction is known in sandalwood. In regards to oil production, it is unknown whether the influence of the environment is the dominant factor or if the observed quantitative variation in heartwood oils is controlled
principally by genetic factors. In Pinus monticola Doug for example, the levels of terpenes and total terpenes extracted have been reported as strongly associated with genotype (Hanover, 1966, Wallis et al, 2011, Sampedro et al, 2010). Genotypically identical plants (clones) grown in diverse environments did not show a differences in terpene levels, indicating that oil composition in this species was stable with respect to a changing environment (Hanover, 1966). In S. spicatum, the influence of the environment and host species on oil production, much like the role of genetics, has received little attention. The nature and number of hosts S. spicatum attaches to is highly variable in its natural habitat (Woodall and Robinson, 2003) and each host contributes differently in regards to the major solutes and nutrients it provides (Radomiljac et al., 1999a; Radomiljac et al., 1999b). Leguminous hosts are generally considered better hosts in terms of growth and development because of their nitrogen contribution through nitrogen fixation (Radomiljac et al., 1998). It has been established that S. spicatum growth and survival is relatively high when parasitized to Acacia plants (A. acuminata and A. saligna) compared to other species such as Allocasuarina huegeliana and Eucalyptus loxophleba which do not promote the same level of growth (Brand et al., 2003; Brand et al., 1999). Secondary metabolites and chemical stimulants generated by a host species affect haustorial development and underground growth of the parasites roots (Cook et al., 1966). Manipulation of a host species may play a significant role in growth and heartwood oil production in Santalum and currently, the contribution of hosts and other environmental conditions (soil composition, rainfall, temperatures) in the natural system has not been explored to a great extent. This project will pave the way to find a key tool to quickly evaluate the effect of host and climatic conditions on sandalwood.
1.5. Research Hypotheses and Brief Outline of Research Chapters

The issues of heartwood-oil variability, oil biosynthesis and chemotype heritability in *S. spicatum* (with one chapter on *S. album*), led to the testing of several main hypotheses:

i. **HYPOTHESIS 1:** *S. spicatum* essential oil is variable in terms of total oil content and chemical composition across its natural range (Chapters Two and Four). Chapter Four compares the oil quality (farnesol/santalol amounts) and total oil content between wild and 11 year old plantation trees to identify differences in quality as well as summarising the oil biosynthetic genes from just *S. spicatum*. Chapter two aims to characterise heartwood composition, including other major components, across the wheatbelt (where most trees were sampled), with additional individuals from the goldfields and north regions of WA, to see whether patterns in oil composition were evident across the species' natural range.

ii. **HYPOTHESIS 2:** Major heartwood oil components in Santalum are synthesized via specific terpene synthases (TPS), and orthologous TPS genes are present in divergent species (*S. spicatum*, *S. album* and *S. austrocaledonicum*). Differences in these gene sequences may help to explain the compositional variation in heartwood components across the genus (Chapters Three and Six).

iii. **HYPOTHESIS 3:** Within *S. spicatum*, disparities in oil profiles may be explained by mutations in terpene synthase gene sequences, giving rise to allelic TPS-variants which may be present in the natural population (Chapter Six).
To test these main hypotheses which link to oil diversity, chapter three of this thesis describes the characterisation of three key TPS genes from phylogenetically divergent sandalwood species; *S. spicatum* (Jessie Moniodis), *S. album* (laboratory work performed by Dr. Chris Jones) and *S. austrocaledonicum* (laboratory work performed by Dr. Katherine Zulak), which includes characterising a santalene synthase. Chapter six of this thesis is an extension of the TPS gene-mining in *S. spicatum*, which sought to find additional sequences involved in total oil production. An EST library was constructed from the xylem tissue and sequences from the MEV, DXP pathways, cytochrome P450 superfamily, allylic phosphatases and terpene synthases were mined. An additional multi-product terpene synthase was characterised from *S. spicatum* using the generated sequence information. This fourth TPS sequence was used to discover whether allelic variants were present in a small-subsample of the population. Allelic variants may be able to explain some of the natural variability in terpene levels in *S. spicatum*.

iv. **HYPOTHESIS 4**: Another enzyme class present in Santalum, the cytochrome P450s, introduce additional chemical diversity by modifying terpene substrates; and that these enzymes hydroxylate the santalenes to produce the valuable α- and β-santalol sandalwood oil components (Chapter Five). Chapter five of this thesis describes the characterisation of a group of CYPs (in the CYP76 family) which were mined from an *S. album* 454 and EST database, capable of converting α- and β-santalene to their hydroxylated analogues.

v. **HYPOTHESIS 5**: Individuals which group together genetically (measuring allelic variability between individuals using microsatellites) will share more
similarities in heartwood-oil composition or trees exposed to similar environments in Western Australia - given the diverse growing conditions across the state - will also contain more similar oil characteristics (Chapter Seven).

Chapter seven investigates the genetic structure of *S. spicatum* across the wheatbelt, with additional trees from the goldfields and north regions of WA using neutral genetic markers (microsatellites). A script was written in a statistical program (R) to determine whether any significant correlations existed between genotype, chemotype, morphological measurements (tree height, diameter) and environmental variables (soil composition, pH, number and species of surrounding hosts) to address the above questions.
1.6. References


Byrne, M., Macdonald, B., Brand, J.E., 2003b. Regional genetic differentiation in Western Australian sandalwood (Santalum spicatum) as revealed by nuclear RFLP analysis. Theoretical and Applied Genetics 107, 1208-1214.


Clarke, M., 2006. Australia's sandalwood industry, an overview and analysis of research needs. Rural Industries Research and Development Corporation (RIRDC): Publication No 06/131


Green, S., Baker, E.N., Laing, W., 2011. A non-synonymous nucleotide substitution can account for one evolutionary route to sesquiterpene synthase activity in the TPS-b group. FEBS letters 585, 1841-1846.


Herbert, D.A., 1925. The root parasitism of Western Australian Santalaceae. Journal of Royal Society Western Australia 11, 127-149.


Moretta, P., 2001. Extraction and Variation of the Essential Oil from Western Australian Sandalwood (*Santalum spicatum*). Australia: The University of Western Australia, Doctor of Philosophy (PhD).


Schnuch, A., Uter, W., Geier, J., Lessmann, H., Frosch, P.J., 2004. Contact allergy to farnesol in 2021 consecutively patch tested patients. Results of the IVDK. Contact Dermatitis 50, 117-121.


CHAPTER TWO

Essential Oil Variation in West Australian Sandalwood

(Santalum spicatum R. Br. DC)

Moniodis J\textsuperscript{a,b,c*}, Jones CG\textsuperscript{a}, Renton M\textsuperscript{a}, Plummer JA\textsuperscript{a}, Barbour EL\textsuperscript{a}, Ghisalberti EL\textsuperscript{b} and Bohlmann J\textsuperscript{c}

\textsuperscript{a} School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley WA 6009 Australia

\textsuperscript{b} School of Chemistry and Biochemistry (M310), University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

\textsuperscript{c} Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T1Z4, Canada

* corresponding author
Santalum spicatum heartwood oil composition varies substantially within and between northern, south-eastern and south-western regions of Western Australia. Total oil yields were similar across the three regions. Trees from the north contained higher amounts of the most valuable fragrance compounds α- and β-santalol and less of the undesired E,E-farnesol than trees from south-eastern regions. Several high quality trees were also found in the south-west where wild harvesting of sandalwood still occurs.
2.1. Abstract

West Australian sandalwood (*Santalum spicatum*) has long been exploited for its fragrant, sesquiterpene-rich heartwood. Essential oil quality varies substantially, which is of interest to the sandalwood industry. We investigated trees from the arid northern and south-eastern and semi-arid south-western regions of West Australia for patterns in oil composition and co-occurrence of sesquiterpenes. Total oil content was similar across the entire sample collection; however sesquiterpene composition was highly variable. Northern populations contained the highest levels of desirable fragrance compounds, α- and β-santalol, as did individuals from the south-west. South-eastern populations were higher in \(E,E\)-farnesol, an undesired allergenic constituent, and low in santalols. These trees generally also contained higher levels of α-bisabolol. \(E,E\)-farnesol co-occurred with dendrolasin. Contrasting α-santalol and \(E,E\)-farnesol chemotypes revealed potential for future genetic tree improvement. Although chemical variation was evident both within and among regions, variation was generally lower within regions. Cluster analysis of 13 dominant sesquiterpenes indicated that at least four separate enzymes contribute major oil constituents. Our results showed distinct patterns in chemical diversity of *S. spicatum* across its natural distribution, consistent with earlier investigations into sandalwood population genetics. These results are relevant for plantation tree improvement and conservation efforts.
Keywords

Sandalwood, *Santalum spicatum*, *Santalum album*, sesquiterpene fragrance, α-santalol, β-santalol, *E,E*-farnesol, naturally occurring chemical variation, chemical diversity

### 2.2. Introduction

West Australian sandalwood (*Santalum spicatum*) belongs to the genus of hemi-parasitic *Santalum* (Santalaceae) trees widely exploited for their fragrant heartwood which is used in perfumes, pharmaceuticals, incense and ornamental carvings. The essential oil contained in the heartwood of mature sandalwood trees (> 10 years) consists of a complex mixture of sesquiterpenoids, with unique compositions apparent across, and occasionally within species (Jones et al., 2006; Moretta, 2001; Page et al., 2010). Historically, Indian sandalwood (*S. album*) has provided the bulk of sandalwood products; however, *S. spicatum* is frequently used as a supplement to incense matrix and to a lesser extent as extracted oil. The international standard for *S. album* oil requires 41% to 55% α-santalol and 16% to 24% β-santalol (Howes et al., 2004; ISO, 2002). At present, *S. spicatum* does not yield an oil which meets these industry standards for two reasons; the combined santalol content is too low, and levels of *E,E*-farnesol, a suspected allergen, are too high (Brand and Pronk, 2011; Hostynek and Magee, 1997; Schnuch et al., 2004). The oil of *S. spicatum* is generally considered less valuable than that of *S. album* due to the lower overall oil content and more variable sesquiterpene composition (Brand and Pronk, 2011).
*S. spicatum* is traditionally harvested from natural stands of undomesticated tree populations throughout south western Australia. To reduce pressure on wild stands, plantations have been established on former grazing and marginal agricultural land. Significant potential exists for the improvement of *S. spicatum* plantations through selection of trees with desired growth and essential oil characteristics. By first quantifying the extent and nature of essential oil variation throughout the tree’s area of distribution, key selection parameters may be specified. Moreover, by better understanding the process of essential oil formation in sandalwood, relevant molecular markers may be developed to speed up the selection process. Continued development of *S. spicatum*, a species naturally adapted to the arid-zone of Western Australia, will further conservation goals as well as improve its commercial potential.

The oil of *S. spicatum* contains over 100 different sesquiterpenes (Adams et al., 1975, Brophy et al., 1991, Valder et al., 2003, Baldovini et al., 2011). The mixture of sandalwood sesquiterpenes are likely to serve as protection against wood-rotting fungal pathogens, given the oil is localised specifically to the innermost heartwood (Jones et al. 2008). The most abundant compounds of the *S. spicatum* oil are the sesquiterpene alcohols *E,E*-farnesol, α- and β-santalol, lanceol, nuciferol and α-bisabolol, as well as a variety of sesquiterpene olefins such as the santalenes, bergamotene, and several curcumenes (Brophy et al., 1991; Valder et al., 2003). Moretta (2001) studied the heartwood cores of 87 *S. spicatum* trees from 12 geographic sites in Western Australia and, in the form of a PhD thesis, reported the levels of α- and β-santalol to vary from 3% to 67% percent, and *E,E*-farnesol to range from 5% to 30% across the entire distribution (Fig. 2.1). Although variation in major oil components has been reported for *S. spicatum*, there have been no further investigations into the oil composition across
populations within its natural distribution. A more comprehensive evaluation of the extent and nature of variability of wild *S. spicatum* trees is needed to inform efforts to increase total santalol content through tree improvement.

![Figure 2.1 - Biosynthesis of major sesquiterpene alcohols in *S. spicatum*. Proposed formation of farnesol from farnesyl diphosphate (FPP) via a terpene synthase (TPS) or phosphatase activity, and the known pathway for santalol biosynthesis in sandalwood via TPS and cytochrome P450 (P450) activities.](image)

Chemical variation in *S. spicatum* might be explained by genetic differences amongst populations, while the species is also growing across a range of diverse environments. Discrete chemical differences in heartwood composition appear to exist throughout the species’ natural distribution (Moretta, 2001) and several populations have been shown to bear strong genetic similarities (Byrne et al., 2003a; Byrne et al., 2003b). *S. spicatum* is distributed throughout the semi-arid southern regions (300 to 600 mm annual precipitation) and arid northern areas of Western Australia (150 to 300 mm), with large differences in soil moisture levels, soil types, root haustorial connectivity, as well as
presence of pests and pathogens (Applegate et al.; 1990). These biotic and abiotic environmental factors may influence the oil profile of a tree.

In this study, the diversity and variation of sesquiterpenes of *S. spicatum* heartwood was characterised and examined for patterns indicative of unique chemotypes. Heartwood core samples were taken from 194 trees across the three main regions where *S. spicatum* naturally occurs. While the majority of samples were taken from the south-west (aka Wheatbelt region) of Western Australia, where a long history of harvesting is known, chemical analysis was extended to populations in the arid north and the south-east (aka Goldfields region) of Western Australia (Fig. 2.2). It was hypothesised that chemically distinct populations of *S. spicatum* would be present, and that the chemical diversity of heartwood oil would be lower within populations, while being higher between populations.

2.3. Experimental

2.3.1. Substrates and reagents

Chemicals, substrates and reagents were purchased from Sigma Aldrich (St. Louis, MO, USA) unless otherwise indicated.
2.3.2. Plant Material Collection

Collection of *S. spicatum* plant material was done in the spring of 2009. Plant material was obtained from trees growing throughout the southern half of Western Australia. Samples collected from the south-west Wheatbelt (N = 152), south-east Goldfields (N = 19) the Northwest (N = 23) (Fig. 2.2). Xylem was collected by drilling into the stems of trees at 30 cm stem height above ground level with a 25 mm drill bit and cordless drill. Only trees with a stem diameter greater than 7 cm were sampled. Wood shavings to be extracted for oil analysis were air dried in paper bags for two weeks.

Figure 2.2 - Map of Western Australia showing locations for collection of 194 *S. spicatum* heartwood-cores for chemical analysis
2.3.3. S. spicatum oil extraction and GC-FID and GC-MS

Oil was extracted from wood shavings (3-5 g) in 25 ml volumetric flasks using ethanol for 7 – 14 days. 10 mM isobutyl benzene (IBB) was added as an internal standard for quantification. The essential oil was separated by gas chromatography (GC) and quantified by flame ionization detection (GC-FID) and identification of components was done using mass spectrometry (MS) (Shimadzu). An external standard plot using a dilution series of authentic S. spicatum oil (Mt. Romance, Albany WA) was generated to determine total oil yield from extracts. Gas chromatography conditions were as follows: stationary phase; DB-Wax column, 30 m x 0.25 mm ID x 0.25 µM film thickness (Agilent). Carrier gas was helium at 1 ml per minute. One microliter of extract was injected using a split ratio of 10:1. Injector temperature; 200 °C, detector was set to 250 °C. Oven temperature started at held at 40 °C and raised at 3 °C for to 220 °C where it was held for a further 25 min. Column conditions for GC-MS were similar to that of GC-FID, except the detector was turned on 5 min after injection. Scan mode was used over the range of 41 to 250 m/z.

2.3.4. Statistical analysis

Hierarchal cluster analysis was performed to explore co-occurence patterns of major oil constituents across the 194 S. Spicatum individuals using Primer 6 (Clarke, 1993, Clarke and Gorley, 2006). Oil data was standardised prior to analysis. A Bray-Curtis
similarity was performed for construction of the dendogram using group and single-linkage results (both were comparable). Principal Component Analysis (PCA) was performed in R (The R Team, 2013) using percentage composition data to explore which individuals were similar and different in terms of sesquiterpene composition, investigate correlations in oil components, and to identify possible *S. spicatum* chemotypes.

### 2.4. Results and Discussion

#### 2.4.1. Variability of total oil content

Total essential oil yield from heartwood of *S. spicatum* individuals sampled at a stem height of 30 cm above ground was not significantly different across the three regions (Table 2.1). Trees in the Wheatbelt (3.22 ± 1.08%, N = 152), Goldfields (2.40 ± 0.94%, N = 19) and North (3.18 ± 1.15%, N = 23) contained similar total oil contents, with individuals ranging from 0.01 to 6.41 % total oil by dry-weight. High standard deviations suggest oil-yield variation is highest between individuals, rather than among regions. Reported oil contents are similar to previous studies on mature *S. spicatum* trees from natural stands (Brand et al., 2007; Moretta, 2001). Tree to tree variation in total terpene concentrations has also been reported in other terpene accumulating species (e.g., Egerton-Warburton et al., 1998). The observed variation in *S. spicatum* can be explained in part by tree maturity, as higher heartwood oil concentrations are more common in older trees (Brand and Pronk, 2011). Mean tree-age should be lower in more frequently harvested regions. Sandalwood in the Wheatbelt and Goldfields has
been more heavily exploited for a longer period of time than those in the North. Biometric parameters such as diameter and tree height or growth rings do not appear to be reliable proxies for tree age in wild stands of *S. spicatum*, given the broad distribution of the species, the wide range of growing conditions and the hemi-parasitic nature of sandalwood (Byrne et al., 2008; Loneragan, 1990; Steffen et al., 2009). Other oil-bearing sandalwoods such as *S. austrocaledonicum* have also displayed variation in total oil yield (Page et al., 2010). Even *S. album*, the oil of which can consistently contain more than 80% α- and β-santalol (Verghese et al., 1990) varies widely in oil yield from no detectable oil to nearly 9% by dry weight. Tree age has been reported to be a significant component of oil yield variation in plantations of *S. album* (Jones et al., 2006). However, genetic and environmental factors contributing to heartwood development in general and oil production in particular in any *Santalum* species are poorly understood.
Table 2.1 - Sesquiterpene composition of *S. spicatum* heartwood oil from three different regions of West Australia (Wheatbelt, Goldfields and North)

<table>
<thead>
<tr>
<th>Region</th>
<th>South West of W.A. (Wheatbelt)</th>
<th>South East of W.A. (Goldfields)</th>
<th>North of W.A. (Carnarvon/Shark Bay)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Components</strong></td>
<td>Average (%)</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td><em>E,E</em>-farnesol</td>
<td>14.4</td>
<td>2.0</td>
<td>48.2</td>
</tr>
<tr>
<td>Total α- and β-Santalol</td>
<td>12.5</td>
<td>0.9</td>
<td>54.3</td>
</tr>
<tr>
<td><em>Z</em>-nuciferol, <em>Z</em>-Y-cumene-12-ol</td>
<td>11.2</td>
<td>3.1</td>
<td>23.3</td>
</tr>
<tr>
<td><em>α</em>-santalol</td>
<td>9.3</td>
<td>0.6</td>
<td>41.3</td>
</tr>
<tr>
<td><em>Z</em>-β-cumene-12-ol</td>
<td>6.9</td>
<td>0.3</td>
<td>17.1</td>
</tr>
<tr>
<td><em>β</em>-santalol</td>
<td>3.2</td>
<td>0.3</td>
<td>13.0</td>
</tr>
<tr>
<td><em>Z</em>-α-trans-bergamotol</td>
<td>2.8</td>
<td>0.0</td>
<td>14.0</td>
</tr>
<tr>
<td><em>α</em>-bisabolol</td>
<td>1.8</td>
<td>0.0</td>
<td>18.4</td>
</tr>
<tr>
<td><em>Z</em>-lanceol</td>
<td>1.4</td>
<td>0.3</td>
<td>8.7</td>
</tr>
<tr>
<td><em>α</em> -dendrolasin</td>
<td>1.3</td>
<td>0.2</td>
<td>5.3</td>
</tr>
<tr>
<td><em>epi</em>-β-santalol</td>
<td>0.6</td>
<td>0.0</td>
<td>1.9</td>
</tr>
<tr>
<td><em>α</em>-santalenol</td>
<td>0.5</td>
<td>0.0</td>
<td>2.1</td>
</tr>
<tr>
<td><em>trans</em>-nerolidol</td>
<td>0.4</td>
<td>0.0</td>
<td>1.7</td>
</tr>
<tr>
<td><em>β</em>-santalenol</td>
<td>0.4</td>
<td>0.0</td>
<td>1.6</td>
</tr>
<tr>
<td><em>E</em> -<em>β</em>-farnesene</td>
<td>0.3</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td><em>epi</em>-β-santalenol</td>
<td>0.2</td>
<td>0.0</td>
<td>1.1</td>
</tr>
<tr>
<td><em>α</em>-trans-bergamotene</td>
<td>0.2</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>54.8</td>
<td>24.6</td>
<td>81.9</td>
</tr>
<tr>
<td><strong>Yield (%)</strong></td>
<td>3.22</td>
<td>0.01</td>
<td>6.41</td>
</tr>
</tbody>
</table>

2.4.2. **Essential oil quality and geographic variation**

The heartwood essential oil composition of *S. spicatum* trees sampled in this study varied both within and between regions (Table 2.1). The two main sesquiterpene alcohols imparting value to the oil, α- and β-santalol, were found to be highest in trees from the North (33.0 ± 11.6% santalols), followed by the Goldfields (12.5 ± 10.9%). Wheatbelt trees contained oil with the least amount of santalols (10.3 ± 7.0%). Large standard deviations are indicative of the natural variation in santalol content, which across the three regions ranged from 0.8% to 55.6% of the total oil makeup. Approximately one quarter of all trees surveyed comprised less than 5% combined α- and β-santalol levels, which is considered very low quality (Howes et al., 2004) and
more than half the trees had oil with less than 10% to 15% combined santalol content (Fig. 2.3). A few individuals (0.5%) had oil with greater than 50% total santalol content, mainly those from the North. These levels are more similar to those found in *S. album* heartwood. Although the sample size was small for the North (N = 23) and the Goldfields regions (N = 19), all tree sampling was done over a wide area at each site in an effort to minimize sample selection bias. All individuals from the North contained at least twice as much α- and β-santalol as farnesol in the extracted oil. No individuals in this study reached the ISO standard (ISO, 2002) for *S. album* (41% to 55% α-santalol and 16% to 24% β-santalol), however several high quality trees were identified, with one northern tree containing 55.6% combined α- and β-santalol content and less than 5% farnesol. Despite the oil composition varying across the natural distribution, the identification of *S. spicatum* trees with high levels of α- and β-santalol and simultaneously low levels of farnesol was important as it indicates the potential to select individual trees with superior oil from these populations for future improvement programs.

Trees from the North of Western Australia contained generally lower levels of *E,E*-farnesol than trees from the Wheatbelt and Goldfields regions, and variation within regions was also evident. Trees from the Goldfields had on average the highest amounts of *E,E*-farnesol (19.4 ± 5.6%) followed by trees from the Wheatbelt (14.4 ± 8.5%), while those in the north had the least (9.1 ± 5.4%). Variation in *E,E*-farnesol content across all regions was high, comprising anywhere between 2.0% and 46.2% of the total oil profile. A frequency distribution of *E,E*-farnesol chemotypes highlights this variation (Fig 2.3). Ten percent of sampled trees contained high levels of farnesol (> 35%) while trees with less than 5% farnesol were similarly infrequent. Some
chemotypes of sandalwood appear to exist with high or low levels of $E,E$-farnesol, however the bulk of the distribution (more than half of the trees analysed) contained between 10% and 35% $E,E$-farnesol, suggesting a continuous distribution. It is perhaps unsurprising that distinct chemotypes of *S. spicatum* could exist in different regions of Western Australia; Byrne et al. (2003b) identified that the northern arid populations of *S. spicatum* were genetically distinct from the southern semi-arid populations (including the Wheatbelt). Moretta (2001) also suggested chemotypes of *S. spicatum* may exist in geographically distinct pockets of Western Australia. Of the 12 regions sampled, Moretta (2001) found $E,E$-farnesol to be lowest in the north-west region (Shark Bay) (< 5%) and highest (0% to 35%) in a south-west population (Katanning) of Western Australia. Additionally, Moretta (2001) reported that three populations from the northern parts of Western Australia had the highest average santalol content compared to other regions of the state. Despite limited by small sampling size, these results are in strong agreement with the findings of the present study.
Figure 2.3 - Frequency of *S. spicatum* individuals with different relative amounts of (A) α- and β-santalol concentration and (B) *E,E*-farnesol in heartwood samples of 194 trees growing in natural stands in Western Australia: 152 trees were from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the North regions (Carnarvon and Shark Bay).

Future efforts into tree improvement of *S. spicatum* for plantations would benefit from an understanding of the genetic and molecular underpinning of the accumulation of *E,E*-farnesol and its variation, as it directly impacts on the quality and marketability of the extracted oil (Adams et al., 1975; Brand et al., 2007; Brophy et al., 1991; Howes et al., 2004). Moreover, abiotic environmental conditions such as temperature, rainfall and soil structure, as well as biotic factors including tree age, growth rate, neighbouring host species and disease prevalence vary widely across sandalwood distribution, and are likely to contribute to variation in oil profiles between trees (Applegate et al., 1990). To
date, little research has focused on specific environmental, genetic and physiological factors and their function in oil production largely due to difficulties in maintaining suitable controls in field experiments. Clonally propagated sandalwood would reduce genetic variation and would allow for the testing of environmental factors affecting oil production; however, *S. spicatum* is difficult to clonally propagate from somatic tissue (McComb and Jones, 1998). Future research directed towards understanding the genetic and environmental origins of chemical variability within this species will aid tree improvement programs which can assist in the selection of superior genotypes with increased production of α- and β-santalol and conversely, low amounts of *E,E*-farnesol.

### 2.4.2. Variation of other sesquiterpene components

Gas chromatography-mass spectrometry (GC-MS) analysis of *S. spicatum* heartwood revealed 17 major sesquiterpene hydrocarbons and alcohols, which contribute up to 80% of the total oil composition. Even where oil yields were barely detectable, the main components α- and β-santalol, *E,E*-farnesol, dendrolasin, Z-lanceol, Z-nuciferol/Z-γ-curcumen-12-ol and Z-β-curcumen-12-ol were always present (Table 2.1). Most of these compounds are also found in other sandalwood species (Baldovini et al., 2011). The apparent fixation of these compounds across different *Santalum* species suggested they may contribute to fitness. Conversely, other components such as α-bisabolol, dendrolasin and nerolidol were more variable across sampling locations (Table 2.1). Mean α-bisabolol content in the North and Wheatbelt (1.50 ± 1.7% and 1.80 ± 3.3% respectively) were markedly lower than in Goldfields trees (9.70 ± 4.9%), suggesting an
α-bisabolol chemotype might be present in this region. Trees sampled from the North also contained higher levels of the olefins α- and β-santalene (1.2 ± 0.5%) compared to Wheatbelt (0.5 ± 0.3%) and Goldfields trees (0.2 ± 0.1%). Differences in the relative abundances of the santalenes may be due to differential expression or activity of santalene synthases (Jones et al., 2011), which produce the santalenes, and cytochromes P450 (Diaz-Chavez et al., 2013), which convert the santalenes to the santalols. Expression of the santalene synthase and P450 genes may be developmentally regulated or affected by environmental factors as is known for terpene synthases and terpene converting P450s in other systems (Chen et al., 2011; Hamberger et al., 2011). Sandalwood may be unusual in this regard as sesquiterpenes appear to be produced, and stored indefinitely, in heartwood xylem. As accumulation of oil generally initiates by 10 years of age and continues for the life of the tree, and little to no seasonal variation is evident (Moretta, 2001), differential biosynthesis of individual metabolites may be more subtle than in short-lived species or in species with strong seasonal variation of metabolic activity.

2.4.3. Statistical analysis of chemotypes

Principle component analysis (PCA) indicated that distinct chemotypes of S. spicatum existed across its natural range. The first two principle components, which explained 83% of the variance is shown in a two dimensional plot (Fig. 2.4a) and indicated that trees within regions tended to share a more similar heartwood chemistry than geographically distant populations, although there were similarities among regions as
well, particularly with north and wheatbelt trees. Overlap in the terpenoid profiles showed that the essential oil composition across regions were not sufficiently differentiated based on genetic or geographic differences. PCA revealed high levels of correlation between oil components, with the first three principle components accounting for 95% of the variance (Fig. 2.4). The analysis indicated that \textit{E,E}-farnesol and \textit{\alpha}-santalol are two largely independent drivers of variability and are the variables mainly responsible for the ordination in the oil of \textit{S. spicatum}. The \textit{\alpha}-santalol chemotypes with low levels of \textit{E,E}-farnesol were mostly represented by trees from the North as well as several individual from the south-west (Wheatbelt) which displayed some remarkable similarities in total oil profiles despite their geographic separation. The PCA also suggested that the sampled trees from the south-east (Goldfields) did not contain any individuals with an \textit{\alpha}-santalol chemotype, however \textit{E,E}-farnesol in addition to \textit{\alpha}-bisabolol were more important contributors to the total oil profile of these trees. These results further support the potential for chemical selection of \textit{S. spicatum} trees for tree improvement programs.
Figure 2.4 - Two dimensional principal component analysis (PCA) ordination scores of \textit{S. spicatum} sesquiterpene oil samples from the heartwood of 194 trees. Each point represents an individual tree, and points close together are similar in terms of oil composition. The first three components represent 95% of the total variance A) represents components 1 and 2 and B) represents components 1 and 3. Lines indicate ordination scores of the variables used.

Variation in terpenoids can be used to some degree to predict pathways of biosynthesis within the plant (Jones et al., 2006). In sandalwood, sesquiterpenes are synthesized via terpene synthases and cytochrome P450s (Diaz-Chavez et al., 2013; Jones et al., 2008; Jones et al., 2011). While some terpene synthases produce a single product, many reported terpene synthase enzymes catalyse the formation of multiple products (Chen et al., 2011). A hierarchical cluster analysis identified patterns of accumulation and provided an approximation to the number of different terpene synthase enzymes which may produce the major \textit{S. spicatum} components (Fig. 2.5). The four main clades of the dendogram indicate at least four separate sets of enzymes may be responsible for producing the major sesquiterpenes and their variations in \textit{S. spicatum}. This is in part confirmed by the knowledge of a previously characterized santalene synthase from \textit{S.}}
spicatum (SspiSSy) which is responsible for producing the four compounds, \( \alpha \)-, \( \beta \)-, epi-\( \beta \)-santalene and \( \alpha \text{-} \text{trans} \)-bergamotene represented in clade IV (Jones et al., 2011). In addition, the hydroxylated analogues of the santalenes in clade III of the dendogram, were recently reported to be the products of a cytochrome P450 CYP76F family from S. album, which produced \( \alpha \)-, \( \beta \)-, epi-\( \beta \)-santalol, \( \alpha \text{-} \text{trans} \)-bergamotol and isomers thereof (Diaz-Chavez et al., 2013).

In S. spicatum, the additional characterization of the biosynthetic pathway of \( \text{E,}\text{E} \)-farnesol remains an important objective of future work, since this component reduces the market potential of the oil. Clade III of the dendogram contains a cluster of \( \text{E,}\text{E} \)-farnesol, dendrolasin, \( \text{E} \)-nerolidol and \( \text{E-}\beta \)-farnesene which are structurally similar (Supplementary Figure 2.S1). \( \text{E,}\text{E} \)-farnesol is highly variable in S. spicatum, so if a correlation exists between \( \text{E,}\text{E} \)-farnesol and other structurally similar oil components, a single enzyme might produce both components, stemming from a common intermediate. \( \text{E,}\text{E} \)-farnesol and dendrolasin showed a positive correlation \((r^2 = 0.53)\), suggesting some level of shared biogenesis. Dendrolasin is a furan sesquiterpenoid structurally similar to farnesol, so it is possible that the two compounds share a common origin. Alternatively, \( \text{E,}\text{E} \)-farnesol may be a minor component of reactions leading to dendrolasin, \( \text{E} \)-nerolidol and \( \text{E-}\beta \)-farnesene \((P < 0.05)\), thus complicating the linear relationship and producing lower coefficients of correlation (Supplementary Figure 2.S1). The possibility for multiple biosynthetic pathways for one component could also help explain some of the quantitative variation that exists in oil components across the geographic range of S. spicatum. Identification of additional pathways is important to understanding total oil production and its regulation in Santalum and whether it can be manipulated to increase expression of certain components.
Figure 2.5 - Dendrogram showing co-occurrence patterns of 13 major components using hierarchical cluster analysis from the heartwood-oil samples of 194 S. spicatum individuals distributed in Western Australia: 152 trees from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the north regions (Carnarvon and Shark Bay).

The genetic heritability of various economically important traits (such as oil composition, vigour and growth form) has not been examined extensively in S. spicatum. Other terpenoid rich plant species such as culinary basil (Ocimum, Lamiaceae) have been successfully bred for desirable chemical characteristics based on studies into chemical composition and field experiments. Like S. spicatum, many cultivars of basil vary in their aroma, with chemotypes of citral, eugenol, linalool, methylchavicol and methylcinnamate well represented (Charles et al., 1990); and variation of composition under different conditions needed to be taken into consideration when devising selection criteria (Vieira and Simon, 2000). Chemical selection and breeding resulted in basil cultivars with up to 15% more methyl cinnamate than the wild types indicative of a high heritability of aroma compound biosynthesis. A
similar approach could be tested for *S. spicatum* based on oil composition data presented here using seeds sourced from individuals where the desired oil characteristics are evident: high total oil content, high levels of α- and β-santalol and low *E,E*-farnesol amounts. While breeding cycles will be lengthy, the resulting seedstock would be considerably more valuable. If associations of genetic and genomics could complement such efforts in future works, the length of breeding cycles could be reduced.

Terpene biosynthesis in *S. spicatum* may be under tight genetic control. Expression and regulation of certain components is likely to be affected by the environment, since inter-population similarities in oil profiles were generally evident, but not always the case. Future work should focus on determining the exact genetic and environmental contributors influencing chemotype heritability and further explore the pathways for sesquiterpene biosynthesis. This will enable a more comprehensive understanding of factors involved in total oil production across *Santalum*, and whether this can be manipulated within a tree to increase expression of certain genes or initiate early heartwood-oil development. The future aim is to ultimately reduce the overall domestication time and increase the market potential of sandalwood oil.
2.5. Supplementary Materials

Supplementary Figure 2.1 - The linear correlation between \(E,E\)-farnesol, dendrolasin, \(E\)-nerolidol and \(E\)-\(\beta\)-farnesene based on GC-MS analysis 194 heartwood cores of \(S.\) spicatum trees in natural stands of Western Australia: 152 trees from the south-west (Wheatbelt), 19 trees from the south-east (Goldfields) and 23 trees from the north regions (Carnarvon and Shark Bay) using percent composition data

<table>
<thead>
<tr>
<th>Label</th>
<th>Compound</th>
<th>(R^2)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>dendrolasin</td>
<td>0.53</td>
<td>(&lt; 2 \times 10^{-16})</td>
</tr>
<tr>
<td>b</td>
<td>(E)-nerolidol</td>
<td>0.1</td>
<td>0.0002467</td>
</tr>
<tr>
<td>c</td>
<td>(E)-(\beta)-farnesene</td>
<td>0.002</td>
<td>0.0269267</td>
</tr>
</tbody>
</table>

2.6. Conclusions

This study highlighted variations in oil quality (assessed by \(\alpha\)- and \(\beta\)-santalol content and \(E,E\)-farnesol) of \(S.\) spicatum across the semi-arid and arid-regions of Western Australia and showed the existence of different \(S.\) spicatum chemotypes in three regions in Western Australia: the semi-arid south west Wheatbelt and the arid south-east Goldfields regions (both of which past and current harvesting practises occur) and the north Carnarvon region. The northern trees showed, on average, better quality oil than
the south; however yields varied within regions rather than across. PCA analysis indicated that α-santalol and \(E,E\)-farnesol chemotypes existed and that it may be possible to select for trees predisposed to increased santalol production. The variation in oil yield and composition within and across regions could be influenced by a range of factors including age of the trees, growth rate, and environmental variations such as soil composition, host tree availability, rainfall or the presence of pathogens which may stimulate production of certain oil components. Tree improvement and better management of natural stands and plantations will be achieved through continued research into genetic and environmental factors which influence oil production, as well as factors which might predispose specific chemical phenotypes.

2.7. Acknowledgements

The authors are grateful for the financial assistance of the Australian Research Council and the Forest Products Commission of Western Australia through linkage project LP0882690. The authors also wish to thank Mr. Len Norris for assistance on field trips and other technical advice, Mr. Jon Brand of FPC for providing access to plantation trees. Work in the laboratory of Joerg Bohlmann was supported, in part, by a grant from the Natural Sciences and Engineering Research Council of Canada.
2.8. References


Brand, J.E., Pronk, G.M., 2011. Influence of age on sandalwood (Santalum spicatum) oil content within different wood grades from five plantations in Western Australia. Austr. Forestry 74, 141-148.


Chen, F., Tholl, D., Bohlmann, J., Pichersky, E., 2011. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 66, 212-229


Moretta, P., 2001. Extraction and variation of the essential oil from Western Australian sandalwood (*Santalum spicatum*). Doctor of Philosophy (PhD). Chemistry, University of Western Australia, Publisher, Perth, 1-239.


Schnuch, A., Uter, W., Geier, J., Lessmann, H., Frosch, P.J., 2004. Contact allergy to farnesol in 2021 consecutively patch tested patients. Results of the IVDK. Contact Dermatitis 50, 117-121.


CHAPTER THREE
Sandalwood Fragrance Biosynthesis Involves Sesquiterpene Synthases of Both the Terpene Synthase (TPS)-a and TPS-b Subfamilies, Including Santalene Synthases

Christopher G. Jones‡, Jessie Moniodis‡, Katherine G. Zulak‡§, Adrian Scaffidi¶, Julie A. Plummer‡, Emilio L. Ghisalberti¶, Elizabeth L. Barbour‡, and Jorg Bohlmann§

‡School of Plant Biology (M084), Faculty of Natural and Agricultural Sciences, University of Western Australia, Crawley, WA 6009, Australia
¶School of Biomedical, Biomolecular, and Chemical Sciences (M313), Faculty of Life and Physical Sciences, University of Western Australia, Crawley, WA 6009, Australia
§The Forest Products Commission of Western Australia, Rivervale, WA 6103, Australia
¶Michael Smith Laboratories, University of British Columbia, 301-2185 East Mall, Vancouver, BC V6T 1Z4, Canada
3.1. Abstract

Sandalwood oil is one of the world's most highly prized fragrances. To identify the genes and encoded enzymes responsible for santalene biosynthesis, we cloned and characterized three orthologous terpene synthase (TPS) genes *SaSSy*, *SauSSy*, and *SspiSSy* from three divergent sandalwood species; *Santalum album*, *S. austrocaledonicum*, and *S. spicatum*, respectively. The encoded enzymes catalyze the formation of α-, β-, *epi*-β-santalene, and α-*exo*-bergamotene from (*E*,*E*)-farnesyl diphosphate (*E*,*E*-FPP). Recombinant *SaSSy* was additionally tested with (*Z*,*Z*)-farnesyl diphosphate (*Z*,*Z*-FPP) and remarkably, found to produce a mixture of α-endo-bergamotene, α-santalene, (*Z*)-β-farnesene, *epi*-β-santalene, and β-santalene. Additional cDNAs that encode bisabolene/bisabolol synthases were also cloned and functionally characterized from these three species. Both the santalene synthases and the bisabolene/bisabolol synthases reside in the TPS-b phylogenetic clade, which is more commonly associated with angiosperm monoterpene synthases. An orthologous set of TPS-a synthases responsible for formation of macrocyclic and bicyclic sesquiterpenes were characterized. Strict functionality and limited sequence divergence in the santalene and bisabolene synthases are in contrast to the TPS-a synthases, suggesting these compounds have played a significant role in the evolution of the *Santalum* genus.
3.2. Introduction

*Santalum album* L. is a slow-growing hemi-parasitic tree, which has long been exploited for its fragrant heartwood. Other species such as *S. spicatum*, native to the arid and semi-arid regions of Western Australia (WA) and *S. austrocaledonicum* from Vanuatu and New Caledonia, have also contributed substantially to the fragrance market (1). Unsustainable demand for sandalwood has led to the establishment of plantations to add supply. Knowledge of the molecular and physiological underpinnings of sandalwood oil biosynthesis will advance plantation development through improved management and selection. Sandalwood heartwood contains a complex mixture of sesquiterpene olefins and alcohols (Fig. 3.1) with (+)-α-santalene, (-)-β-santalene, (-)-α-exo-bergamotene, (+)-epi-β-santalene, and β-bisabolene frequently representing about 1-2% by weight of oil (2, 3). The more odor intensive compounds, α-- and β-santalol, α-exo-bergamotol, and epi-β-santalol, as well as (Z)-lanceol and α-bisabolol make up the bulk of *S. album* oil, but are often lower and more variable in *S. austrocaledonicum* (4) and *S. spicatum* (5). The compounds found in sandalwood oil may play an important role in warding off pathogens, as sesquiterpenoids have been implicated in plant pathogen defense strategies (6). Santalol biosynthesis is proposed to proceed by multiple Wagner-Meerwein rearrangements of transoid farnesyl diphosphate, (*E*,*E*)-FPP,3 followed by oxidation at C12, probably via a cytochrome P450 (Fig. 3.1). Co-occurrence patterns of the santalenes and bergamotene, as well as other olefins and bisabolol, indicated that multi-product TPS enzymes may be responsible for the production of several sandalwood oil components (2). Multiple product formation from terpene synthase (TPS) enzymes in *S. album* was recently confirmed by characterization of two TPSs; however these were not responsible for santalene biosynthesis (7). The commercially exploited sandalwoods possess unique...
chemical phenotypes (3), whereas others such as *S. accuminatum* and *S. murrayanum*, endemic to southwestern WA, produce little or no oil (8). Despite this variation, the santalenes and their hydroxylated equivalents remain key components. To better understand essential oil biosynthesis as well as the molecular origins of chemical diversity in the *Santalum* genus, we cloned and functionally characterized several TPS cDNAs from three divergent oil-producing *Santalum* species, *S. album*, *S. austrocaledonicum*, and *S. spicatum*. We also compared genomic TPS sequences of the three commercial *Santalum* species to those of *S. murranyanum* to test if the absence of, or open reading frame mutations in the genes accounted for the oil-deficient phenotype of this species.

![Figure 3.1 - Biosynthesis of sesquiterpenes in sandalwood commences with the TPS-catalyzed rearrangements of farnesyl diphosphate. Specific oxidation at C12 is proposed to occur via a cytochrome P450 enzyme](image-url)

63
3.3. Experimental

3.3.1. Chemicals and Reagents

All reagents, solvents, antibiotics, cloning kits, modifying enzymes, and precursor chemicals were purchased from commercial sources. \((E,E)\)-FPP and geranyl diphosphate (GPP) were from Sigma. \((Z,Z)\)-FPP was synthesized using standard methods (9, 10) and products verified by GC-MS, NMR, and IR spectroscopy. Monoterpene standards were from an in-house collection of commercially available standards. An olefin fraction of sandalwood oil yielded a santalene standard (see below under GC-MS analysis and product identification).

3.3.2. Plant Material Collection and RNA Extraction

Several 25 mm holes were drilled into the lower stems of mature \(S. \text{album}\), \(S. \text{austrocaledonicum}\), \(S. \text{spicatum}\), and \(S. \text{murrayanum}\) trees growing on land managed by the Forest Products Commission of Western Australia. Wood shavings from the heartwood-sapwood transition zone were collected and frozen immediately in liquid nitrogen. The samples were transported to the lab where RNA was extracted from 10 g of tissue using an established protocol (11). After precipitation by LiCl, RNA was stored at 80°C until needed for reverse transcription and RACE. \(Santalum \text{album}\) RNA was transported to UBC Vancouver, Canada for cDNA library construction.
3.3.3. Santalum album cDNA Library Construction

Xylem total RNA from *S. album* (1.4 µg) was reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) and a cDNA library was constructed using the SMART-Creator kit with the pDNR-LIB vector (Clontech). The ligation mixture was transformed by electroporation into 25 µl of phage resistant electro-competent *Escherichia coli* cells and Sangar-sequenced at the Michael Smith Genome Sciences Centre, Vancouver, Canada. Reads were assembled using the CAP3 program with default settings. The sequences (~6000 unique reads) were compared with the GenBankTM database for key specialized metabolism genes, particularly prenyltransferases and TPS genes. The following sequences have been deposited into the GenBankTM database: *SaSSy* (HQ343276); *SauSSy* (HQ343277); *SspiSSy* (HQ343278); *SauBS* (HQ343279); *SspiBS* (HQ343280); *SauSesquiTPS* (HQ343281); *SspiSesquiTPS* (HQ343282), *SaSTPS gDNA* (JF746807); *SauSTPS gDNA* (JF746808); *SspiSTPS gDNA* (JF746809); *SmSTPS gDNA* (JF746810); *SmSSy* (JF746811); *SaSSy* (JF746812); *SauBS gDNA* (JF746813); *SmBS* (JF746814); *SaMonoTPS1 gDNA* (JF746815).

3.3.4. TPS Gene Discovery and RACE

cDNA was generated for *S. austrocaledonicum*, *S. spicatum*, and *S. murrayanum* in the same manner as before, except the cDNA was used directly as template for PCR. Primers based on the open reading frame (ORF) of previously identified *S. album* TPS genes were used for amplification (7) (supplemental Fig. 3.S1). Where products could not be amplified, 5’- and 3’-RACE using the Clontech SMARTerTM kit was used to
obtain the respective untranslated regions for more specific primer design. *SaSesquiTPS1* gene orthologs were amplified in two rounds using a nested primer approach. All products were first cloned into a high-copy storage vector (TOPO Zero Blunt, Invitrogen, or pJET1.2, Fermentas) for sequencing before being subcloned into pET28b(+). Several individuals of *S. album* and *S. spicatum* were studied to examine if polymorphisms were present in the ORFs across populations of these two species. No TPS genes could be amplified from *S. murrayanum* cDNA, although this could be due to exceedingly low transcript abundance. Genomic DNA sequences for all three TPS genes were also cloned and sequenced for all three *Santalum* species. The same ORF primers used for successful cDNA amplifications were used on genomic DNA extracted from the same individuals from which RNA extractions were performed. These larger gDNA fragments (3-4 kb) were cloned into pJET1.2 vectors and sent for sequencing (Macrogen, Korea). Genomic DNA sequences of TPS genes from the three oil-bearing species were spliced *in-silico* with introns following the usual 3’ n/GT and 5’ AG/n pattern (12). *Santalum murrayanum* intron-exon patterns were determined by comparing the gDNA sequences to cDNAs of *S. spicatum*.

### 3.3.5. Phylogenetic Analysis and Alignment of TPS Genes

Sequences were aligned using ClustalX 2.1.0, trees were constructed using Phylip (13) and visualized with the Phylodraw 0.8 program. Sequences used in the phylogenetic analysis are listed in full under supplemental Fig. 3.S2. Genomic DNA sequences of TPS genes from the three oil-bearing species were spliced *in-silico* with introns following the usual 3’ n/GT and 5’ AG/n pattern (12). *S. murrayanum* intron-exon patterns were determined by comparing the gDNA sequences to cDNA sequences of
S. spicatum.

3.3.6. Bacterial Expression and Protein Isolation

TPS genes were cloned into the pET28b(+) expression vector (Novagen, San Diego CA) with a polyhistidin tag in-frame. Depending on the restriction sites available, the His6 tag was either N-terminal or C-terminal. Primers with appropriate restriction sites (supplemental Fig. 3.S1) were used to amplify each gene and cloned into the pET28b(+) vector. Vectors containing the TPS genes were transformed into chemically competent C41 E. coli cells (Avidis, Saint-Beauzire, France) containing the pRARE 2 plasmid isolated from Rosetta 2 competent cells (Novagen). Colonies were grown on LB plates containing kanamycin and chloramphenicol (50 µg ml\(^{-1}\)). Three independent colonies were picked and grown in a shaker overnight at 37 °C in 5 ml of LB with the same antibiotics and this culture was used to inoculate 400 ml of Terrific Broth. Cell suspensions were grown at 37 °C with shaking until the \(A_{600} = 0.8\) and induced with isopropyl-D-thiogalacto-pyranoside (IPTG) to a final concentration of 0.2 mM, and shaken overnight at 16 °C. Cell suspensions were centrifuged at 4 °C and pellets (~1 g) were frozen at 80 °C for future use. Cell pellets were resuspended in 5 ml of lysis buffer containing 1 mg ml\(^{-1}\) lysozyme, 1 mM MgCl\(_2\), 5 mM DTT, 0.01 mg ml\(^{-1}\) DNAse1, and RNAse1, 100 µl of protease inhibitor mixture (Sigma) and made in His-trap binding buffer (20 mM Na\(_2\)HPO\(_4\), pH 7.4, 500 mM NaCl, 30 mM imidazole, pH 7.4). Cells were stirred thoroughly on ice with a glass rod for 30 min followed by homogenized using a high pressure cell cruncher. The lysate was centrifuged at 12,000 x g at 4 °C for 1.25 h before being decanted. Cleared lysate (~12 ml) was purified using Ni\(^{2+}\) affinity chromatography spin-columns (GE Healthcare) and eluted in 600 l of elution buffer (20 mM Na\(_2\)HPO\(_4\) pH 7.4, 500 mM NaCl, 500 mM imidazole, pH 7.4). The eluted protein
was desalted on a PD-10 desalting column (GE Healthcare) using 25 mM 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES) pH 7.4, 10% (v/v) glycerol and 100 mM KCl. Approximately 1 ml fractions from a 3.5 ml elution were collected. Protein concentrations were determined using a Nano-Drop spectrophotometer with extinction coefficients calculated by amino acid composition (14). SDS-PAGE followed by Coomassie Blue staining visualized the purified recombinant proteins (supplemental Fig. 3.S3).

3.3.7. Enzyme Functional Characterization and Kinetic Assays

Enzyme assays for all recombinant proteins were done in triplicate using the GC vial method described by O’Maille et al. (15). For enzyme assays where only product identification was assessed, 10 µg of protein was used in a final volume of 500 µl of reaction buffer (25 mM HEPES, 10% (v/v) glycerol, 5 mM DTT, and 10 mM of either Mg^{2+} or Mn^{2+}. Substrates (FPP and GPP) were added to a final reaction concentration of 100 µM. Vials were overlaid with 500 µl of hexane to trap volatile products and incubated at 30 °C for 2 h. Mixtures were vortexed for 1 min to extract all volatiles and the vials were centrifuged to separate the organic layer. For determination of steady-state enzyme kinetic constants, conditions were as described previously except the enzyme concentration was kept at 10 nM. Substrate concentrations ranged from 1 µM to 100 µM, and reactions were incubated at 30 °C for exactly 5 min. Critically, reactions were quenched at 5 min by the addition of 500 µl 0.5 M EDTA, pH 8.0 and vortexed, then stored at 80 °C immediately.
3.3.8. *In Vivo Santalene Production in E. coli*

An approach similar to the previously published method for diterpene production (16) was used for *in vivo* production of santalenes. The *Sa*FPPS gene was identified in the xylem EST library (GenBankTM accession no. HQ343283) and was amplified from cDNA using primers with NcoI and NotI ends amenable to pCDFDuet-1 vector multiple cloning site 1. Likewise, *Sa*SSy was amplified with NdeI and KpnI sites adapted for use with the In-Fusion cloning system (Clontech) and cloned into the second multiple cloning site (supplemental Fig. 3.S1). The dual expression vector construct, pCDFDuet-1:*Sa*FPPS:*Sa*SSy, as well as an empty pCDFDuet-1 vector control were transformed into chemically competent C41(DE3) cells and grown on 50 µg ml\(^{-1}\) streptomycin selective media. Individual colonies were grown in 5 ml overnight cultures, and these were used to inoculate 200 ml shaking cultures of Terrific Broth. As cells were approaching the log-phase of growth (*A*\(_{600}\) = 0.6) at 37 °C the incubator was cooled to 20 °C for 1 h before inducing with IPTG at a final concentration of 0.5 mM. Cultures were shaken for a further 16 h before being centrifuged to pellet the cells. Cell pellets were lysed with 0.2 M NaOH and neutralized with acetic acid before being extracted twice with hexanes and reduced by evaporation. Media was extracted twice with 100 ml of hexane and these fractions were reduced by rotary evaporation to yield a thin oily residue. The residues were resuspended in 2 ml of hexane and analyzed by GC-MS using conditions described below.
3.3.9. GC-MS Analysis and Product Identification

Product mixtures were analyzed by GC-MS in scan mode for product identification. A standard containing the three santalenes and \( \alpha \text{-exo-bergamotene} \) was prepared by flash chromatography of 2 ml of neat \( S. \text{album} \) oil over silica and eluted in hexane. A final yield of 25 mg was resuspended in EtOH, and purity was confirmed by GC. Monoterpene identification was aided through an in-house standard collection. All mass spectra were compared with the NIST 2005 library and the literature (17-23). The Kovat retention indices (KIs) were determined for all compounds (supplemental Figs. 3.S4, 3.S7, 3.S10) using an n-alkane standard and compared with the literature wherever DB-WAX or similar phase column data were available (17-21, 23, 24). Where reliable retention data and pure standards were not available, a combination of mass spectral and retention properties was used to infer the candidate compound. GC-MS was performed on a Shimadzu GC2010 with a DB-WAX column and He as carrier gas. Conditions were as follows: Injector 200 °C, MS interface 240 °C, ion source 200 °C, Oven program: 40 °C for 3 min, then 8 °C min\(^{-1}\) to 180 °C, held 5 min, then 10 °C min\(^{-1}\) to 220 °C, held 10 min. Solvent cut time was set to 5 min. For product identification, total ion monitoring was used, scanning from m/z 41 to m/z 250. Incubations with (\( Z,Z \))-FPP were analyzed on a HP5890 with a DB-WAX column and an initial oven temperature of 40 °C then ramped at 10 °C min\(^{-1}\) to 230 °C and held for 20 min. For kinetic assays, single ion monitoring (SIM) of the sesquiterpene base ions m/z 91, 93, and 94 were used. Likewise monoterpene base ions (m/z 69, 71, and 93) were monitored for GPP assays. An internal standard (isobutyl benzene, 30 µM) was added to the hexane used to overlay each reaction. Detector response factors were calculated based on the santalene standard which was prepared earlier. Product losses due to
extraction inefficiency was accounted for by first adding the standard to assay buffer as an EtOH stock and extracting into the hexane layer as per sample assays.

3.4. Results

A full-length cDNA, SaSSy (S. album santalene synthase) homologous to previously reported angiosperm TPS-b group genes (25) was identified in the S. album xylem EST library and cloned. It encoded a 569 amino acid protein with 56 % identity to SamonoTPS1 (7) and appeared to lack an N-terminal transit peptide. Upon heterologous expression in E. coli, the affinity purified recombinant His6-tagged protein had a molecular mass of ~ 66 kDa, similar to that of most monomeric TPS enzymes (supplemental Fig. 3.S3). When incubated in the presence of 10 mM Mg$^{2+}$ the enzyme converted (E,E)-FPP into α-, β- and epi-β-santalene, α-exo-bergamotene, as well as traces of α- and β-farnesene (Fig. 3.2). Orthologous TPSs identified in two other oil-bearing species (SspiSSy from S. spicatum and SauSSy from S. austrocaledonicum) were also found to convert (E,E)-FPP into the santalenes in very similar proportions as SaSSy (Fig. 3.2). In all three orthologs, incubations with Mn$^{2+}$ yielded mainly α-exo-bergamotene (supplemental Fig. 3.S4). The larger metal ion likely distorts the active site, causing premature quenching of the bergamotyl carbocation intermediate. Each santalene synthase had an apparent Km of 1.4 (± 0.3) µM, indicating a biologically relevant, high affinity for (E,E)-FPP. Catalytic turnover rates were similar for all three enzymes, with a kcat for SaSSy of 0.34 s$^{-1}$, 0.91 s$^{-1}$ for SauSSy, and 2.6 s$^{-1}$ for SspiSSy. All three santalene synthases produced linalool, geraniol, and terpineol along with traces of α-pinene and camphene when incubated with geranyl diphosphate (GPP) (supplemental Fig. 3.S4). Although conversion of GPP did occur, a linear relationship
existed between substrate concentration and $V_0$, even at high (> 100 µM) concentrations, rather than an asymptotic curve indicative of active site saturation. Monoterpenes have only been reported in sandalwood oil at very low concentrations (23, 26).

Figure 3.2 - GC-MS chromatogram of in vitro assays with recombinant santalene synthases; SaSSy (black trace), SauSSy (blue trace), and SspiSSy (red trace)

Incubations of SaSSy with (Z,Z)-FPP indicated the enzyme was also catalytically active on this isomer. In the presence of 10 mM Mg$^{2+}$ (Z,Z)-FPP was converted into α-endo-
bergamotene, α-santalene, (Z)-β-farnesene, epi-β-santalene, and β-santalene (Fig. 3.3A). The product profile resulting from incubations of SaSSy with (E,E)-FPP (Fig. 3.3B) resemble the authentic essential oil of S. album (Fig. 3.3C) more closely than that of the (Z,Z)-FPP incubations.

Figure 3.3 - GC-MS chromatogram of incubations of (Z,Z)-FPP with SaSSy (A), (E,E)-FPP with SaSSy (B), and the olefin fraction of authentic sandalwood oil for comparison (C). Peaks: 1) α-endo-bergamotene, 2) α-santalene, 3) α-exo-bergamotene, 4) (Z)-β-farnesene
Continuity of santalene biosynthesis from FPP produced by a *S. album* FPP synthase (*SaFPPS*) was confirmed using an *in vivo* *E. coli* expression system. *SaFPPS* and *SaSSy* were cloned into the dual expression vector pCDFDuet-1 (Novagen) and transformed into C41 chemically competent cells. No detectable levels of the santalenes or bergamotene were found in the cell pellet extraction, or in the empty vector control extracts, but all four compounds were detected in hexane extracts of media from overnight cultures (Fig. 3.4). These results validated the *in vivo* activity of *SaSSy* and demonstrate the feasibility of metabolically engineering a santalene synthase into an appropriate host microorganism for *in vivo* production of santalenes using established methods (16, 27).

![Figure 3.4 - In vivo production of santalenes in overnight *E. coli* cultures. Peak 1, α-santalene; peak 2, α-exo-bergamotene; peak 3, epi-β-santalene; peak 4, β-santalene](image)

To further explore the origins of chemical diversity in the genus *Santalum*, the orthologous TPS gene pair *SauBS* (*S. austrocaledonicum*, β-bisabolene synthase) and *SspiBS* (*S. spicatum*, α-bisabolol synthase) were cloned and characterized using cDNA as PCR template with primers originally developed from *S. album* (7). Recombinant
SauBS produced almost exclusively β-bisabolene and only traces of α-bisabolol with (E,E)-FPP, while only limonene and terpineol were produced when incubated with GPP (supplemental Figs. 3.S5 and 3.S6). SspiBS produced a mixture of β-bisabolene and α-bisabolol, along with traces of α-bisabolene and farnesene isomers. The functions of these two enzymes were very similar to the previously identified SaMonoTPS1 from S. album which produced mostly β-bisabolene and traces of α-bisabolol from (E,E)-FPP, but could also convert GPP into monoterpenes analogous in structure to the bisabolenes (7). Extending our investigation of Santalum sesquiterpene synthases into the TPS-a subfamily, two orthologous cDNAs; SspiSesquiTPS and SauSesquiTPS, from S. spicatum and S. austrocaledonicum, respectively, were cloned and characterized. The translated amino acid sequences of these genes were similar to the previously characterized SaSesquiTPS1 from S. album (7) but the recombinant TPSs of these species yielded markedly different sesquiterpene profiles. SauSesquiTPS produced α-humulene and γ-cadinene, along with β-elemene, which is the thermal rearrangement product of germacrene A (28) and several other bicyclic sesquiterpenes when incubated with (E,E)-FPP (supplemental Fig. 3.S7). The cadinenes may also be the result of heat-induced dehydration rearrangements (7). In contrast, SspiSesquiTPS produced only three main compounds; β-elemol (the thermal rearrangement product of hedycaryol (29)), guaiol and bulnesol (supplemental Fig. 3.S8). Both enzymes produced only traces of linalool with GPP (supplemental Fig. 3.S9).

Each set of orthologous TPS across the Santalum genus showed sequence homology to previously reported TPS with key domains being well conserved (supplemental Fig. 3.S10). As with all angiosperm TPSs, the R(R/P)X8W motif implicated in prenyl diphosphate ionization (30) and the aspartate-rich divalent metal ion binding domain.
(31) (DDXXD) are present. Residues likely to be responsible for product specificity and substrate preference were identified in the α14 helix, and the α18 - α19 helix turn respectively, based on those identified by Kamparinas et al. (32).

Genomic sequences of a set of *Santalum* TPS genes were compared to determine whether differences at the genomic level might be related to the chemotypic differences observed among oil producing (*S. album, S. spicatum, S. austrocaledonicum*) and oil-deficient (*S. murrayanum*) species. All four species have genomic copies of the target TPS genes (supplemental Fig. 3.S11) although cDNAs could not be amplified from a xylem-derived RNA pool of *S. murrayanum*. The deduced amino acid sequences of the santalene synthase, bisabolene/bisabolol synthase, and TPS-a sesquiterpene synthase gDNAs in *S. murrayanum* are highly homologous to those of the functionally characterized TPS cDNAs of the other species (Fig. 3.7) and possibly encode functional TPS enzymes. Intron-exon structure of sandalwood TPS genes were typical of angiosperm TPS-a and TPS-b genes; each bearing 6 introns and 7 exons (12) (supplemental Fig. 3.S11). Intron boundaries observed a 3’ N*GT, 5’ AG*N pattern. An ortholog of *SauBS* in *S. murrayanum*, labeled *SmBS*, showed a mutation encoding a stop codon in exon 3. We also cloned a variant *SaSSy* gDNA sequence from *S. album* with a 10-nucleotide deletion in exon 3, producing a frameshift and subsequent stop codon. No frame shifts or premature stop codons were present in the other TPS gDNAs sequenced.
Figure 3.5 - Detailed mechanism explaining the formation of products resulting from incubations of SaSSy with \((E,E)\)-farnesyl diphosphate in the presence of Mg\(^{2+}\)
3.5. Discussion

The genes and encoded enzymes for santalene biosynthesis have been isolated from three divergent sandalwood species. SaSSy, SauSSy, and SpiSSy are bona fide sesquiterpene synthases for santalene formation, despite being phylogenetically aligned with the TPS-b subfamily, which comprises mainly of angiosperm monoterpene synthases. Low $K_m$ values for each santalene synthase indicated a high affinity for $(E,E)$-FPP, while no active site saturation was apparent with GPP, indicating they are genuine sesquiterpene synthases. Vestigial activity with GPP suggests that these santalene synthases may have evolved from a monoterpene synthase ancestor through loss of the plastid target peptide and subsequent specialization of the active site for $(E,E)$-FPP. The putative ancestral enzyme which gave rise to SaSSy, SauSSy, and SpiSSy likely had latent plasticity in the active site, enabling it to accommodate both GPP and FPP. Similarly, it has been shown in the case of fruit flavor biosynthesis in wild and cultivated strawberry (33) that molecular evolution of the corresponding TPSs involved plasticity of the active site to accommodate alternative substrates when subcellular localization of the enzymes is changed through loss of a plastid target peptide. Most intriguing, SaSSy was able to convert $(Z,Z)$-FPP into $\alpha$, $\beta$-, and $\text{epi}$-$\beta$-santalenes and $\alpha$-$\text{endo}$-bergamotene. Sallaud et al. (34) discovered a santalene/bergamotene synthase in wild tomato, which was able to accommodate $(Z,Z)$-FPP but was unreactive toward $(E,E)$-FPP (Fig. 3.5). SaSSy is uniquely able to accommodate both transoid and cisoid isomers of FPP and surprisingly, produce a similar suite of compounds. These rearrangements are likely to proceed via the initial ionization of $(Z,Z)$-FPP into either (3S) or (3R)-nerolidyl diphosphate (Fig. 3.6). Cyclization may occur through an endo- conformation, resulting in $\alpha$-$\text{endo}$ bergamotene and (-)-$\text{epi}$-$\beta$-santalene, while the exo-cyclization would lead to $\alpha$- and $\beta$-
Sandalwood oil contains traces of \( \alpha\)-endo-bergamotene (26) thus it is plausible that the native enzyme is able to access both isomers of FPP in the cell. To our knowledge this level of plasticity in the active site of a TPS is unprecedented.

Figure 3.6 - Detailed mechanism explaining the formation of products resulting from incubations of SaSSy with \((Z,Z)\)-farnesyl diphosphate in the presence of Mg\(^{2+}\)
In addition to characterizing the santalene synthases, continuity of the biosynthetic pathway from FPP to the santalenes (Fig. 3.4) was demonstrated using an *in vivo* production system, highlighting the potential for metabolic engineering of micro-organisms to produce precursors of valuable fragrance compounds of limited availability from plants. This system will also serve as a useful platform for exploring further downstream metabolic processes such as hydroxylation to the santalols.

Like the santalene synthases, the active sites of the bisabolene/bisabolol synthases are sufficiently plastic to accommodate the same series of carbocation rearrangements for both C10 and C15 substrates, as has been found with a cineole synthase mutant from *Salvia fruticosa* (32). *Santalum spicatum* oil contains variable amounts of α-bisabolol relative to other sesquiterpenoids (5) and allelic variation (35) in the ORF of *Sspi*BS, particularly in the α14 helix may be partly responsible for the diversity of phenotypes observed across the distribution of *S. spicatum*. Within-species variation in the santalene synthase ORFs were not evident for the species studied here.

Phylogenetically, *Sau*BS and *Sspi*BS also cluster with the santalene synthases in the TPS-b group (Fig. 3.7). Thus, TPSs of two phylogenetic clades, TPS-a and TPS-b, contribute to the sesquiterpene profiles of sandalwood oils.
Figure 3.7 - Neighbor-joining phylogenetic tree of the TPSs compared in this study. The santalene synthase and bisabolene/bisabolol synthase orthologs group with the TPS-b clade. Abbreviations and GenBankTM accession numbers are listed under “Experimental Procedures”. The asterisk indicates a bootstrap value greater than 95%.

The TPS-a sesquiterpene synthases from the three oil-bearing Santalum species all produced markedly different combinations of sesquiterpenes (supplemental Figs. 3.S7-3.S9). This was true to a lesser extent for the bisabolene/bisabolol synthases (supplemental Figs. 3.S5 and 3.S6) yet in both cases the diversity of function is in contrast to the highly conserved product profiles of the three orthologous santalene synthases (Fig. 3.2).
Genomic sequences revealed that all four species of Santalum studied here had copies of TPS genes, and that no frame shifts or mutations in the open reading frames were apparent (supplemental Fig. 3.S11). Even the genome of S. murrayanum contains a complete ORF of the santalene synthase, despite possessing an oil-deficient phenotype. This suggests factors controlling the spatial or temporal patterns of TPS expression, rather than the absence of, or mutations in the ORFs of the genes themselves, are likely to be responsible for the low- or no-oil phenotype.

All oil producing species of Santalum, and even ancestral genera within the Santalaceae family (36), contain detectable levels of the santalenes, α-bergamotene, bisabolene, and bisabolol, and all species studied here contain complete genomic copies of the santalene synthase ORF. The uniformity of santalene product profiles of SaSSy, SauSSy, and SspiSSy is in contrast to the variety of compounds produced by the bisabolene/bisabolol synthases, and more so by the TPS-a group sesquiterpene synthases in the Santalum genus. The higher amino acid sequence identity of the santalene synthase orthologs (94-98%) compared with the TPS-a genes (89-93%) further support a genetic basis of functional conservation of santalene biosynthesis (supplemental Fig. 3.S10). Function of SauBS and SaMonoTPS (7) is also well conserved between S. album and S. austrocaledonicum (99% identity) however SspiBS differs in both sequence identity (92%) and catalytic function (supplemental Fig. 3.S2). These findings may suggest positive selection for santalene and to a lesser extent bisabolene biosynthesis in the Santalum genus, and possibly in the Santalaceae family more generally. As sesquiterpenes have been implicated in defense against pathogens (6) it is possible that selection pressure in the form of disease has allowed for the survival of populations with the santalene and bisabolene phenotype.
3.6. Conclusions

In conclusion, we have identified several terpene synthases responsible for the production of key fragrant compounds in three commercially exploited species of sandalwood. These findings are of great significance for the flavor and fragrance industry, as well as the growing plantation sandalwood industry.

3.7. Acknowledgements

We thank Len Norris, Karen Reid, Lina Madilao, Maria Li, Hanna Henderson, and Harpreet Kaur for technical support. We thank Drs. Christopher Keeling and Dawn Hall for critical comments and discussion.
3.8. **Supplementary Materials**

3.51 - **Primers used in this study.** Restriction sites are underlined

<table>
<thead>
<tr>
<th>Open reading frame primers (5'→3')</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
<td><strong>Forward primer</strong></td>
<td><strong>Reverse primer</strong></td>
</tr>
<tr>
<td>SoSSy</td>
<td>ATGGATCTTCACCACCCGACGAGC</td>
<td>CCAACGTTACTCTTCCTGCGGAGAG</td>
</tr>
<tr>
<td>SoSSs</td>
<td>ATGGATCTTCACCACCCGACGAGC</td>
<td>CCAACGTTACTCTTCCTGCGGAGAG</td>
</tr>
<tr>
<td>SpSSs</td>
<td>ATGGATCTTCACCACCCGACGAGC</td>
<td>CCAACGTTACTCTTCCTGCGGAGAG</td>
</tr>
<tr>
<td>SoBS</td>
<td>GCTGATTCCGCCGAGGATGCTACCGGACT</td>
<td>CTCAACTCTCTCTGCGGAGAGAG</td>
</tr>
<tr>
<td>SpBS</td>
<td>GCTGATTCCGCCGAGGATGCTACCGGACT</td>
<td>CTCAACTCTCTCTGCGGAGAGAG</td>
</tr>
<tr>
<td>SoSSequi TPS (outer)</td>
<td>CACAGCTCCCGTCTATGCGTCTAGAGC</td>
<td>CATCTATAACGGTAGAAATCCACGAGATG</td>
</tr>
<tr>
<td>SoSSequi TPS (inner)</td>
<td>CATCATACTGAACTGGTGATCGATGAGG</td>
<td>TCAAAACATACAGATTATCAAGTACAG</td>
</tr>
<tr>
<td>SpSSequi TPS (outer)</td>
<td>GCTCCGCTCTATGCGTCTAGAGCAGATC</td>
<td>CTAGAGGAAAAGCTATATATCCACGAGATG</td>
</tr>
<tr>
<td>SpSSequi TPS (inner)</td>
<td>CATCATACTGAACTGGTGATCGATGAGG</td>
<td>TCAAAACATACAGATTATCAAGTACAG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pEF128b(+) cloning primers (5'→3')</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
<td><strong>Forward primer</strong></td>
<td><strong>Reverse primer</strong></td>
</tr>
<tr>
<td>SoSSy</td>
<td>ATCATGATCATCTTCACCACCGACG</td>
<td>ATCTGAGCTCTCTGCCGAGAG</td>
</tr>
<tr>
<td>SoSSs</td>
<td>ATCATGATCATCTTCACCACCGACG</td>
<td>ATCTGAGCTCTCTGCCGAGAG</td>
</tr>
<tr>
<td>SpSSs</td>
<td>ATCATGATCATCTTCACCACCGACG</td>
<td>ATCTGAGCTCTCTGCCGAGAG</td>
</tr>
<tr>
<td>SoBS</td>
<td>TGCTGATTCCGCCGAGGATGCTACCGGACT</td>
<td>CGAATGAACTGACCTCTCTGCTAGG</td>
</tr>
<tr>
<td>SpBS</td>
<td>TGCTGATTCCGCCGAGGATGCTACCGGACT</td>
<td>CGAATGAACTGACCTCTCTGCTAGG</td>
</tr>
<tr>
<td>SoSSequi TPS</td>
<td>AGCTACATGACCGAAAATACAAAGGCTGCTTTGCTTTT</td>
<td>AGAAATGAAATCAACATGAGATCTTATAAGTA</td>
</tr>
<tr>
<td>SpSSequi TPS</td>
<td>AGCTACATGACCGAAAATACAAAGGCTGCTTTT</td>
<td>AGAAATGAAATCAACATGAGATCTTATAAGTA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pCDFDuet-1 cloning primers (5'→3')</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
<td><strong>Forward primer</strong></td>
<td><strong>Reverse primer</strong></td>
</tr>
<tr>
<td>SpFFPS</td>
<td>ATGCACTGATCATCTTCACCACCGACG</td>
<td>AGTTGGAGCGCTACCTTCCTGCGGAGAGAG</td>
</tr>
<tr>
<td>SpSSy</td>
<td>AAGAGATATATACATGATGATGATCTCCCGAGG</td>
<td>TTACGAGCTCGAGGAGATTTGCTTACACCAGCTACCTCCCGAGG</td>
</tr>
</tbody>
</table>

(Accession number for forward primer) 

(Accession number for reverse primer)
3.52 - Sequences and abbreviated names used for phylogenetic tree (Fig 3.7): SaSesquiTPS, *S. album* sesquiterpene synthase (ACF24768); SauSesquiTPS, *S. austrocaledonicum* sesquiterpene synthase (HQ343281); SmSesquiTPS, *S. murrayanum* sesquiterpene synthase (deduced from gDNA sequence JF746810); SspiSesquiTPS, *S. spicatum* sesquiterpene synthase (HQ343282); PtrGDS, *Populus trichocarpa x deltoides* (-)-germacrene D synthase (AAR99061); CanEAS, *Capsicum annuum var. annuum* 5-epi-aristolochene synthase (CAA06614); AanADS, *Artemisia annua* amorph-4,11-diene synthase (AAF61439); OsaSTPS, *Oryza sativa* sesquiterpene synthase (ACF05529); PabCS, *Picea abies* (+)-3-carene synthase (AAO73863); AgrPSA, *Abies grandis* β-phellandrene synthase (AAO61453); AgrDSS, *Abies grandis* delta-selinene synthase (AACO5727); AgrHS, *Abies grandis* γ-humulene synthase (AACO5728); PabLAS, *Picea abies* levopimaradiene/AS synthase, (AAC47691); FaNES2, *Fragaria ananassa* nerolidol synthase (CAD57081); AmaAMS, *Antirrhinum majus* myrcene synthase (AA041727); AthLIS, *Arabidopsis thaliana* linalool synthase (NP176361); SauSSy, *S. austrocaledonicum* santalene synthase (HQ343277); SmSSy, *S. murrayanum* santalene synthase (deduced from gDNA sequence JF746811); SaSSy, *S. album* santalene synthase (HQ343276); SspiSSy, *S. spicatum* santalene synthase (HQ343278); SaMonoTPS, *S. album* monoterpenoid synthase (ACF24767); SauBS, *S. austrocaledonicum* β-bisabolene synthase (HQ343279); SmBS, *S. murrayanum* bisabolene synthase (deduced from gDNA JF746814) SspiBS, *S. spicatum* α-bisabolol synthase (HQ343280); BcLIS, *Backhousia citriodora* linalool synthase (BAG82825); ObAZS, *Ocimum basilicum* α-zingiberene synthase (AAV63788); SfCINS, *Salvia fruticosa* cineole synthase (DQ785793); VvTS, *Vitis vinifera* α-terpineol synthase (AA579351); MgATS, *Magnolia grandifolia* α-terpineol synthase (ACC662822); EgMTPS, *Eucalyptus globulus* monoterpenoid synthase (BAF02832)
3.53 - SDS-PAGE of Ni$^{2+}$ affinity purified recombinant proteins. A; *S. spicatum* TPS enzymes. B; *SaSSy* and *S. austrocaledonicum* TPS enzymes. Expression of *SauSesquiTPS* and *SspiSesquiTPS* (approximate size 68 kDa) was particularly poor.
3.54 - Product profiles of *SaSSy*, *SauSSy* and *SspiSSy*. Positive identification of products was based on comparison to authentic standards and mass spectral data, along with comparison to Kovats retention indices quoted in the literature. * = run on a different DB-WAX column under slightly different oven conditions: 1 min at 40°C then 10°C min^{-1} to 230°C held 25 min. nd = none detected

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Product</th>
<th>Ret. time (min)</th>
<th>KI (DB-WAX)</th>
<th>Lit. KI</th>
<th>Ng[^a]</th>
<th>Mn[^b]</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>SaSSy</em></td>
<td>(E,E)-FPP</td>
<td>α-santalene</td>
<td>15.10</td>
<td>1581</td>
<td>1577</td>
<td>45.3</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-exo-bergamotene</td>
<td>15.29</td>
<td>1592</td>
<td>1586</td>
<td>16.0</td>
<td>74.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>epil-β-santalene</td>
<td>16.13</td>
<td>1645</td>
<td>1637</td>
<td>4.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-santalene</td>
<td>16.36</td>
<td>1659</td>
<td>1651</td>
<td>31.5</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-β-farnesene</td>
<td>16.47</td>
<td>1666</td>
<td>1668</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E,E)-α-farnesene</td>
<td>16.93</td>
<td>1695</td>
<td>1698</td>
<td>1.7</td>
<td>7.8</td>
</tr>
<tr>
<td><em>SaSSy</em></td>
<td>(Z,Z)-FPP</td>
<td>α-exo-bergamotene</td>
<td>12.79</td>
<td>1573</td>
<td>1570</td>
<td>16.0</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-santalene</td>
<td>12.84</td>
<td>1577</td>
<td>1577</td>
<td>29.4</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-exo-bergamotene</td>
<td>12.99</td>
<td>1590</td>
<td>1586</td>
<td>0.6</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Z)-β-farnesene</td>
<td>13.53</td>
<td>1636</td>
<td>1636</td>
<td>24.4</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>epil-β-santalene</td>
<td>13.60</td>
<td>1641</td>
<td>1640</td>
<td>23.0</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-santalene</td>
<td>13.77</td>
<td>1656</td>
<td>1654</td>
<td>6.5</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-pinene</td>
<td>4.85</td>
<td>1019</td>
<td>1013</td>
<td>2.4</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>camphene</td>
<td>5.59</td>
<td>1061</td>
<td>1054</td>
<td>0.3</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-pinene</td>
<td>6.39</td>
<td>1106</td>
<td>1096</td>
<td>2.4</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>myrcene</td>
<td>7.49</td>
<td>1163</td>
<td>1158</td>
<td>8.2</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limonene</td>
<td>8.17</td>
<td>1198</td>
<td>1189</td>
<td>14.7</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Z)-β-ocimene</td>
<td>8.87</td>
<td>1235</td>
<td>1225</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-β-ocimene</td>
<td>9.19</td>
<td>1252</td>
<td>1242</td>
<td>2.0</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpinolene</td>
<td>9.82</td>
<td>1285</td>
<td>1279</td>
<td>0.7</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>linalool</td>
<td>14.49</td>
<td>1550</td>
<td>1552</td>
<td>22.6</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpineol</td>
<td>16.98</td>
<td>1706</td>
<td>1695</td>
<td>6.6</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-geranool</td>
<td>19.14</td>
<td>1853</td>
<td>1840</td>
<td>39.0</td>
<td>40.4</td>
</tr>
<tr>
<td><em>SauSSy</em></td>
<td>(E,E)-FPP</td>
<td>α-santalene</td>
<td>15.09</td>
<td>1581</td>
<td>1575</td>
<td>51.3</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-exo-bergamotene</td>
<td>15.28</td>
<td>1592</td>
<td>1586</td>
<td>14.5</td>
<td>73.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>epil-β-santalene</td>
<td>16.12</td>
<td>1645</td>
<td>1637</td>
<td>4.8</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-santalene</td>
<td>16.35</td>
<td>1659</td>
<td>1651</td>
<td>27.4</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-β-farnesene</td>
<td>16.46</td>
<td>1666</td>
<td>1668</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E,E)-α-farnesene</td>
<td>16.92</td>
<td>1695</td>
<td>1698</td>
<td>1.4</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-pinene</td>
<td>4.85</td>
<td>1019</td>
<td>1013</td>
<td>1.5</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>camphene</td>
<td>5.59</td>
<td>1061</td>
<td>1054</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-pinene</td>
<td>6.39</td>
<td>1106</td>
<td>1098</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>myrcene</td>
<td>7.49</td>
<td>1163</td>
<td>1158</td>
<td>5.4</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limonene</td>
<td>8.17</td>
<td>1198</td>
<td>1189</td>
<td>12.5</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Z)-β-ocimene</td>
<td>8.87</td>
<td>1235</td>
<td>1225</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-β-ocimene</td>
<td>9.19</td>
<td>1252</td>
<td>1242</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpinolene</td>
<td>9.82</td>
<td>1285</td>
<td>1279</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>linalool</td>
<td>14.49</td>
<td>1550</td>
<td>1552</td>
<td>14.7</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpineol</td>
<td>16.98</td>
<td>1706</td>
<td>1695</td>
<td>4.8</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-geranool</td>
<td>19.14</td>
<td>1853</td>
<td>1849</td>
<td>42.0</td>
<td>41.4</td>
</tr>
<tr>
<td></td>
<td>(E,E)-FPP</td>
<td>SipSSy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-santalene</td>
<td>15.09</td>
<td>1581</td>
<td>1575</td>
<td>47.3</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-endo-bisnaphthyl</td>
<td>15.28</td>
<td>1592</td>
<td>1586</td>
<td>19.2</td>
<td>78.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-pinene</td>
<td>16.13</td>
<td>1645</td>
<td>1637</td>
<td>4.2</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-santalene</td>
<td>16.35</td>
<td>1659</td>
<td>1651</td>
<td>36.3</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-β-farnesene</td>
<td>16.46</td>
<td>1666</td>
<td>1668</td>
<td>0.7</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E,E)-α-farnesene</td>
<td>16.92</td>
<td>1695</td>
<td>1698</td>
<td>2.2</td>
<td>8.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-pinene</td>
<td>4.35</td>
<td>1019</td>
<td>1013</td>
<td>3.2</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>camphene</td>
<td>5.59</td>
<td>1051</td>
<td>1054</td>
<td>1.0</td>
<td>tr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-pinene</td>
<td>6.30</td>
<td>1106</td>
<td>1098</td>
<td>0.8</td>
<td>tr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>myrcene</td>
<td>7.49</td>
<td>1153</td>
<td>1158</td>
<td>25.7</td>
<td>27.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>limonene</td>
<td>8.17</td>
<td>1189</td>
<td>1189</td>
<td>14.9</td>
<td>16.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Z)-ocimene</td>
<td>8.37</td>
<td>1225</td>
<td>1225</td>
<td>11.3</td>
<td>11.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-ocimene</td>
<td>9.10</td>
<td>1242</td>
<td>1242</td>
<td>14.0</td>
<td>14.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-pinanol</td>
<td>9.82</td>
<td>1297</td>
<td>1297</td>
<td>2.9</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linalool</td>
<td>14.49</td>
<td>1552</td>
<td>1552</td>
<td>12.3</td>
<td>13.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-terpinol</td>
<td>16.98</td>
<td>1695</td>
<td>1695</td>
<td>5.5</td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-geraniol</td>
<td>19.14</td>
<td>1849</td>
<td>8.4</td>
<td>8.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.55 - GC-MS chromatogram and mass spectra of sesquiterpenes produced by *SauBS* and *SspiBS* when incubated with *(E,E)*-FPP and Mg$^{2+}$
3.56 - Product profiles of SauBS and SspiBS. tr = trace

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Product</th>
<th>Ret. Time (min)</th>
<th>KI (DB-WAX)</th>
<th>Lit. KI</th>
<th>Mg$^{2+}$</th>
<th>Mn$^{2+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SauBS</strong></td>
<td>(E,E)-FPP</td>
<td>(E)-β-farnesene</td>
<td>16.74</td>
<td>1656</td>
<td>1668</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-bisabolene</td>
<td>17.68</td>
<td>1717</td>
<td>1726</td>
<td>88.7</td>
<td>86.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-bisabolene</td>
<td>18.33</td>
<td>1761</td>
<td>1778</td>
<td>1.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-bisabolol</td>
<td>25.73</td>
<td>2188</td>
<td>2153</td>
<td>9.0</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>myrcene</td>
<td>7.65</td>
<td>1139</td>
<td>1158</td>
<td>5.5</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limonene</td>
<td>8.35</td>
<td>1176</td>
<td>1189</td>
<td>25.2</td>
<td>24.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unknown</td>
<td>12.08</td>
<td>1377</td>
<td>9.1</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>linalool</td>
<td>14.57</td>
<td>1532</td>
<td>1552</td>
<td>7.1</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpineol</td>
<td>17.16</td>
<td>1683</td>
<td>1695</td>
<td>50.9</td>
<td>47.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-geraniol</td>
<td>19.36</td>
<td>1831</td>
<td>1849</td>
<td>2.2</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>SspiBS</strong></td>
<td>(E,E)-FPP</td>
<td>(E)-β-farnesene</td>
<td>16.74</td>
<td>1656</td>
<td>1668</td>
<td>4.6</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-bisabolene</td>
<td>17.68</td>
<td>1717</td>
<td>1726</td>
<td>19.8</td>
<td>18.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unknown</td>
<td>17.99</td>
<td>1738</td>
<td>8.9</td>
<td>7.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E,E)-α-farnesene</td>
<td>18.12</td>
<td>1746</td>
<td>1751</td>
<td>2.7</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-bisabolene</td>
<td>18.33</td>
<td>1761</td>
<td>1778</td>
<td>10.5</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-bisabolol</td>
<td>25.73</td>
<td>2188</td>
<td>2153</td>
<td>53.6</td>
<td>57.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>limonene</td>
<td>8.34</td>
<td>1174</td>
<td>1189</td>
<td>tr.</td>
<td>tr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>linalool</td>
<td>14.60</td>
<td>1523</td>
<td>1552</td>
<td>tr.</td>
<td>tr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>α-terpineol</td>
<td>17.29</td>
<td>1691</td>
<td>1695</td>
<td>tr.</td>
<td>tr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(E)-geraniol</td>
<td>19.35</td>
<td>1762</td>
<td>1849</td>
<td>tr.</td>
<td>tr.</td>
</tr>
</tbody>
</table>
3.57 - GC-MS trace and mass spectra of sesquiterpenes produced by SauSesquiTPS when incubated with \((E,E)\)-FPP and Mg\(^{2+}\)
3.8 - GC-MS trace and mass spectra of sesquiterpenes produced by *Ssp/SesquiTPS* when incubated with \((E,E)\)-FPP and Mg\(^{2+}\)
### 3.59 - Product profiles of SauSesiTPS and SspiSesiTPS. tr = trace, nd = none detected

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Product</th>
<th>Ret. time (min)</th>
<th>KI (DB-WAX)</th>
<th>Lit. KI</th>
<th>composition (% TIC)</th>
<th>Mg&lt;sup&gt;2+&lt;/sup&gt;</th>
<th>Min&lt;sup&gt;2+&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SauSesiTPS</strong> (E,E)-FPP</td>
<td>β-elemene</td>
<td>16.05</td>
<td>1600</td>
<td>1595</td>
<td>5.7</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-caryophyllene</td>
<td>16.23</td>
<td>1613</td>
<td>1608</td>
<td>6.9</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>16.80</td>
<td>1653</td>
<td></td>
<td></td>
<td>2.3</td>
<td>tr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>α-humulene</td>
<td>17.26</td>
<td>1686</td>
<td>1680</td>
<td>25.7</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-cadinene?</td>
<td>17.80</td>
<td>1725</td>
<td></td>
<td></td>
<td>7.2</td>
<td>tr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cadina-1,4-diene</td>
<td>17.90</td>
<td>1733</td>
<td>1769</td>
<td>3.6</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>γ-cadinene</td>
<td>18.40</td>
<td>1770</td>
<td>1776</td>
<td>11.2</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ-cadinene</td>
<td>18.47</td>
<td>1775</td>
<td>1765</td>
<td>20.7</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>α-cadinene</td>
<td>18.89</td>
<td>1807</td>
<td>1785</td>
<td>3.7</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>germacrene D-4-ol</td>
<td>22.69</td>
<td>2031</td>
<td>2047</td>
<td>9.2</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPP</td>
<td>linalool</td>
<td>15.52</td>
<td>1564</td>
<td>1552</td>
<td>tr</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SspiSesiTPS</strong> (E,E)-FPP</td>
<td>α-bulnesene</td>
<td>17.86</td>
<td>1730</td>
<td>1729</td>
<td>3.6</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-elemol</td>
<td>23.21</td>
<td>2094</td>
<td>2069</td>
<td>45.2</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>guaiol</td>
<td>23.45</td>
<td>2104</td>
<td>2077</td>
<td>19.4</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>bulnesol</td>
<td>26.00</td>
<td>2227</td>
<td>2202</td>
<td>31.8</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPP</td>
<td>linalool</td>
<td>15.08</td>
<td>1552</td>
<td>1552</td>
<td>tr</td>
<td>tr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.510 - ClustalX protein alignment of TPS genes in this study. Germacrene D synthase from *Populus trochodora x deltoides*, 5-epi-aristolochene synthase from *Capsicum annuum var. annum* and cineole synthase from *Salvia fruticosa* are included for comparison. R(P/R)X₉W and DDXXD motifs are indicated with a black overhead line. Red overhead lines indicate positions most likely to influence product profiles and substrate utilisation, as identified by Kampranis et al. (1). Gray histogram indicates degree of conservation of residues.
3.511 - Genomic structures of deduced open reading frames of TPS genes amplified from *S. album*, *S. austrocaledonicum*, *S. spicatum* and *S. murrayanum*. Exons are in red. R(P/R)X₈W and DDXXD motifs are highlighted. Double dagger indicates a 10 nucleotide deletion in the *SaSSy* gDNA sequence and asterisk indicates a stop codon in exon 3 of *SmBS*; an *S. murrayanum* bisabolene synthase-like gene.
3.9. References


CHAPTER FOUR

Essential Oil Components and their Biosynthesis in West Australian Sandalwood Oil (*Santalum spicatum*)

Jessie Moniodis¹,²*, Christopher G Jones¹, Julie A Plummer¹, Liz Barbour¹, Emilio Ghisalberti², Joerg Bohlmann³

1. School of Plant Biology, University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

2. School of Chemistry and Biochemistry, University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

3. Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T1Z4, Canada

* corresponding author
4.1. Abstract

Western Australian sandalwood, *Santalum spicatum* (R.Br.) A.DC. is a major export industry in Western Australia. It produces an essential oil which is less valuable than other commercial species of sandalwood due to considerable variation in oil quality and yield across its natural distribution. There is good potential for tree improvement through a better understanding of the natural chemical diversity of this species, and processes involved in sesquiterpene production. The heartwood-oil quality of 147 wild *S. spicatum* trees distributed throughout the semi-arid region of Western Australia, and nine individuals from an 11 year-old plantation was investigated. There was high variation in total α- and β- santalol content across all individuals ranging from 1-54% of the total mixture, with some trees approaching the quality expected for *S. album*. Most individuals contained considerable levels of *E,E*-farnesol (an undesired sesquiterpene component) in quantities ranging from 2-46%, however a few trees contained less than 2%. The total oil content ranged from 0.02-5.9% with considerably lower yields in younger plantation trees compared to wild-sourced trees. A summary of the enzymes involved in oil production in *S. spicatum* is provided to guide future research into the improvement of this tree species.
4.2. Introduction

Sandalwood is highly valued for the fragrant essential oils contained in its heartwood (Adams 1995). Several members of the genus are used for carved ornaments, joss sticks (incense) while the essential oil of *S. album* is used predominantly in perfumery, medicine and aromatherapy (Rai 1990; Brennan and Merlin 1993; Jones 2001). Sandalwood products are preferentially derived from *S. album* (Srinivasan et al. 1992), however, natural resources are in decline (Clarke 2006). To fill the gap, additional species of sandalwood are contributing to the international sandalwood market. Western Australian sandalwood, *S. spicatum* provides an alternative source of sandalwood, which contributes to the local economy and export market, mainly in the form of ground wood for incense sticks (Statham 1990). It is among the least valuable of the sandalwood species currently harvested. The two main reasons why *S. spicatum* does not meet the international standard of oil quality (ISO:3518 2002) for *S. album* are that (i) it contains high levels of *E*,*E*-farnesol, a potential allergen (Lapczynski et al. 2008), and (ii) the α- and β-santalol content, which imparts the ultimate value of the extracted oil, is often low and highly variable. Other sandalwood species such as *S. album*, *S. yasi*, *S. austrocaledonicum* and *S. macgregorii* have or continue to be exploited instead. Hence, there is a need to improve the market value of this species, particularly as it develops into managed plantations, and to ensure a sustainable sandalwood industry into the future.

*Santalum spicatum* is a hemi-parasitic tree or shrub native to South-western Australia. Its distribution extends from the semi-arid southern regions (300–600 mm annual rainfall) to the arid northern areas of Western Australia (150–300 mm rainfall) and
some parts of South Australia (Fox and Brand 1993; Loneragan 1990). At present, *S. spicatum* is harvested mainly from natural stands in the semi-arid regions of Western Australia (Brand et al. 2007). Over the past ten years, there has been a rapid expansion of plantations in Southwest Australia with over 13,000 ha dedicated to *S. spicatum* (McKinnell et al. 2008). This will enable a sustainable supply of sandalwood into the future, however, the selection of superior trees with consistent heartwood quality needs to occur. Current plantations of *S. spicatum* remain variable and a minimum of 25 years has been suggested to obtain high-grade wood (Brand et al. 2007; Brand and Pronk 2011). The butt and roots are the most valuable parts of the tree, as they generally contain the highest concentration of oil. However, in an early examination of a single tree, Piggot et al. (1997) found compositional changes of major sesquiterpenes in different sections, with santalol content being highest in the base of the stem, decreasing in concentration further up the tree. A similar study in *S. album* indicated no significant variation in composition along different sections of the trees (Jones et al. 2006). This highlights challenges for tree improvement in *S. spicatum* which will be enhanced by a more complete understanding of natural chemotypes and the mechanisms involved in oil production.

The oil of *S. spicatum* contains most of the same sesquiterpenes as *S. album*, however *E,E*-farnesol, lanceol and nuciferol are also abundant (Piggot et al. 1997). The main components in *S. spicatum* oil are *E,E*-farnesol, α- and β-santalol (the hydroxylated products of α- and β-santalene), dendrolasin and α-bisabolol (Brophy et al. 1991) (Fig. 4.1). Marketing opportunities are limited while this remains the case, particularly for perfume and cosmetic products. The value of *S. spicatum* is less predictable than in *S. album* due to variation in α- and β-santalol, the presence of *E,E*-farnesol, in addition to
the generally lower oil content. To better understand heartwood production and know whether oil composition within a tree can be manipulated, the biosynthetic pathway of sesquiterpenes in the oil must first be elucidated.

Figure 4.1 - Sesquiterpenes found in *S. spicatum* essential oil

Sesquiterpenes are derived from the five carbon precursors fundamental to all isoprenoids; isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Successive head to tail condensation of IPP and two molecules of DMAPP results in an acyclic (C15) precursor for sesquiterpenes, farnesyl pyrophosphate (FPP). FPP is a substrate for a very large family of enzymes, the terpene synthases (TPS), which are ubiquitous across the plant and fungi kingdoms (Croteau and Cane 1985; Tholl 2006). Recently, several TPS were characterized from divergent sandalwood species, *S. spicatum*, *S. album* and *S. austrocaledonicum* including a santalene and bisabolol synthase (Jones et al. 2011). In this study, the gene sequences were found to be highly conserved, suggesting the production of these sesquiterpenes may have been critical adaptations to various biotic selection pressure. While our understanding of oil production in *Santalum* is enhanced by these results, further work into sesquiterpene biosynthesis and causes of chemical variation within *S. spicatum* is needed to improve
its commercial value and management of plantations into the future. The aim of this study was to (i) compare oil quality; \( \alpha \)- and \( \beta \)-santalol and \( E,E \)-farnesol concentrations of individuals distributed throughout the semi-arid region of Western Australia (the wheatbelt) where sandalwood is currently sourced and plantations are being developed, (ii) compare the oil quality of these wild-sourced sandalwood trees to those of nine individuals from an 11 year-old plantation based on seed from these regions, (iii) summarise the current information of enzymes involved in oil production within \( S. \) spicatum. These results will provide a foundation for the further development of the Australian sandalwood industry and improve plantation management for enhanced oil quality.
4.3. Experimental

4.3.1. Plant collection

Mature *S. spicatum* trees (147) were sampled from natural stands throughout the semi-arid wheatbelt regions of Western Australia, as well as nine individuals from a southern wheatbelt plantation (Narrogin). The latitude and longitude range of the collection site was 30 to 34°S and 116 to 119°E.

4.3.2. Oil extraction

Mature sandalwood trees were drilled at 30 cm (butt) using a manual hand drill. Wood shavings were ground to a fine powder and extracted in 25 ml volumetric flasks in hexane for 7 days (Fig 4.2). Samples were analysed using gas chromatography mass spectrometry (GC-MS) on a Shimadzu GC MS–QP2010 with a DB-WAX capillary column (30 m, 0.25 mm ID, 0.25 µm film thickness). Injection volume was 1µl, split 10:1. Injector temperature was 200° C, detector 250° C. Oven temperature program was held at 40° C for 3 min, then ramped at 4° C per minute to 240° C, where it was held for a further 10 min. An external standard curve was generated using authentic *S. spicatum* oil (Mount Romance, Albany, Western Australia) as well as an internal standard (30 µM of isobutyl benzene) for consistent quantification. Mass spectra were analyzed using the Shimadzu GC-MS software (ssi.shimadzu.com) and compared to the
2008 NIST library (nist.gov). Kovats retention indices were calculated for all compounds using a C10-C30 alkane standard, and compared to the literature.

Figure 4.2 - Collection and extraction of *S. spicatum* oil; A. Sandalwood tree drilled at 30 cm; B. Grind wood shavings; C. Ground wood shavings extracted in 25 mL volumetric flasks using hexane for 7 days followed by D. Analysis by GC-MS
4.4. Results and Discussion

4.4.1. Heartwood oil analysis

Across the wheatbelt, the mean oil concentration of wild-sourced individuals at 30 cm (butt) was 3.27 ± 1.02% and ranged from 0.71-6% (Table 4.1). For comparison, 9 individuals from an 11 year-old plantation yielded less oil with an average of 1.78 ± 0.79% and ranged from 0.02-2.5%. Considerable variation in oil yield was noted for trees distributed naturally in the wheatbelt region both within and among collection sites. These results are similar to previous investigations into natural stands of *S. spicatum* with 2.3-3.1% total oil yields (Brand et al. 2007). A plantation study of 8-11 year-old trees reported 1-1.9% oil by dry weight, and 2.6 ± 0.2% in a 26 year-old plantation, indicating an increase in oil content with age (Brand and Pronk 2011). The age of sampled trees in the wild was uncertain, so variation associated with age is not accounted for in this study. Since the sampling region may have undergone harvesting and restoration since establishment of the sandalwood industry, the range of tree-ages may be substantial. In the case of *S. austrocaledonicum*, Page et al. (2010) also found considerable variation in oil yield and santalol content in wild populations across Vanuatu. Even *S. album*, which has displayed the highest consistency in santalol content (Verghese et al. 1990), has shown variation in oil concentrations from none to nearly 9%, and this is also influenced by age (Jones et al. 2006, 2007). Variation in heartwood oil concentration is complex and age is almost certainly a compounding factor. Further research into genetic and environmental factors which contribute to the overall oil variability needs to be done.
Heartwood quality may broadly be determined by assessing the overall levels of α- and β-santalol, as well as \( E,E \)-farnesol. Total santalol content for wild-sourced trees ranged from less than 1% to 54%; more than half the total oil composition (Table 4.2). The α- and β-santalol levels for plantation trees similarly varied from 7-43%. No individuals met the ISO standard (ISO:3518 2002) for \( S. \) album (41-55% \( \alpha \)-santalol and 16-24% \( \beta \)-santalol), although one individual contained 41% \( \alpha \)-santalol and 13% \( \beta \)-santalol. The most variable component was \( \alpha \)-santalol followed by \( E,E \)-farnesol. Most individuals contained levels of \( E,E \)-farnesol, similar to those of \( \alpha \)- and \( \beta \)-santalol (2-46%) (Table 4.2). Plantation trees contained similar levels of \( E,E \)-farnesol with concentrations varying from 12-31%. Across all individuals from natural stands, mean \( E,E \)-farnesol and santalol contents were similar with 14 (± 8%) for \( E,E \)-farnesol and 12 (± 10%) total santalols. Variation in chemical composition has been noted throughout the root and stem, and \( \alpha \)- and \( \beta \)-santalol concentrations are higher in the mid-trunk and roots and \( E,E \)-farnesol increases up the height of the tree (Piggot et al. 1997). This spatial variation must also be considered when comparing the individual trees.
Table 4.2 - Total santalols and E-E-farnesol as percentage of total oil composition of *S. spicatum* heartwood samples extracted from wild trees (*N* = 147) and individuals from an 11 year-old plantation (*N* = 9)

<table>
<thead>
<tr>
<th>Compound</th>
<th>Mean (%)</th>
<th>Max (%)</th>
<th>Min (%)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total santalol (wild trees)</td>
<td>11.77</td>
<td>54.27</td>
<td>0.93</td>
<td>10.28</td>
</tr>
<tr>
<td>E,E-farnesol (wild trees)</td>
<td>13.95</td>
<td>46.22</td>
<td>1.99</td>
<td>8.26</td>
</tr>
<tr>
<td>Total santalol (plantation trees)</td>
<td>26.21</td>
<td>43.55</td>
<td>7.12</td>
<td>12.50</td>
</tr>
<tr>
<td>E,E-farnesol (plantation trees)</td>
<td>20.69</td>
<td>30.56</td>
<td>12.56</td>
<td>6.56</td>
</tr>
<tr>
<td>Total</td>
<td>53.83</td>
<td>81.33</td>
<td>24.56</td>
<td>9.4</td>
</tr>
</tbody>
</table>

*E,E*-farnesol is not a major component of other oil bearing sandalwoods such as *S. album, S. insulare* and *S. macgregorii* (Baldovini et al. 2011) which have overall higher concentrations of α- and β-santalol. Braun and Hammerschmidt (2005) reported a 1% concentration of *E,E*-farnesol in a single *S. austrocaledonicum* tree while *S. album* has ≤ 0.1% (Brand et al. 2007). *S. macgregorii* does not contain farnesol (Baldovini et al. 2011) but like *S. spicatum* variation in α- and β-santalol content (0.9-46% α-santalol and 0.8-20% β-santalol) exists among different provenances. In *S. austrocaledonicum* distributed across Vanuatu, large variations in α- and β-santalol levels are evident (0.8-47% α-santalol and none to 24% β-santalol) (Page et al. 2010).

A combination of genetic and environmental factors such as host type, number of surrounding hosts, age and climatic conditions (soil composition, microbes, rainfall) are all likely to contribute to total oil composition and yield. There was considerable variation in the environment of the different collection sites in the trees sourced from the wheatbelt (some examples in Fig. 4.3). To date, little research has been conducted into this area, presumably because maintaining suitable controls is very difficult.
Clonally propagated sandalwood would account for environmental factors, but *S. spicatum* is notoriously difficult to produce from somatic tissue (McComb and Jones 1998). Future research directed towards understanding the causes of chemical variation within *Santalum* will improve current management of plantations and possibly enable the selection of superior trees prior to harvesting. *S. spicatum* remains a valuable source of sandalwood and an important contributor to the economy of Western Australia providing additional environmental advantages. There was also no indication that trees with high levels of \( \alpha \)- and \( \beta \)-santalol contained less *E*,*E*-farnesol. Such a negative co-occurrence pattern would indicate a dominance of santalol production over other components (Jones et al. 2006). Evidently, *E*,*E*-farnesol is being generated through a different set of enzymes, with similarly unrelated regulation. Characterization of enzymes involved in *E*,*E*-farnesol biosynthesis may help identify the source of this variation and assist in understanding the mechanisms which trigger oil production and their regulation.
4.4.2. Review of Heartwood Oil Production in *S. spicatum*

The oil profile of *S. spicatum* is a complex mixture with more than 100 sesquiterpenes (Adams et al. 1975, Baldovini et al. 2011). To date, three TPS genes from *S. Spicatum*, responsible for the production of major components in the oil, have been characterized (Jones et al. 2011). The GC-MS oil profile of the tree from which TPS genes *SspiSSy,*
SspiBS and SspiSesqui were recently characterized is illustrated in Fig 4.4a with the enzymes respective metabolic products labelled in Fig 4.4b. E,E-farnesol is a major component in the heartwood extract of this individual, whereas α- and β-santalol are less abundant. Based on the heartwood composition of all 156 individuals in this study, the olefins produced by SspiSSy when incubated with FPP (α, β-, epi-β-santalene and α-trans-bergamotene) account for 1-5% of the total oil. For SspiBS (the enzyme which produces α- and β-bisabolene, as well as α-bisabolol) the products contribute from 1-18% of the total oil composition. The sesquiterpene products of SspiSesqui, account for less than 1% of the total oil. These results are fairly consistent with the homologous enzymes characterized in S. album (Jones et al. 2011). The TPS enzymes characterized can account for up to 20% of the total components of sandalwood oil, and based on the analysis presented here, this is likely to vary within Santalum species. Although progress has been made towards understanding oil production in Santalum, the enzyme or enzymes responsible for other components such as E,E-farnesol, as well as the hydroxylated analogues of the santalenes are yet to be fully elucidated. With this in mind, future work on the biosynthesis and exploitation of S. spicatum should include a deeper search for the biosynthesis of E,E-farnesol, as it has a direct impact on the quality and marketability of the extracted oil. There would then exist the potential to produce knock-out or knock-down lines in plantations.
Figure 4.4. a) GC-MS oil profile of *S. spicatum*. Sesquiterpenoids known to be produced by characterised TPS enzymes are highlighted with arrows according to their retention time. The percent composition of each compound is based on extracts of 156 *S. spicatum* individuals from the semi-arid wheatbelt region of Western Australia. b) Summary of TPS enzymes from *S. spicatum* (Jones et al. 2011) and the products resulting from incubations with FPP (abbreviations: FPP, farnesyl pyrophosphate; *Sspi*SSy, *S. spicatum* santalene synthase; *Sspi*BS, *S. spicatum* bisabolol synthase; *Sspi*Sesqui, *S. spicatum* sesquiterpene synthase; TPS, terpene synthase)
4.5. Conclusions

Western Australian sandalwood, *S. spicatum*, has long provided an alternative source of sandalwood products while sources of the preferred *S. album* are in decline (Clarke 2006). Currently, the market potential for *S. spicatum* is limited, however, great potential exists to improve oil quality, leading to sustainable production in plantations. These plantations are already having a positive impact on the agricultural regions of South Western Australia. Improvement of *S. spicatum* will be achieved through a better understanding of natural oil diversity and the mechanisms involved in the oil production, including enzymes such as TPS which produce sesquiterpenes in *Santalum*. This study sought to highlight variation in oil quality (assessed by α- and β-santalol content and *E,E*-farnesol) across the species distribution in the semi-arid regions of Western Australia, where past and current harvesting practises occur. Substantial progress has been made towards understanding the oil production in this species and identification of new TPS involved in oil production is on-going. Tree improvement and better management of natural stands and plantations will progress through continued research into physiological factors which control oil production, as well as any genetic predispositions to specific chemical phenotypes.
4.6. Acknowledgements

The work was supported by ARC Lingage project LP0882690. The support and assistance of the Forest Products Commission (Perth, Western Australia) and the University of British Columbia (Vancouver, Canada) is gratefully acknowledged.
4.7. References

Adams DR, Bhatnagar SP, Cookson RC (1975) Sesquiterpenes of Santalum album and Santalum spicatum. Phytochem 14:1459-1460


Baldovini, N, Delasalle C., Joulain D (2011) Phytochemistry of the heartwood from fragrant Santalum species: a review. Flavour Frag J 26:7–26


Brand JE, Pronk GM (2011) Influence of age on sandalwood (Santalum spicatum) oil content within different wood grades from five plantations in Western Australia. Aust Forestry 74:141-148


Loneragan OW (1990) Historical review of sandalwood (*Santalum spicatum*) research in Western Australia. Published by the Department of Conservation and Land Management (CALM). Perth, Australia


Srinivasan VV, Sivaramakrishnan VR, Rangaswamy CR, Ananthapadmanabha HS, Shankaranarayana KH (1992) Sandal (Santalum album). In Indian Council of Forestry Research and Education. Dehradun, India


Vergheese J, Sunny TP, Balakrishnan KV (1990) (Z)-(+)α-santalol and (Z)-(−)β-santalol concentration, a new quality determinant of east Indian sandalwood oil. Flavour Frag J 5:223-226
CHAPTER FIVE

Biosynthesis of Sandalwood Oil: *Santalum album* CYP76F

Cytochromes P450 Produce Santalols and Bergamotol

Maria L. Diaz-Chavez¹, Jessie Moniodis¹,²,³, Lufiani L. Madilao¹, Sharon Jancsik¹, Christopher I. Keeling¹, Elizabeth L. Barbour³, Emilio L. Ghisalberti³, Julie A. Plummer², Christopher G. Jones², and Joerg Bohlmann¹,*

¹ Michael Smith Laboratories, University of British Columbia, 301-2185 East Mall, Vancouver, BC, V6T 1Z4, Canada

² School of Plant Biology, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

³ School of Chemistry and Biochemistry, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

* corresponding author
5.1. Abstract

Sandalwood oil is one of the world’s most highly prized essential oils, appearing in many high-end perfumes and fragrances. Extracted from the mature heartwood of several Santalum species, sandalwood oil is comprised mainly of sesquiterpene olefins and alcohols. Four sesquiterpenols, α-, β-, and epi-β-santalol and α-exo-bergamotol, make up approximately 90% of the oil of Santalum album. These compounds are the hydroxylated analogues of α-, β- and epi-β-santalene and α-exo-bergamotene. By mining a transcriptome database of S. album for candidate cytochrome P450 genes, we cloned and characterized cDNAs encoding a small family of ten cytochrome P450-dependent monooxygenases annotated as SaCYP76F37v1, SaCYP76F37v2, SaCYP76F38v1, SaCYP76F38v2, SaCYP76F39v1, SaCYP76F39v2, SaCYP76F40, SaCYP76F41, SaCYP76F42 and SaCYP76F43. Nine of these genes were functionally characterized using in vitro assays and yeast in vivo assays to encode santalene/bergamotene oxidases and bergamotene oxidases. These results provide a foundation for production of sandalwood oil for the fragrance industry by means of metabolic engineering, as demonstrated with proof-of-concept formation of santalols and bergamotol in engineered yeast cells, simultaneously addressing conservation challenges by reducing pressure on supply of sandalwood from native forests.
5.2. Introduction

Sandalwood is the general name for woody perennials of the *Santalum* genus (Santalaceae), which are exploited for their fragrant heartwood. Sandalwoods are slow growing hemi-parasitic trees distributed throughout the tropical and temperate regions of India, Indonesia, Australia and the Pacific Islands [1,2]. The oil extracted from the stems and roots are highly sought after by the fragrance and perfume industry. *Santalum album*, also known as tropical or Indian sandalwood, is the most valuable of the commercially used species due to the high heartwood oil content (6-10% by dry weight) and desirable odour characteristics. Approximately 90% of *S. album* essential oil is composed of the sesquiterpene alcohols α-, β-, and *epi*-β-santalol and α-*exo*-bergamotol (Figure 5.1). The α- and β-santalols are the most important contributors to sandalwood oil fragrance [3-5]. Lanceol and α-bisabolol are also found in modest concentrations [6]. While the demand for sandalwood oil is increasing, disease, grazing animals and unsustainable exploitation of sandalwood trees has led to the demise of many natural populations. Plantations provide a more sustainable alternative to wild harvesting; however, slow growth rates, high potential for disease and substantial variation in oil yield hamper productivity. Alternatively, chemical approaches to synthesize the santalols have been described [7-9], but multiple low-recovery steps make chemical synthesis unfeasible at an industrial scale.

Investigations into alternative, more sustainable strategies to produce sandalwood oil include improved plantation systems through development of predictive marker systems for oil biosynthesis in developing heartwood of the slow growing trees, and metabolic
engineering of heterologous production systems. Key to these approaches is the elucidation of the biosynthesis of the santalols, bergamotols, and other sesquiterpene compounds characteristic of sandalwood oil. The first step in santalol and bergamotol biosynthesis is the generation of farnesyl diphosphate (FPP) from dimethylallyl diphosphate and isoprenyl diphosphate, catalyzed by FPP synthase (FPPS). FPP is cyclized by santalene synthase (SaSSy), a previously characterized sesquiterpene synthase [10], which produces a mixture of santalenes (α-, β- and epi-β-santalene) and α-exo-bergamotene. Since SaSSy generated four structurally similar products, it seemed plausible that a single, multi-substrate cytochrome P450 dependent monooxygenase (P450) could oxidize α-, β-, epi-β-santalene and bergamotene to produce α-, β-, epi-β-santalols and bergamotol, respectively (Figure 5.1). Alternatively, different cytochromes P450 could be involved in the oxidation of the different santalenes and bergamotene.
Figure 5.1 - Schematic biosynthetic pathway for santalols and bergamotol in sandalwood. Compounds identified with numbers are: α-santalene (1), α-exo-bergamotene (2), epi-β-santalene (3), β-santalene (4), (Z)-α-santalol (5), (E)-α-santalol (7), (Z)-α-exo-bergamotol (6), (E)-α-exo-bergamotol (8), (Z)-epi-β-santalol (9), (E)-epi-β-santalol (11), (Z)-β-santalol (10), (E)-β-santalol (12). Numbers match the numbers in Table 5.1. DMADP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; FPP, farnesyl diphosphate; FPPS, farnesyl diphosphate synthase; SaSSy, S. album santalene synthase

Here, we describe the discovery, cloning and functional characterization of a family of ten S. album P450s of the new CYP76F subfamily and an NADPH-dependent cytochrome P450 reductase (CPR) involved in santalol/bergamotol biosynthesis.
5.3. Experimental

5.3.1. Materials

The Saccharomyces cerevisiae yeast strain used in this study was BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). Escherichia coli α-Select Chemically Competent Cells (Bioline) were used for routine cloning and plasmid propagation. The sesquiterpene olefins α-, β- and epi-β-santalene, and α-exo-bergamotene are not commercially available, but can be produced by expression of SaSSy in yeast [10]. A sesquiterpene oil containing α-, β- and epi-β-santalene, and α-exo-bergamotene was produced in an industrial scale fermentation system by Allylix, Inc. (Kentucky, USA). The mixture was separated using silver nitrate impregnated TLC plates according to Daramwar et al. [25]; fractions were scraped from TLC plates and sequiterpenes eluted with pentane followed by GCMS analysis for purity. Other sesquiterpenes, specifically bisabolol, trans-β-farnesene and trans-nerolidol were purchased from SIGMA. Zingiberine, α-curcumene, β-bisabolene and β-sesquiphellandrene were from our in house collection of sesquiterpene standards isolated from natural sources.

5.3.2. Transcriptome Sequences

A cDNA library made from Santalum album xylem was sequenced with Sanger technologies generating 11,520 paired end sequences [10]. 454 Titanium sequencing of the cDNA library generated an additional 902,111 sequence reads. The transcriptome assembly was done using both the 454 and Sanger sequences with Roche Newbler assembler ver 2.6 under default parameters, which generated a total of 31,461 isotigs.
5.3.3. Cloning of P450 and CPR FLcRNAs and Yeast Transformation

FLcDNAs were amplified by PCR using Phusion Hot Start II DNA Polymerase (Thermo Scientific) with gene specific primers (Table 5.S2) and cDNA prepared from S. album wood cores and leaves as template. PCR conditions included initial denaturing at 98°C for 3 min, two cycles at 98°C for 10 sec, Tm-2°C for 20 sec, and 72°C for 30 sec, followed by 30 cycles at 98°C for 10 sec, Tm for 20 sec and 72°C for 30 sec, and termination for 7 min at 72°C. PCR products were gel purified and cloned into the pJET1.2 vector (Fermentas). Constructs designated pJET1.2-SaCYP76F37 through pJET1.2-SaCYP76F43, pJET1.2-SaCPR1 and pJET1.2-SaCPR2 were sequence verified. SaCYP76F FLcDNAs were subcloned into yeast expression vector pYEDP60 following the User Cloning method [26]. SaSSY (HQ343276) and SaFPPS (HQ343283) cDNAs [10] were cloned, respectively, into the NotI-Bgl II and BamHI-XhoI sites of the dual expression vector pESC-LEU2d by In-Fusion Cloning (Clontech). SaCPR1 and SaCPR2 were cloned individually into the EcoRI-NotI sites of the dual expression vector pESC-HIS (Stratagene). Plasmid transformation of yeast strain BY4741 was done using the LiCl method Gietz et al. [27]. Transformed yeast strains were selected on plates with appropriate synthetic complete drop-out selection medium and grown at 30°C for 48 h.

5.3.4. Microsome Preparation

For microsome isolation, BY4741 cells were transformed with plasmids harbouring P450 or CPR. Microsome membranes were prepared from 250 ml cultures according to Pompom et al. [28]. In brief, a 5 ml overnight culture was used to inoculate 50 ml of SD-
selective media starting at an OD$_{600}$ of 0.2 and grown at 30°C, 170 rpm for 24 h. A volume of 200 ml YPDE medium (1% yeast extract, 2% bacto-peptone, 5% ethanol, 2% dextrose) was inoculated with the 50 ml culture and incubated for another 24 h at 30°C, 170 rpm. Cells were collected by centrifugation for 10 min at 1,000 x g and induced with 2% galactose in 250 ml YP medium at 30°C, 170 rpm for 12-16 h. Yeast cells were pelleted by centrifugation at 2,000 x g for 10 min, washed once with 5 ml TEK (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM KCl) and suspended in TES2 buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 600 mM sorbitol, 5 mM DTT and 0.25 mM PMSF). All subsequent steps were performed at 4°C. Yeast cell were disrupted mechanically using acid-washed glass beads (425-600 μm, Sigma) and vigorous manual shaking for 3 x 30 sec. The cell homogenate was centrifuged at 10,000 x g for 15 min followed by ultracentrifugation of the supernatant at 100,000 x g for 1 h. Microsomes were suspended and homogenized in a buffer containing 50 mM Tris-HCl buffer pH 7.5, 1 mM EDTA and 30% (v/v) glycerol, and used directly for enzyme assays or stored at -80°C.

5.3.5. CPR Activity and P450 CO Spectra

Activity of recombinant SaCPRs was assayed using the Cytochrome C Reductase (NADPH) assay kit (Sigma). CO difference spectra of recombinant P450s were measured according to Guengerich et al. [29].
5.3.6. *In Vitro P450 Assays*

Microsome preparations containing candidate P450 and CPR were assayed for their capacity to oxidize sesquiterpenes. The reaction mixtures contained 50 mM potassium phosphate pH 7.5, 0.8 mM NADPH and 40 μM of substrate in a total volume of 400 μl. Enzyme reactions were initiated by adding 50 μl of the microsome preparation, incubated at 30°C for 2 h with shaking and stopped by adding 500 μl of hexane. The organic layer was transferred to a new GC vial and concentrated under N₂ gas to about 100 μl followed by GCMS analysis. For kinetic analysis, enzyme assays were performed as above with the following modifications: Assays were performed in a total volume of 400 μl with either 17 pmol of *Sa* CYP7639v1 protein or 35 pmol of *Sa* CYP7637v1 protein, and substrate concentrations of 12 to 138 μM of α-santalene or β-santalene; assays were incubated for 20 min.

5.3.7. *Yeast Metabolic Engineering*

To assess the production of santalols/bergamotol in a yeast system, the yeast strain BY4741 was co-transformed with plasmids containing cDNAs for *Sa* FPPS, *Sa* SSY, *Sa* CPR, and a candidate CYP76F. Recombinant yeast was initially grown overnight at 30°C in 5 ml of 2% dextrose in minimal selective media. The next day, a 50 ml culture was initiated at a starting OD₆₀₀ of 0.2 and grown at 30°C with shaking at 170 rpm until the culture reached an OD₆₀₀ of 0.6-0.8. Expression was initiated by transfer into minimal selective media with 2% galactose and grown for 14-16 h. Yeast cells were harvested by centrifugation at 1,000 x g for 10 min and washed once with 5 ml sterile ddH₂O. Cells were extracted twice by vortexing for 1 min with 2 ml hexane and 250
μl acid-washed glass beads (425-600 μm, Sigma). Pooled extracts were transferred to a clean test-tube containing anhydrous Na₂SO₄ and evaporated under a gentle stream of N₂ gas to about 200 μl. The samples were transferred to a GC glass vial for GCMS analysis or stored at -80°C.

5.3.8. **GC-MS Analysis**

GCMS analysis was carried out on an Agilent 7890A/5975C GC-MS system operating in electron ionization selected ion monitoring (SIM)-scan mode. Samples were analyzed on both an HP5 (non-polar; 30 m x 0.25 mm ID x 0.25μm thickness) and a DB Wax fused silica column (polar; 30 m x 0.25 mm ID x 0.25μm thickness). In both cases, the injector was operated in pulsed splitless mode with the injector temperature maintained at 250°C. Helium was used as the carrier gas with a flow rate of 0.8 ml min⁻¹ and pulsed pressure set at 25 psi for 0.5 min. Scan range: m/z 40-500; SIM: m/z 93, 94, 105, 107, 119, 122 and 202 [dwell time 50 msec]. The oven program for the HP5 column was: 40°C for 3 min; ramp of 10°C min⁻¹ to 130°C, 2°C min⁻¹ to 180°C, 50°C min⁻¹ to 300°C; 300°C for 10 min. The oven program for the DB-wax column was: 40°C for 3 min; ramp of 10°C min⁻¹ to 130°C, 2°C min⁻¹ to 200°C, 50°C min⁻¹ to 250°C; 250°C for 15 min. Chemstation software was used for data acquisition and processing. Compounds were identified by comparison of mass spectral with authentic standards and the NIST/EPA/NIH mass spectral library v2.0 and by comparison of retention indices with those appearing in other publications [15,16].
5.3.9. *Phylogenetic Analysis*

Phylogenetic analysis was performed using the software MEGA version 4 [30] employing the neighbor-joining (NJ) algorithm with default parameters. Bootstrap (500 replications) confidence values over 50% are displayed at branch points.

5.3.10. *Accession Numbers*

The cDNA sequences described in this paper have been submitted to GenBankTM/EBI with accession numbers: *SaCYP76F37v1* (KC533717); *SaCYP76F37v2* (KC698966); *SaCYP76F38v1* (KC533715); *SaCYP76F38v2* (KC533718); *SaCYP76F39v1* (KC533716); *SaCYP76F39v2* (KC698967); *SaCYP76F40* (KC698968); *SaCYP76F41* (KC698969); *SaCYP76F42* (KC698965); *SaCYP76F43* (KC533719); *SaCPR1* (KC842187); *SaCPR2* (KC842188).
5.4. Results

5.4.1. Gene discovery and Full-Length (FL) cDNA cloning

A *S. album* transcriptome assembly of 31,461 isotigs was blastx searched for candidate CPRs and P450s potentially involved in the hydroxylation of santalenes and bergamotene. Two *Sa*CPRs were identified using *Arabidopsis thaliana* CPRs (CAB58575.1, CAB58576.1) as search sequences. FLcDNAs *Sa*CPR1 and *Sa*CPR2 were 70% identical and 82% similar at the amino acid level. Searches for P450s were performed with a set of known plant P450s of the CYP71, CYP72 and CYP76 families, which include P450s with known functions in terpenoid biosynthesis [11-13].

Transcripts of the CYP76 family were among the most abundant P450s in the *S. album* transcriptome and assembled into two different isogroups and two individual isotigs (Table 5.S1). Isogroup 1 consisted of 2,143 reads including 1,107 unique reads assembled into three isotigs. It generated a consensus sequence of 1,917 base pairs and an open reading frame (ORF) of 1,530 bp. Isogroup 2 consisted of 228 reads including 140 unique reads assembled into two isotigs. Both isotigs share a consensus ORF of 1,530 bp. A separate isotig consisted of 11 reads generating a partial sequence of 1,200 bp. Another separate isotig contained one partial sequence of 277 bp with several stop codons. Isogroups 1 and 2 were selected for FLcDNA cloning. PCR amplification with primers designed according to isogroup 1 resulted in a single unique FLcDNA clone designated as *Sa*CYP76F38v1. PCR amplification with primers based on isogroup 2 resulted in nine different cDNAs clones designated as *Sa*CYP76F37v1, *Sa*CYP76F37v2, *Sa*CYP76F38v1, *Sa*CYP76F38v2, *Sa*CYP76F39v1, *Sa*CYP76F39v2, *Sa*CYP76F40,
SaCYP76F41, SaCYP76F42, and SaCYP76F43. The predicted CYP76F proteins were 94-99% identical to each other and contained motifs characteristic of eukaryotic P450s including a proline-rich region near the N-terminal membrane anchoring domain, the oxygen-binding domain and the highly conserved heme binding motif (Figure 5.S1). A blastp search of the deduced amino acid sequences against the NCBI GenBank protein database identified best matches to a putative P450 from Vitis vinifera (XP_002281735) with 62-64% identity, and CYP76B6 geraniol hydroxylase (CAC80883) from Catharanthus roseus [14] with 53-54% identity. A phylogeny with related plant P450s (Figure 5.2) showed the S. album CYP76F proteins form two separate clades, I and II, and are closest to the CYP76B cluster of other species.
Table 5.1 - Retention indices of sesquiterpenes and sesquiterpenols identified in the enzyme assays with cytochromes P450 of the *S. album* CYP76F subfamily and of sesquiterpene alcohols of *S. album* oil

<table>
<thead>
<tr>
<th>No</th>
<th>Compound</th>
<th>LRI&lt;sup&gt;2&lt;/sup&gt; DBwax</th>
<th>LRI&lt;sup&gt;3&lt;/sup&gt; HP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>α-santalene</td>
<td>1579</td>
<td>1423</td>
</tr>
<tr>
<td>2</td>
<td>α-exo-bergamotene</td>
<td>1592</td>
<td>1437</td>
</tr>
<tr>
<td>3</td>
<td>epi-β-santalene</td>
<td>1637</td>
<td>1450</td>
</tr>
<tr>
<td>4</td>
<td>β-santalene</td>
<td>1652</td>
<td>1463</td>
</tr>
<tr>
<td>5</td>
<td>(Z)-α-santalol</td>
<td>2343</td>
<td>1676</td>
</tr>
<tr>
<td>6</td>
<td>(Z)-α-exo-bergamotol</td>
<td>2353</td>
<td>1692</td>
</tr>
<tr>
<td>7</td>
<td>(E)-α-santalol</td>
<td>2382</td>
<td>1697</td>
</tr>
<tr>
<td>8</td>
<td>(E)-α-exo-bergamotol</td>
<td>2389</td>
<td>1711</td>
</tr>
<tr>
<td>9</td>
<td>(Z)-epi-β-santalol</td>
<td>2409</td>
<td>1703</td>
</tr>
<tr>
<td>10</td>
<td>(Z)-β-santalol</td>
<td>2423</td>
<td>1717</td>
</tr>
<tr>
<td>11</td>
<td>(E)-epi-β-santalol (tentative)</td>
<td>2452</td>
<td>1726</td>
</tr>
<tr>
<td>12</td>
<td>(E)-β-santalol</td>
<td>2465</td>
<td>1738</td>
</tr>
</tbody>
</table>

<sup>1</sup>These numbers are used as identifiers for compounds and corresponding peaks in.
<sup>2</sup>GC traces throughout the paper and figures.
<sup>3</sup>Linear retention indices (LRI) measured on a DBwax column.
<sup>3</sup>Linear retention indices (LRI) measured on an HP5 column.
doi:10.1371/Journal.pone.0075053.t001
Figure 5.2 - Phylogenetic tree of *S. album* CYP76F proteins and related terpene-modifying P450s. The neighbor-joining tree was constructed with members of the CYP71 clan, using *Picea sitchensis* PsCYP720B4 (ADR78276) as an outgroup. *S. album* CYP76F proteins fell into two clades, clade I santalene/bergamotene oxidases and clade II bergamotene oxidases. CaCYP76B4, *Camptotheca acuminata* putative geraniol-10-hydroxylase (AES93118); CrCYP76B6, *Catharanthus roseus* geraniol-10-hydroxylase (Q8VWZ7); SmCYP76B4, *Swertia mussotii* geraniol-10-hydroxylase (D1MI46); OsCYP76M7 *Oryza sativa* ent-cassadiene C11-α-hydroxylase (NP_001047185); MpCYP71A32, *Mentha x piperita* menthofuran synthase (Q947B7); PaCYP71A1, *Persea americana* (P24465); CiCYP71AV8, *Chlorodium intybus* valencene oxidase (ADM86719); MpCYP71D13, *Mentha x piperita* (-)-limonene-3-hydroxylase (AY281027); NtCYP71D20, *Nicotiana tabacum*, 5-epi-aristolochene-1,3-dihydroxylase (AF368376); GaCYP706B1, *Gossypium arboreum* (+)-delta-cadinene-8-hydroxylase (AAK60517). This work: SaCYP76F37v1 (KC533717); SaCYP76F37v2 (KC698966); SaCYP76F38v1 (KC533715); SaCYP76F38v2 (KC533718); SaCYP76F39v1 (KC533716); SaCYP76F39v2 (KC698967); SaCYP76F40 (KC698968); SaCYP76F41 (KC698969); SaCYP76F42 (KC698965); SaCYP76F43 (KC533719)
5.4.2. Expression of recombinant SaCYP76Fs in yeast

SaCYP76F FLcDNAs were expressed together with SaCPR2 in yeast cells, and microsomes were isolated for *in vitro* P450 enzyme assays. Microsome preparations for all ten SaCYP76Fs, except SaCYP76F43, displayed characteristic P450 CO difference spectra. The P450 content of the microsomal preparations ranged from 0.2 to 1.6 μM (Figure 5.S2).

5.4.3. *In vitro* functional identification of clade I SaCYP76Fs using a blend of sesquiterpenes

Microsome preparations were screened for sesquiterpene oxidase activity using NADPH and a defined sesquiterpene mixture of α-, β- and *epi*-β-santalene and α-*exo*-bergamotene as substrate. These sesquiterpenes are not commercially available and were produced by expression of SaSSy in yeast (Figure 5.S3). Product formation was measured by gas chromatography mass spectrometry (GCMS).

Focusing initially on the clade I SaCYP76F39v1, we found that microsomes containing this P450 catalyzed the hydroxylation of the three santalenes and α-*exo*-bergamotene, leading to eight different compounds identified as (Z)- and (E)-α santalol (5 and 7), (Z)- and (E)-α-*exo*-bergamotol (6 and 8), (Z)- and (E)-*epi*-β-santalol (9 and 11) and (Z)- and (E)-β-santalol (10 and 12) (Figure 5.3A and Table 5.1; peak numbers in Figures match the numbers in Table 5.1). Products were identified based on matches of their mass spectra (Figure 5.S4) with entries in the NIST and Wiley libraries and by
matches of retention indices obtained on two different GC matrices (Table 5.1) with retention indices reported in the literature [15,16]. In addition, comparison of the product profile with the profile of an authentic sandalwood oil sample (Figure 5.3B) showed identical retention times and nearly identical mass spectra (Figure 5.S4) for all eight compounds that were present both in the product of the enzyme assay and in the oil, albeit in different proportions. No product formation was detected in the absence of NADPH or with microsomes from yeast carrying an empty vector (Figure 5.3C).

The sesquiterpenol profile produced in vitro by microsomes containing SaCYP76F39v1 matched the overall sesquiterpenol composition of S. album oil; however, the relative amounts of individual stereoisomers differed (Figure 5.3). CYP76F39v1 produced (E)-\(\alpha\)-santalol (7) and (Z)-\(\alpha\)-santalol (5) in a ratio of approximately 5:1, and (E)-\(\beta\)-santalol (12) and (Z)-\(\beta\)-santalol (10) in a ratio of approximately 4:1, while (Z)-\(\alpha\)-santalol (5) and (Z)-\(\beta\)-santalol (10) are the more dominant isomers in the oil.
Figure 5.3 - GCMS analysis of products formed in vitro with *Sa* CYP76F39v1. A sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Supplementary Figure 5.S3) was incubated with microsomes containing *Sa* CYP76F39v1 and *Sa* CPR produced in yeast. (A) Product profile (extracted ion chromatogram, EIC) of assays with *Sa* CYP76F39v1. (B) Authentic *S. album* oil. (C) Control assays were performed with microsomes isolated from yeast cells transformed with the empty vector. Mass spectra of compounds corresponding to peaks 5-12 identified in assays with *Sa* CYP76F39v1 (left panel) and *S. album* oil (right panel) are shown in Figure 5.S4. Peak numbers match the numbers in Table 5.1 and Figure 5.1

Other clade I P450s, specifically *Sa* CYP76F39v2, *Sa* CYP76F40, *Sa* CYP76F41, and *Sa* CYP76F42, gave product profiles similar to that observed with CYP76F39v1 (Figure 5.4 A-D). The major products formed by microsomes containing *Sa* CYP76F40 or *Sa* CYP76F42 were (*E*)-α-exo-bergamotol (8) and (*E*)-β-santalol (12).
Figure 5.4 - GCMS analysis of products formed in vitro with clade I SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of products formed in vitro with (A) SaCYP76F39v2; (B) SaCYP76F40; (C) SaCYP76F41; (D) SaCYP76F42. Assays were performed with a sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Figure 5.S3) as substrate and microsomes prepared from yeast transformed with SaCPR, individual clade I candidate SaCYP76F cDNAs, or (E) empty vector as control. Peak numbers match the numbers in Table 5.1 and Figure 5.1
5.4.4. *In vitro functional identification of clade II SaCYP76Fs using a blend of sesquiterpenes*

In contrast to the clade I *Sa*CYP76Fs, which each gave the same eight sesquiterpenol products, microsomes containing clade II members *Sa*CYP76F37v1, *Sa*CYP76F37v2, *Sa*CYP76F38v1, and *Sa*CYP76F38v2 gave only three products identified as (E)-α-exo-bergamotol (8) as the major product and (E)-α-santalol (7) and (E)-β-santalol (12) as minor products (Figure 5.5A-D). No activity was found with *Sa*CYP76F43 (Figure 5.5E) possibly due to low expression in yeast as evidenced by the corresponding CO difference spectrum (Figure 5.S2).
Figure 5.5 - GCMS analysis of products formed in vitro with clade II SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of products formed in vitro with (A) SaCYP76F38v1; (B) SaCYP76F38v2; (C) SaCYP76F37v1; (D) SaCYP76F37v2. Assays were performed with a sesquiterpene mixture of α-, β- and epi-β-santalene and α-exo-bergamotene (Figure 5.53) as substrate and microsomes prepared from yeast transformed with SaCPR, individual clade II candidate SaCYP76F cDNAs, or (E) empty vector as control. Peak numbers match the numbers in Table 5.1 and Figure 5.1
5.4.5. Characterization of clade I and clade II SaCYP76Fs using individual sesquiterpenes

Although the authentic candidate substrates are not available in pure form, we could partially separate the sesquiterpenes of the mixture of α-, β- and *epi*-β-santalene and α *exo*-bergamotene (Figure 5.3S). Three different fractions containing mainly α-santalene (1) (Figure S5.3B), α-*exo*-bergamotene (2) (Figure S5.3C), or *epi*-β-santalene (3) and β santalene (4) (Figure S5.3D) were used as individual substrates in assays with microsomes containing *Sa* CYP76F39v1, representing clade I, or *Sa* CYP76F37v1, representing clade II. *Sa* CYP76F39v1 with α-santalene produced both (Z) - and (E)-α-santalol (5 and 7; Figure S5.5A), while only (E)-α-santalol (7) formation was detected with *Sa* CYP76F37v1 (Figure S5.5D). With α-*exo*-bergamotene, *Sa* CYP76F39v1 produced (Z)- and (E)-α-*exo*-bergamotol (6 and 8; Figure S5.5B), while only (E)-α-*exo*-bergamotol (8) formation was detected with *Sa* CYP76F37v1 (Figure S5.5E). *Sa* CYP76F39v1 gave four products, (Z)- and (E)-*epi*-β-santalol (9 and 11) and (Z)- and (E)-β-santalol (10 and 12), in assays with *epi*-β-santalene and β-santalene (Figure S5.5C), whereas only (E)-β-santalol (12) was detected in assays with *Sa* CYP76F37v1 (Figure S5.5F). These results confirmed the activities seen with microsome in vitro assays with the mixture of santalenes and bergamotene.

5.4.6. Substrate specificity and kinetic properties of SaCYP76Fs

To test the range of substrates potentially converted by the clade I and clade II *Sa* CYP76F enzymes, we assayed *Sa* CYP76F37v1 and *Sa* CYP76F39v1 with a set of
sesquiterpenes which resemble santalenes in the acyclic isoprenyl side chain (Table 5.2). Of the nine different substrates tested, SaCYP76F39v1 efficiently converted only the two santalenes, while it showed low activity with α-bisabolol and was not active with α-curcumene, zingiberene, β-bisabolene, β-sesquiphellandreene, farnesene, and trans-nerolidol. These results demonstrated a narrow substrate selectivity of SaCYP76F39v1 with sesquiterpenes relevant for sandalwood oil biosynthesis. Similarly, SaCYP76F37v1 was selectively active with the two santalenes and trans-nerolidol.
Table 5.2 - Relative activities of *SaCYP76F39v1* and *SaCYP76F37v1* with different sesquiterpenes

<table>
<thead>
<tr>
<th>Substrate</th>
<th>CYP76F39v1 [%]</th>
<th>CYP76F37v1 [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-santalene</td>
<td>99.8</td>
<td>17.3</td>
</tr>
<tr>
<td>β-santalene</td>
<td>100</td>
<td>17.7</td>
</tr>
<tr>
<td>α-curcumene</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>α-zingiberene</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>β-bisabolene</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>β-sesquiphellandrene</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>α-bisabolol</td>
<td>9.4</td>
<td>0</td>
</tr>
<tr>
<td>trans-β-farnesene</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>trans-nerolidol</td>
<td>0</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Apparent Kₘ values of *SaCYP76F39v1* and *SaCYP76F37v1* were, respectively, 25.92 (±0.11) μM and 133 (±0.41) μM with α-santalene, and respectively, 34.82 (±0.41) μM and 157 (±0.17) μM with β-santalene. The kₗ values obtained for *SaCYP76F39v1* were 1.12 s⁻¹ with α-santalene and 1.17 s⁻¹ with β-santalene. The kₗ/Kₘ values for *SaCYP76F39v1* were 4.3 x 10⁴ s⁻¹ M⁻¹ with α-santalene and 3.3 x 10⁴ s⁻¹ M⁻¹ with β-santalene. The kₗ values obtained for *SaCYP76F37v1* were 0.2 s⁻¹ with α-santalene and 0.13 s⁻¹ with β-santalene. The kₗ/Kₘ values for *SaCYP76F37v1* were 1.5 x 10³ s⁻¹ M⁻¹ with α-santalene and 8.1 x 10² s⁻¹ M⁻¹ with β-santalene.
5.4.7.  **Formation of santalols and bergamotol in transformed yeast cells**

To test the potential for using *SaCYP76F* cDNAs to produce santalols and bergamotol *in vivo*, we first expressed the previously characterized *SaSSy* and *SaFPPS* cDNAs [10] in yeast to form the known *SaSSy* products α-santalene (1), α-*exo*-bergamotene (2), *epi*-β-santalene (3) and β-santalene (4). These four sesquiterpenes were detected in transformed yeast cells (Figure 5.S6), but were not released with detectable amounts into the culture medium. No differences were observed between cells expressing *SaSSy* with or without the additional *SaFPPS* suggesting that endogenous yeast FPP is accessible for *SaSSy* to produce santalenes and bergamotene. We then tested product formation with the additional expression of *SaCPR2* and *SaCYP76F* candidate cDNAs. GCMS analysis of yeast cells expressing *SaCYP76F39v1, SaCPR2* and *SaSSy* showed a product profile of eight sesquiterpenols identified as (Z)- and (E)-α-santalol (5 and 7), (Z)- and (E)-α-*exo*-bergamotol (6 and 8), (Z)- and (E)-*epi*-β-santalene (9 and 11) and (Z)- and (E)-β-santalene (10 and 12) (Figure 5.6A), similar to the product profile of the *in vitro* assays (Figure 5.3A). The product peak for (Z)-α- *exo*-bergamotol (6) overlapped with a peak corresponding to (E,E)-farnesol, which was produced in yeast independent of the *SaCYP76F39v1* (Figure 5.6B).

 Apparently, a fraction of the sesquiterpenol produced by recombinant yeast expressing *SaSSy, SaCPR2* and *SaCYP76F39v1* were modified to unknown compounds (identified with hash marks in Figure 5.6A). When untransformed yeast cells were incubated with authentic sandalwood oil, we found the same unknown compounds (Figure 5.S7), implying that these compounds are not direct products of *SaCYP76F39v1*, but are produced by an endogenous activity of yeast converting sandalwood sesquiterpenols.
In vivo analysis of the other SaCYP76F clade I members gave product profiles with nearly identical ratios (Figure 5.S8) as observed with the corresponding *in vitro* assays with the microsomal preparations (Figure 5.4). Yeast cells expressing clade II SaCYP76Fs produced mostly (E)-α-exo-bergamotol (8) similar to the products formed in the *in vitro* assays, but only traces of santalols (7 and 12) (Figure 5.S9). Again, no activity was found with CYP76F43.
Figure 5.6 - GCMS analysis of products formed in vivo with SaCYP76F39v1. GCMS analysis (extracted ion chromatograms) of compounds formed in vivo in yeast cells expressing SaSSY, SaCPR2 and (A) SaCYP76F39v1 or (B) an empty vector. (C) Mass spectra of compounds corresponding to peaks 5-12 identified in (A). Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks in (A) and (B) marked with symbol (*) correspond to farnesol also produced in yeast cells without SaCYP76F. Peaks in (A) marked with symbol (†) represent yeast in vivo modifications of santalols (see Figure 5.56)
5.4.8. **Effect of CPR1 and CPR2**

To test if substituting SaCPR1 and SaCPR2, which are 70% identical at the protein level, could affect changes in product profiles, we tested both CPRs in yeast *in vivo* experiments with representative class I and class II *Sa*CYP76F, CYP76F39v1 and CYP76F38v1. No differences were observed in the products and their relative abundances.

5.5. **Discussion**

Using transcriptome analysis, cloning and functional characterization of recombinant P450s, we identified a new CYP76F subfamily in *S. album* involved in the biosynthesis of α-, β- and *epi*-β-santalols and bergamotols. The different *Sa*CYP76Fs catalyze hydroxylations of santalenes and/or bergamotene products of *Sa*SSy at the terminal allylic methyl groups. Clade I *Sa*CYP76F enzymes produced both (Z) and (E) stereoisomers of α-, β- and *epi*-β-santalols and bergamotols. The P450 product ratios of (Z) and (E) stereoisomers of α- and β-santalol were approximately 1:5 and 1:4, respectively, while the oil harvested from the mature heartwood of *S. album* trees contained mainly the (Z) alcohols [17,18]. There are several possible explanations for the difference in the ratio of stereoisomers found in the enzyme product profile and in the oil extracted from trees. Importantly, we excluded the possibility that the activity of *Sa*CYP76Fs was non-specific towards a range of different substrates, since only products of *Sa*SSy were preferred substrates when compared with other similar sesquiterpenes. However, it is important to note that conditions of yeast cells and
in vitro assays are different compared to the physiological conditions in planta, which might explain the differences of product stereoisomers observed. It is possible that subtle changes in the shape and size of the active site under different conditions might result in the olefin precursors being oxidized in different configurations. It is also important to note that the products detected in in vitro microsome assays and in yeast in vivo assays were formed and accumulated over a period of minutes to hours. In contrast, the oil extracted from mature heartwood is the product of biosynthesis and accumulation that occurs over a much longer time period of many years. Isomerization, perhaps catalyzed by an isomerase, may be possible in the trees, however may not have been mimicked with the conditions of the in vitro or yeast in vivo enzyme assays used here. Although the ten P450s isolated in this work are the most abundant P450s in the sandalwood transcriptome sequences, it is also possible that additional sandalwood P450s exist that are similarly active on the santalenes and bergamotene substrates, but generating predominantly the (Z) stereoisomer. We will be exploring this possibility with further screening of the S. album P450 family.

The CYP76 gene family is part of the CYP71 clan, which includes P450 families involved in plant primary and secondary metabolism. Previously functionally characterized CYP76 members are involved in xenobiotic detoxification [19], oxidation of iridoid monoterpenoids [14,20], and oxidation of diterpenes [21,22]. The CYP76F members described here for sesquiterpene hydroxylation add a new dimension to the known functional space of the CYP76 family. The number of CYP76 genes is highly variable in different plant species. For example, papaya (Carica papaya) contains three CYP76 genes, A. thaliana has nine CYP76 genes, and grapevine (Vitis vinifera) has 24 CYP76 genes [11,23]. The ten S. album CYP76F members described
here were identified based on transcriptome sequencing and may not represent the full complement of CYP76 genes of this species. In the absence of a genome sequence of *S. album*, it is not clear if any of these genes represent pairs of allelic variants. The *S. album* CYP76F members separate into two clades, clade I and II. Although there is overlap in their product profiles, clade I members formed preferentially santalols, whereas clade II members produced preferentially (E)-\(\alpha\)-exo-bergamotol.

The CYP76 and CPR cDNAs described here, combined with previously cloned santalene synthases [10], provide a biotechnology opportunity to produce valuable components of sandalwood oil. Our initial results demonstrate the potential of transformed yeast cells for production of santalols and bergamotols. As a proof-of-concept, we reconstructed the pathways for biosynthesis of santalols and bergamotols in yeast cells using the multi-product *SaSSy* and *SaCPR* in combination with different multi-substrate *SaCYP76Fs*. These results provide a foundation for further metabolic engineering to improve yields and target product specificities.

The cloned terpene synthases [10,24] and P450s (this study) of sandalwood oil biosynthesis can also be explored as biomarkers to monitor the onset of oil formation in sandalwood plantations or for the development of genetic markers for tree improvement. In this context, it is important to note that very little is known about the cell types and the molecular events that control spatial and temporal patterns of the onset of biosynthesis of sandalwood oil. In fact, the spatial and temporal patterns of the onset of sandalwood oil biosynthesis are not well known, beyond the association of oil accumulation in the aging heartwood of sandalwood stems and roots. The aging heartwood of sandalwood trees provides an extremely difficult system to study with
biochemical tools. Thus, the genes described here and in previous work [10] and their possible applications for metabolic engineering of sandalwood oil biosynthesis and the development of molecular markers are likely to become more important as worldwide demand for sandalwood products increase and as natural resources of *S. album* continue to decline.

5.6. Acknowledgements

We thank Richard Burlingame (Allylix Inc., Kentucky) for the sequiterpene oil; Andreas Gesell, Macaire M. Yuen, and Philipp Zerbe (UBC) for helpful discussion and assistance; David Nelson (University Tennessee) for P450 naming.
5.7. Supplementary Materials

Table 5.S1 - Primers designed for amplification of cDNAs from *S. album*

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′ → 3′)</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isogroup 1 Forward</td>
<td>ATGGACTCTTAAAGTTTTATCCTGTTTG</td>
<td>55</td>
</tr>
<tr>
<td>Isogroup 1 Reverse</td>
<td>TTACCCCCGGATCGGGACAG</td>
<td>56</td>
</tr>
<tr>
<td>Isogroup 2 Forward</td>
<td>ATGGACTTCTTAAAGTTTGATCCTC</td>
<td>52</td>
</tr>
<tr>
<td>Isogroup 2 Reverse</td>
<td>TTACCCCCGGATCGGGACAG</td>
<td>54</td>
</tr>
<tr>
<td>SaCPR1 Forward</td>
<td>ATGAGTTGAGCTCGAGCTGATG</td>
<td>57</td>
</tr>
<tr>
<td>SaCPR1 Reverse</td>
<td>TCACCACACATCCCCGTAAATACCTTC</td>
<td>57</td>
</tr>
<tr>
<td>SaCPR2 Forward</td>
<td>ATGCAATTGAGCTCGTCAAG</td>
<td>58</td>
</tr>
<tr>
<td>SaCPR2 Reverse</td>
<td>TCACCACACATCCCCGTAAATACCTTCC</td>
<td>58</td>
</tr>
</tbody>
</table>
Table 5.52 - Summary of transcriptome mining for CYP76 family members in the *S. album* Sanger and 454 sequence data

<table>
<thead>
<tr>
<th>Isogroup/isotig</th>
<th>isotig</th>
<th>Reads</th>
<th>E-value (CrCYP76B6)</th>
<th>Identity [%] (CrCYP76B6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>isotig05182</td>
<td>910</td>
<td>8.34E-142</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>isotig05183</td>
<td>763</td>
<td>2.68E-145</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>isotig05184</td>
<td>470</td>
<td>1.61E-78</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>isotig06871</td>
<td>110</td>
<td>1.23E-126</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>isotig06872</td>
<td>118</td>
<td>9.19E-156</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>isotig14788</td>
<td>11</td>
<td>1.53E-93</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>isotig29133</td>
<td>1</td>
<td>1.49E-52</td>
<td>60</td>
</tr>
</tbody>
</table>

Note: Dotted lines denote isotig members of the same isogroup
Figure S.1 - Amino acid sequence alignment of *S. album* CYP76F genes. SaCYP76F37v1, SaCYP76F37v2, SaCYP76F38v1, SaCYP76F38v2, SaCYP76F39v1, SaCYP76F39v2, SaCYP76F40, SaCYP76F41, SaCYP76F42 and SaCYP76F43. Red, dark grey and light grey shading denote 100% and 80% and 50% conserved residues, respectively. Horizontal arrows denote the proline region (a), O\(_2\) binding motif (b) and heme binding motif (c). Boxes indicate the putative substrate recognition sites (SRS) regions originally described by Gotoh [31]. Multiple sequence alignment was performed with the software CLUSTALW [32] and visualized with Gendoc v2.7.
Figure 5.S2 - Reduced CO-difference spectra of isolated microsomes containing *S. album* CYP76F proteins. CO difference spectra of microsomal fractions from *S. cerevisiae* harboring a cytochrome P450 or an empty vector are shown. Concentration of SaCYP76F proteins are given based on an extinction coefficient of 91,000 M$^{-1}$ cm$^{-1}$.
Figure 5.53 - GCMS analysis (extracted ion chromatogram) of a sesquiterpene mixture and fractions separated by TLC. The sesquiterpene mixture was produced with a recombinant yeast strain expressing $SaSSy$ (10) and was provided to us by Allylix Inc. It contained (A) $\alpha$-santalene (1), $\alpha$-exo-bergamotene (2), epi-$\beta$-santalene (3), $\beta$-santalene (4). The mixture was separated by TLC into three fractions containing mainly (B) $\alpha$-santalene (1); (C) $\alpha$-exo-bergamotene (2); or (D) $\beta$-santalene (4). Mass spectra of peaks 1 to 4 are provided. Peak numbers match the numbers in Table 5.1 and Figure 5.1.
Figure 5.S4 - Mass spectra of products formed \textit{in vitro} with \textit{Sa}CYP76F39v1. Mass spectra of compounds corresponding to peaks 5-12 shown in Figure 5.3 and identified in assays with CYP76F39v1 (left panel) and \textit{S. album} oil (right panel). Peak numbers match the numbers in Table 5.1, Figure 5.1, and Figure 5.3.
Figure 5.55 - GCMS analysis (extracted ion chromatogram) of products formed in vitro with SaCYP76F39v1 or SaCYP76F37v1 using partially purified substrates. Product profile in assays with SaCYP76F39v1 using (A) α-santalene, (B) α-exo-bergamotene, or (C) epi-β-santalene and β-santalene as substrate. Product profile in assays with SaCYP76F37v1 using (D) α-santalene, (E) α-exo-bergamotene, or (F) epi-β-santalene and β-santalene as substrate. (G) Products were identified by comparison to authentic standards. Peak numbers match the numbers in Table 5.1 and Figure 5.1
Figure 5.56 - GCMS analysis (extracted ion chromatogram) and mass spectra of sesquiterpenes produced in yeast expressing SaSSy. (A) GCMS analysis of sesquiterpenes extracted from pelleted yeast cells expressing SaSSy. (B) Mass spectra of peaks 1-4: α-santalene (1), α-exo-bergamotene (2), epi-β-santalene (3), and β-santalene (4). Compounds were identified by comparison to an authentic standard and retention indices. Peak numbers match the numbers in Table 5.1 and Figure 5.1.
Figure 5.57 - Modification of sandalwood oil compounds in yeast cell culture. GCMS analysis of sesquiterpenols of natural sandalwood oil sample before (A) and after (B) overnight incubation with yeast cells, which do not contain a SaCYP76F gene. Peaks in (B) marked with symbol (#) represent yeast in vivo modifications of santalols independent of SaCYP76F. Peak numbers match the numbers in Table 5.1 and Figure 5.1.
Figure 5.8 - GCMS analysis of products formed in vivo with clade I SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of compounds formed in vivo in yeast cells expressing SaSSy, SaCPR2, and (A) SaCYP76F39v2, (B) SaCYP76F40, (C) SaCYP76F41, or (D) SaCYP76F42. Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks marked with symbol (*) correspond to farnesol produced also in yeast cells without SaCYP76F. Peaks in marked with symbol (#) represent yeast in vivo modifications of santalols (see Figure 5.S7).
Figure 5.59 - GCMS analysis of products formed in vivo with clade II SaCYP76Fs. GCMS analysis (extracted ion chromatograms) of compounds formed in yeast cells expressing SaSSy, SaCPR2 and (A) SaCYP76F38v1, (B) SaCYP76F38v2, (C) SaCYP76F37v1, (D) SaCYP76F37v2, or (E) SaCYP76F43. Peak numbers match the numbers in Table 5.1 and Figure 5.1. Peaks marked with symbol (*) correspond to farnesol produced also in yeast cells without SaCYP76F. Peaks in marked with symbol (#) represent yeast in vivo modifications of santalols (see Figure 5.57)
5.8. References


CHAPTER SIX

The transcriptome of sequiterpenoid biosynthesis in heartwood xylem of Western Australian sandalwood 
(Santalum spicatum)

Jessie Moniodis\textsuperscript{a,b,c*}, Christopher G Jones\textsuperscript{a}, Julie A Plummer\textsuperscript{a}, E. Liz Barbour\textsuperscript{a}, Emilio Ghisalberti\textsuperscript{b} and Joerg Bohlmann\textsuperscript{c}

\textsuperscript{a}School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley WA 6009 Australia

\textsuperscript{b}School of Chemistry and Biochemistry (M310), University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

\textsuperscript{c}Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T1Z4, Canada

*Corresponding author
6.1. Graphical Abstract

The transcriptome of sesquiterpenoid biosynthesis in heartwood xylem of Western Australian sandalwood (*Santalum spicatum*)

Jessie Moniodis*, Christopher G. Jones, E. Liz Barbour, Julie A. Plummer, Emilio L. Ghisalberti and Joerg Bohlmann

*Santalum spicatum* heartwood xylem produces fragrant sesquiterpenoids. Sequences of sesquiterpenoid biosynthesis including terpene synthase and cytochrome P450 sequences were identified in a xylem transcriptome.
6.2. Abstract

The fragrant heartwood oil of West Australian sandalwood (*Santalum spicatum*) contains a mixture of sesquiterpene olefins and alcohols, including variable levels of the valuable sesquiterpene alcohols, α- and β-santalol, and often high levels of *E,E*-farnesol. Transcriptome analysis revealed sequences for a nearly complete set of genes of the sesquiterpenoid biosynthetic pathway in this commercially valuable sandalwood species. Transcriptome sequences were produced from heartwood xylem tissue of a farnesol-rich individual tree. From the assembly of 12,537 contigs, we identified seven different terpene synthases (TPSs), several cytochromes P450, allylic phosphatases, as well as transcripts of the mevalonic acid and methylerythritol phosphate pathways. Five of the *S. spicatum* TPS sequences were previously unknown. The full-length cDNA of *SspiTPS4* was cloned and the enzyme functionally characterised as a multi-product sesquisabinene B synthase, which complements the previous characterization of sanatalene and bisabolol synthases in *S. spicatum*. The genes identified in this work will form a basis for future studies on natural variation of sandalwood terpenoid oil profiles.

**Keywords:** Sandalwood; *Santalum spicatum*; fragrance oil; sesquiterpene; α-santalol; β-santalol; *E,E*-farnesol
6.3. Introduction

The heartwood of several members of the sandalwood genus (*Santalum*, Santalaceae) is highly prized for the sesquiterpene-rich oil of mature trees (Adams, 1995; Brennan and Merlin, 1993). *Santalum spicatum* is an important oil-bearing species, which has contributed substantially to the economic development of Western Australia. Major essential oil components include α- and β-santalol, α-bisabolol and *E,E*-farnesol along with olefin components such as santalenes, sesquisabinene, sesquiphellandrene and dendrolasin. Variation in oil composition is present across the natural range of distribution of *S. spicatum* (Moretta, 2001). This variation and the higher levels of *E,E*-farnesol contribute to the lower value of *S. spicatum* oil compared to *S. album* oil in the fragrance industry, where the preferred quality of sandalwood oils is largely defined by high levels of α- and β-santalol (Howes et al., 2004; Verghese et al., 1990). Elucidation of sesquiterpenoid biosynthesis in sandalwood provides a basis to understand composition of oil quality and its variability. This in turn may support efforts towards tree improvement for oil yields and quality in sustainable sandalwood plantations and can afford opportunities for biotechnological sandalwood oil production. In nature, terpenoids are thought to protect sandalwood trees, as heartwood extractives are active against fungi and bacteria (Hammer et al., 1998; Jirovetz et al., 2006).

In plants, terpenoids are produced by terpene synthases (TPSs) from linear prenyl diphosphates, geranyl diphosphate (GPP) for monoterpenes, farnesyl diphosphate (FPP) for sesquiterpenes and geranylgeranyl diphosphate (GGPP) for diterpenes (Chen et al., 2011). GPP, FPP and GGPP are derived from the condensation of isopentyl
diphosphate (IPP) and one, two or three molecules, respectively, of dimethylallyl diphosphate (DMAPP). DMAPP and IPP are produced from two pathways, the mevalonate (MEV) pathway and the methylerithritol phosphate (MEP) pathway (Fig. 6.S1). Chemical diversity of terpenoids in plants results from a manifold different rearrangements of carbocation intermediates in the reactions catalysed by TPSs (Davis and Croteau, 2000, Degenhardt et al., 2009). Many TPSs are multi-product enzymes and variations of as little as a single amino acid substitution in conserved regions of TPSs can alter product profiles. Plant genomes typically contain families of many similar yet functionally diverse TPSs (Chen et al., 2011). Several cytochrome P450 (P450) family enzymes produce additional alcohol functionalities of sandalwood terpenoids, such as the santalols and bisabolols, and contribute another level of terpenoid diversity (Diaz-Chavez et al., 2013). Variations in the genomic and biochemical makeup as well as expression variations of TPSs and P450s may explain some of the variations of sesquiterpenoids present in natural S. spicatum populations.

Substantial progress has been made on the elucidation of TPSs and P450s of sesquiterpene biosynthesis in three different Santalum species, S. spicatum, S. album and S. austrocaledonicum. The primary focus has been on the santalols and bisabolols, while biosynthesis of several other less abundant bicyclic sesquiterpenes has also been explained (Diaz-Chavez et al., 2013; Jones et al., 2008; Jones et al., 2011). TPS cDNAs encoding santalene synthases are highly conserved with 94% to 96% sequence identity across the three species, which was reflected in the conservation of enzyme functions indicative of an important ecological role in the Santalum genus. In contrast, a bisabolol synthase-like TPS showed substantial sequence and functional variations, producing α-bisabolol in S. spicatum and β-bisabolene in S. album and S.
austrocaledonicum. Of these three species, S. album appears to be the least chemically diverse (Jones et al., 2008) while substantial differences in composition have been noted for S. spicatum and S. austrocaledonicum (Butaud et al., 2003; Moretta, 2001; Page et al., 2010).

To establish a larger set of genes that contribute to the biosynthesis and possibly the variation of S. spicatum sequiterpenoids, we used transcriptome sequencing and mining which is a proven approach for the identification of genes of terpenoid biosynthesis in non-model species (Zerbe et al., 2013). We report here on gene discovery in the transcriptome of S. spicatum heartwood xylem, the main sandalwood oil accumulating tissue, and the identification and characterization of SpiTPS4 sesquisabinene B synthase.

6.3. Experimental

6.3.1. Materials

S. spicatum material for RNA extraction was obtained from trees growing in a 12 year-old plantation in the south-west of Western Australia (Narrogin). Heartwood xylem was collected by drilling the stems of trees at 30 cm above ground using a cordless drill. Eight individual trees were sampled which had been previously characterized for oil content and yield (Moniodis et al., submitted). Tissue samples were snap frozen in liquid nitrogen prior to RNA extraction. Chemicals, substrates and reagents were from
Sigma Aldrich (St. Louis, MO, USA) unless otherwise indicated. Restriction endonucleases and T4 DNA ligase were from New England Biolabs (Ipswich, MA).

6.3.2. RNA extraction and generation of cDNA

Total RNA was extracted from xylem shavings according to Kolosova et al. (2004) and Jones et al. (2011). Isolated RNA was stored at -80 °C. cDNA was generated from 5 µg of total RNA using Superscript III (Invotrogen, CA). cDNA libraries for 5’ and 3’ rapid amplification of cDNA ends (RACE) were made according to manufacturers instructions (Clontech, CA).

6.3.3. Transcriptome sequencing of S. spicatum xylem

RNA isolated from a single, E,E-farnesol rich tree was sent to the Australian Genome Research Facility, Melbourne, Australia for cDNA library construction and sequencing on the GLX-454 platform (Roche). Half of one sequencing plate generated 489,364 reads, which were assembled using GS De Novo Assembler into 12,537 contigs of 500 to 1,800 bps length. The 12,537 sequences were grouped using Blast2GO (blast2go.com) to assign Gene Ontology terms (GO terms). Known TPSs from S. spicatum (Jones et al. 2011) were used as BLAST queries to search the library for additional TPS sequences. Additional genes in the MEV and MEP pathway were identified using sequences from Arabidopsis thaliana (AED95638.1, AED95639.1, PI4891.1, AED93690.1, AED93691.1, QO9152.2, O04146, Q43315.1, Q38929.3,
AAB67741.1, Q38854.2, AEE83625.1, AED97658.1, AEC05588.1, Q888C5.1, Q9CAK8.1, AED97353.1, AEE86362.1), Catharanthus roseus (ACC7779661.1, CAC80883.1) and Rhodotorula glutinis (EGU11849.1) as search sequences.

6.3.4. Phylogenetic analysis

TPS sequences were aligned using Clustal W. A phylogenetic tree was constructed using the software MEGA version 5.2 (Tamura et al., 2011), employing the neighbor-joining (NJ) algorithm.

6.3.5. FLcDNA cloning of SspiTPS4

The FLcDNA of SspiTPS4 was cloned using 5’RACE (Clontech, CA). The 5’ sequence was obtained using the gene specific reverse primer 5’-GAAAAGGTCCTTGCTCATCAAGGATGTCC-3’ and universal forward primer (Clonetech). The full-length open reading frame was amplified using forward and reverse primers 5’-ATGGATTTGTGCCAGATCCCGCC-3’ and 5’-TTACTCCTCATCTAGCGTAACTGGGTGAA-3’. PCR was done in 50 µl reactions using a high fidelity DNA polymerase (Phusion® HF, Espoo, Finland), 1.5 mM Mg2+, 200 µM of each dNTP, 1 µM of each primer. Thermocycling involved initial denaturation at 98°C for 1 minute followed by 35 cycles of 10 seconds at 98°C, 30 seconds at 57°C, 1 minute at 72°C, with a final extension at 72°C for 10 minutes. Agarose gel electrophoresis with SYBR® safe stain (Invitrogen) showed a 2 kbp
amplicon which was cloned into pJET1.2 (Fermentas) for sequence verification and subcloned into the pET-28b(+) expression vector (Novagen, San Diego CA) with Nco1 and Xho1 restriction sites adding a C-terminal His<sub>6</sub> tag. Primers for subcloning were 5′-TACCATGGATGGATTTGTGCCAGATCCCG-3′ and 5′-TGGCGGCCGCTTCCTCCTCATCTAGCGTAACTG-3′. All constructs were sequence verified. The pET28b(+)-SspiTPS4 construct was transformed into C41 E. coli cells (Avidis, Saint-Beauzire, France) containing the pRARE plasmid from Rosetta 2 cells (Novagen).

6.3.6. SspiTPS4 protein expression in E. coli and purification

Transformed E. coli were grown overnight in 3 mL Luria-Bertani (LB) medium with 50 µg/mL kanamycin and chloramphenicol at 37°C. The overnight culture was used to inoculate 300 mL of rich LB medium containing both antibiotics at 37 °C with continuous shaking. At OD<sub>600</sub> of 0.8 cultures were cooled down to 16°C and isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.2 mM. Cultures were incubated for additional 12 hours at 16°C before harvesting by centrifugation (4,000 x g; 45 minutes). The pellet was lysed using a high pressure cell cruncher in buffer of 0.5 mg of DNase1 and RNaseA, protease inhibitor cocktail (Roche), 1 mM MgCl<sub>2</sub>, 50 mg lysozyme, 40 mg DTT, 20 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 7.4), 500 mM NaCl and 30 mM imidazole. The cell extracts was cleared by centrifugation (12,000 x g; 75 minutes at 4 °C). Protein in the supernatant was purified using Nickel affinity HisTrap™ columns (GE Healthcare) with 500 mM imidazole elution buffer (pH 7.4) containing 20 mM Na<sub>2</sub>HPO<sub>4</sub> and 500 mM NaCl. The purified protein fraction was
buffer-exchanged using a PD-10 column (GE Healthcare), equilibrated with buffer containing 25 mM HEPES (pH 7.4), 10% glycerol, 100 mM KCl. Protein concentration was determined using a Nanodrop spectrophotometer $A_{280}$ with extinction coefficients calculated by amino acid composition (ExPasy.org). SDS-polyacrylamide gel electrophoresis and Coomassie blue staining was used to confirm the presence of purified recombinant proteins of the expected size (64 kDa).

6.3.7. *SspiTPS4* enzyme assays

Samples containing 10 µg of purified *SspiTPS4* protein were assayed for TPS activity in the presence of Mg$^{2+}$ or Mn$^{2+}$ according to O’Maille et al. (2004). Assays were started by addition of GPP, FPP or GGPP to a final concentration of 100 µM. Assays were overlayed with 500 µL of hexane prior to incubation at 30 °C for 2 hours, after which assay vials were vortexed and stored at -80 °C prior to GC-MS analysis. Assays were thawed and centrifuged to separate the organic/aqueous layers prior to GC-MS analysis. Assays were performed in triplicate.

6.3.8. GCMS analysis

Gas chromatography was performed on a 30 m x 0.25 mm ID x 0.25 µM DB-Wax column (Agilent). Oven temperature started at 40 °C and held for 5 minutes, then raised at 3 °C per minute to 240 °C where it was held for a further 25 min. Carrier gas was helium at a flow rate of 1 ml per minute. One microliter injections (10:1) at 200 °C were
performed and detection was by mass-spectrometry (MS: at 250 °C). Mass spectra were compared to the 2008 NIST library and the literature. Kovats retention indices were calculated for all compounds using an n-alkane standard.

6.3.9. Accession Numbers

The full-length cDNA sequences reported in this manuscript have been deposited in NCBI Genbank with accession numbers KM091271 and KM091272.

6.4. Results and Discussion

6.4.1. Transcripts of the S. spicatum MEV and MEP pathway

To identify the core biosynthetic steps of sesquiterpene formation in S. spicatum, we explored a transcriptome established by 454-sequencing of RNA from xylem tissue containing ray parenchyma cells where sandalwood oil is thought to be synthesised (Jones et al., 2008). The transcriptome library of 489,364 reads was produced from a single farnesol-rich tree and assembled into 12,537 apparently unique contig sequences. Contigs were classified into functional ontology groups (Fig. 6.S2). More than half (55%) of the contigs had matches of known functions in other species. Of these, the majority (28%) were annotated with “cellular” or “metabolic processes”. Genes involved in secondary metabolism were found in “metabolic processes” and “response to stimuli” groups, which comprised 14% and 7% of the transcriptome, respectively.
The transcriptome included candidate TPS, P450 and allylic phosphatase sequences, as well as sequences for genes of the MEV and MEP pathways (Table 6.1). All steps of the MEV pathway, except for phosphomevalonate synthase, were found in the heartwood xylem transcriptome. In contrast, only two enzymes of the MEP pathway were represented. These results are consistent with the MEV pathway being the primary route by which sesquiterpene precursors are produced in plants and the heartwood xylem tissue being particularly rich in sesquiterpenoids. The most abundant transcript of the terpenoid pathway in the xylem transcriptome was HMG-CoA reductase 1 (HMGR1) classified based on sequence relatedness with *Arabidopsis thaliana* HMGR1 and HMGR2 (GenBank accession no. AEE35849 and AEC06618 respectively), a critical step for isoprenoid biosynthesis in plants (Chye et al., 1992; Goldstein and Brown, 1990). The MEP pathway, which provides the isoprenoid building blocks for monoterpenes and diterpenes was underrepresented in the transcriptome of *S. spicatum* heartwood matching the low abundance or lack of these compounds.
Table 6.1 - Summary of genes identified in the xylem transcriptome encoding enzymes of the MEV and MEP pathway, TPS, cytochrome P450, and allylic phosphatase

<table>
<thead>
<tr>
<th>Contig Function</th>
<th>Contig Length (bp)</th>
<th>Read Number</th>
<th>Species</th>
<th>Best Match</th>
<th>Blast E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mevalonic Acid Pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACT1 Acetyl-CoA thiolase</td>
<td>1227</td>
<td>67</td>
<td>Bacopa monnieri</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>AACT2</td>
<td>1805</td>
<td>1</td>
<td>Bacopa monnieri</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>HMGS 3-Hydroxy-3-methylglutaryl-CoA Synthase</td>
<td>1413</td>
<td>34</td>
<td>Camellia sinensis</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>HMGR1 3-Hydroxy-3-methylglutaryl-CoA Reductase</td>
<td>1909</td>
<td>339</td>
<td>Camptotheca acuminata</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>HMGR2</td>
<td>2030</td>
<td>1</td>
<td>Camptotheca acuminata</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>MK Mevalonate Kinase</td>
<td>1100</td>
<td>18</td>
<td>Hevea brasiliensis</td>
<td>7.00E-159</td>
<td></td>
</tr>
<tr>
<td>MVD Mevalonate 5-diphosphate decarboxylase</td>
<td>843</td>
<td>27</td>
<td>Catharanthus roseus</td>
<td>2.00E-172</td>
<td></td>
</tr>
<tr>
<td>MEP Pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS 1-Deoxyxylulose-5-phosphate synthase</td>
<td>777</td>
<td>6</td>
<td>Arabidopsis thaliana</td>
<td>9.00E-159</td>
<td></td>
</tr>
<tr>
<td>DXR 1-Deoxyxylulose-5-phosphate reductoisomerase</td>
<td>1303</td>
<td>1</td>
<td>Hevea brasiliensis</td>
<td>1.00E-114</td>
<td></td>
</tr>
<tr>
<td>IPP isomerasases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDI Isopentyl diphosphate isomerase</td>
<td>789</td>
<td>5</td>
<td>Santalum album</td>
<td>5.00E-161</td>
<td></td>
</tr>
<tr>
<td>Prenyltransferases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPPS Geranyl diphosphate synthase</td>
<td>536</td>
<td>2</td>
<td>Solanum hyspersicum</td>
<td>6.00E-52</td>
<td></td>
</tr>
<tr>
<td>FPPS Farnesyl diphosphate synthase</td>
<td>1154</td>
<td>25</td>
<td>Santalum album</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>GGPPS Geranylgeranyl diphosphate synthase</td>
<td>798</td>
<td>1</td>
<td>Corylus avellana</td>
<td>1.00E-73</td>
<td></td>
</tr>
<tr>
<td>Terpene Synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3W204 Santalene synthase</td>
<td>1713</td>
<td>96</td>
<td>Santalum spicatum</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>ADOB7004 α-bisabolol synthase</td>
<td>1731</td>
<td>57</td>
<td>Santalum spicatum</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>SspTPS4 S. spicatum unknown TPS5</td>
<td>667</td>
<td>1</td>
<td>Santalum austrocaledonicum</td>
<td>3.00E-84</td>
<td></td>
</tr>
<tr>
<td>S. spicatum unknown TPS6</td>
<td>867</td>
<td>4</td>
<td>Santalum murrayanum</td>
<td>2.00E-166</td>
<td></td>
</tr>
<tr>
<td>S. spicatum unknown TPS7</td>
<td>746</td>
<td>1</td>
<td>Quercus ilex</td>
<td>2.00E-93</td>
<td></td>
</tr>
<tr>
<td>S. spicatum unknown TPS8</td>
<td>765</td>
<td>4</td>
<td>Vitis vinifera</td>
<td>1.00E-48</td>
<td></td>
</tr>
<tr>
<td>CYP76 P450s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssp/CYP76 39 S. spicatum CYP76-610</td>
<td>1530</td>
<td>5</td>
<td>Santalum album</td>
<td>0.00E+00</td>
<td></td>
</tr>
<tr>
<td>Phosphatases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. spicatum phosphatase PAP2-like</td>
<td>1181</td>
<td>12</td>
<td>Croton stellatapilosus</td>
<td>1.00E-124</td>
<td></td>
</tr>
</tbody>
</table>

6.4.2. Terpene synthases

Seven different TPS-like partial sequences were identified in the xylem transcriptome, including previously characterised S. spicatum santalene synthase (SspiSSy) and α-bisabolol synthase (SspiBS) (Jones et al., 2011). Five TPS sequences were new for S. spicatum and like other Santalum TPSs were phylogenetically associated with the TPS-a and TPS-b subfamilies (Fig. 6.1) of angiosperm mono- and sesquiterpene synthases (Bohlmann et al., 1998; Bohlmann et al., 1997). The full-length cDNA of SspiTPS4, which was represented with five sequence reads in the transcriptome library, was obtained by 5’-RACE, expressed in E. coli and functionally characterised. For the other
four new TPS sequences, *SpyrTPS5, -6, -7 and -8*, two of which were present only with one read, full-length cDNAs could not be recovered from the available RNA.

Figure 6.1 - Neighbor-joining phylogenetic tree of TPS from *S. spicatum* (TPS-a and b), and select TPS of other species (see Supplemental Table 1). Sequences were aligned using Clustal W and the phylogenetic tree constructed with MEGA5.2 (Tamura et al., 2011). Bootstrap confidence values over 50% are shown at branch points based on 500 replications.
6.4.3. *SspiTPS4 is a multi-product sesquisabinene B synthase*

The predicted protein encoded by *Sspi*TPS4 showed high amino acid homology to a sesquisabinene synthase from *S. album* (93% similarity, 87% identity; GenBank accession number ADP37190) (Jones et al., 2011). *Sspi*TPS4 encodes a 566 amino acid protein, which clustered with the TPS-b subfamily. Like *Sspi*SSy and *Sspi*BS (Jones et al., 2011), *Sspi*TPS4 did not contain an N-terminal transit peptide for plastidial targeting. *Sspi*TPS4 contains the RRX₈W motif near the N-terminus and the aspartate rich (DDxxD) metal ion-binding domain (Fig. 6.S3). We obtained a genomic sequence of *Sspi*TPS4 of 3,405 bps with a gene structure of 6 introns and 7 exons similar to other TPS-b genes (Fig. 6.S4). Intron/exon boundaries were typical of TPS genes with a 3' ~ NˇGT, 5' AGˇN~ pattern.

The recombinant *Sspi*TPS4 protein expressed in *E. coli* had a molecular mass of 66 kDa similar to other sesquiterpene syntheses. *Spi*TPS4 converted GPP and FPP substrates to mono- and sesquiterpene products, respectively. However, only the sesquiterpene products were constituents of *S. spicatum* oil. *Sspi*TPS4 was most active with *E,E*-FPP and Mg²⁺ producing a profile of six main sesquiterpenes (Fig. 6.2).
Figure 6.2 - GC-MS chromatogram of *in vitro* assay products of recombinant *SspITPS4* using FPP as a substrate and Mg$^{2+}$. Peaks: 1 unknown, 2 sesquisabinene B, 3 α-acoradiene, 4 γ-curcumene, 5 β-bisabolene, 6 β-sesquiphellandrene

The most abundant product was sesquisabinene B (58%) in addition to β-bisabolene (18%), γ-curcumene (12%), β-sesquiphellandrene (9%), α-acoradiene (2%) and another unidentified component (1%) (Table 6.2). No activity was detected when *SspITPS4* was assayed with *E,Z*-FPP. Likewise, assays with Mn$^{2+}$ instead of Mg$^{2+}$ showed no activity.

Table 6.2 - Product profile *SspITPS4* with FPP and Mg$^{2+}$ identified by GC-MS analysis using a DB-Wax column, NIST08 library and Kovats retention indices. Peak numbering matches peaks shown in Figure 6.2

<table>
<thead>
<tr>
<th>Peak No.</th>
<th>Retention Index (R.I.)</th>
<th>Literature R.I.</th>
<th>Compound from <em>SspITPS4</em></th>
<th>Total Composition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1538</td>
<td>1521</td>
<td>unknown</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1642</td>
<td>1657</td>
<td>sesquisabinene B</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>1662</td>
<td>1680</td>
<td>α-acoradiene</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1673</td>
<td>1688</td>
<td>γ-curcumene</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>1709</td>
<td>1726</td>
<td>β-bisabolene</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>1754</td>
<td>1766</td>
<td>β-sesquiphellandrene</td>
<td>9</td>
</tr>
</tbody>
</table>
When $SspITPS4$ was incubated with GPP and $\text{Mg}^{2+}$, six different products were formed; β-pinene (24%), myrcene (21%), sabinene (18%), α-pinene (16%), α-terpineol (11%) and linalool (10%). The same compounds were produced when $\text{Mn}^{2+}$ was replaced with $\text{Mg}^{2+}$, however levels of product formation were markedly lower (Fig. 6.3). Monoterpenes have been reported from sandalwood only in very low concentrations (Valder et al., 2003). The major $SspITPS4$ sesquiterpene product sesquisabinene B is also present in $S. \text{accumunatum}$, and β-bisabolene is also found in $S. \text{album}$, $S. \text{accumunatum}$ and $S. \text{insulare}$ oil extracts (Baldovini et al., 2011).

![GC chromatogram of in vitro assay products of recombinant $SspITPS4$ using GPP as a substrate and $\text{Mg}^{2+}$ or $\text{Mn}^{2+}$ (dotted). Peaks: 1 α-pinene, 2 β-pinene, 3 sabinene, 4 myrcene, 5 linalool, 6 α-terpineol](image)

**Figure 6.3** - GC chromatogram of *in vitro* assay products of recombinant $SspITPS4$ using GPP as a substrate and $\text{Mg}^{2+}$ or $\text{Mn}^{2+}$ (dotted). Peaks: 1 α-pinene, 2 β-pinene, 3 sabinene, 4 myrcene, 5 linalool, 6 α-terpineol

Proposed mechanisms of formation of $SspITPS4$ products (Fig. 6.4) begin with the ionization of FPP and isomerization at the C2-C3 double bond to the tertiary nerolidyl cation (Cane and Iyengar, 1979; Davis and Croteau, 2000). The cisoid conformer of the
nerolidyl cation can undergo cyclisation to generate a 1,6-bisabolyl cation. Subsequent hydride shifts and proton abstractions yield the six cyclic sesquiterpenes of \textit{SspiTPS4}.

Figure 6.4 - Proposed reaction mechanism for the formation of sesquiterpene products by \textit{SspiTPS4} involving ionization of \textit{E,E}-FPP, isomerization of the farnesyl cation, followed by different cyclisations, hydride shifts, and deprotonations. Numbering refers to that of \textit{E,E}-FPP

Sesquiterpene synthases produce a large amount of the terpenoid diversity of plants, which stems from their ability to convert a single substrate into a wide array of terpene
structures (Degenhardt et al., 2009; Chen et al., 2011). The sesquiterpene mixtures produced in the heartwood of _Santalum_ species show substantial compositional variation (Moretta 2001). Some of this variation might be attributed to sequence and functional variations in TPS enzyme. In sesquiterpene synthases, the C-terminal domain contains the active site where sequence changes can influence product profiles. For example, in the maize TPS4 and TPS5 enzymes, four amino acid substitutions in the C-terminal region were responsible for altering product proportions, while producing the same suite of compounds (Kollner et al., 2004). _S. spicatum_ predominantly displays quantitative variations in sesquiterpene components. To test whether variants of _SspiTPS4_ could help explain the natural chemical diversity of the oil, _SspiTPS4_ cDNAs were cloned from eight individuals, which differed in their quantitative properties of _SspiTPS4_ oil products. However, variations in the N terminal region at residues 64, 177 and 212 (Fig. 6.S3) may be less likely to influence variation in product profiles than changes in and around the active site of the C terminal domain (Back and Chappell, 1996; Greenhagen et al., 2006; Kollner et al., 2004).

6.4.4. **Cytochromes P450 and SspiCYP76F39 of santalol biosynthesis**

P450s for the hydroxylation of terpenes may be members of the CYP71, CYP72 and CYP76 families (Hamberger and Bak, 2013; Nelson et al.; 1996, Nelson and Werck-Reichhart, 2011; Zerbe et al., 2013). Although several P450 sequences were identified in the _S. spicatum_ xylem transcriptome, only one P450 was found which clustered into the CYP76 family, and no members of the CYP71 and CYP72 groups were found. The CYP76 member was present with 5 sequence reads in the transcriptome. The full-
length sequence was 99% identical at the amino acid level to the recently described \textit{S. album} CYP76F39v1 which hydroxylates several sandalwood sesquiterpenes to produce \(\alpha\)-, \(\beta\)-, \textit{epi}-\(\beta\)-santalol and \(\alpha\)-\textit{E}-bergamotol and isomers thereof (Diaz-Chavez et al., 2013). Identification of \textit{Sspi}CYP76F39 in the heartwood xylem transcriptome confirmed the value of this sequence resource for the discovery of genes involved in sandalwood oil biosynthesis.

\textbf{6.4.5. \textit{Other potential candidate genes for \textit{S. spicatum} terpenoid biosynthesis}}

Despite the abundance of the \(E,E\)-farnesol metabolite in \textit{S. spicatum} heatwood oil, a corresponding abundance of a unique new TPS transcript was not apparent, although it cannot be excluded that one of the low abundance partial TPS sequence may represent an \(E,E\)-farnesol synthase. We searched the transcriptome for allylic phosphatases, which may conceivably produce farnesol by hydrolysis of FPP. One PAP-2 like phosphatase sequence was identified with 12 reads and relatively closely resembled the allylic phosphatase (CsPDP) from \textit{Croton stellatopilosus} (Nualkaew et al., 2012) with 70\% identity and 85\% similarity at the amino acid level (Table 6.1). CsPDP was reported to convert the C20-geranylgeranyl diphosphate (GGPP) into the corresponding alcohol (GGOH). Given its relative transcript abundance, the \textit{S. spicatum} PAP-2 like enzyme may be a candidate for dephosphorylation of FPP to generate the corresponding acyclic sesquiterpene alcohol.
6.5. Conclusions

The sequences of the MEV pathway, TPS and P450 described here provide the most complete map of sesquiterpene biosynthesis of any sandalwood species thus far. These sequences together with the previously characterized sandalwood FPP synthase, TPS and P450 sequences can be used in future work to comprehensively explore sequence variations and monitor gene expression profiles of sesquiterpenoid biosynthesis in *S. spicatum* and other *Santalum* species. Expression profiles could be of substantial diagnostic value as a toolkit of biomarkers to assess onset and progress of sandalwood oil production in the heartwood of maturing trees.

6.6. Acknowledgements

The authors are grateful for the financial assistance of the Australian Research Council and the Forest Products Commission (FPC) of Western Australia through linkage project LP0882690. Work in the laboratory of Joerg Bohlmann was supported with funds from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors thank Mr. Len Norris for field assistance and technical advice, Mr. Jon Brand of FPC for providing access to plantation trees, Mr. Mack Yuen (UBC) for bioinformatics support, and Ms. Lina Madilao (UBC) for GC-MS support.
### Table 6.51 - TPS used to construct the neighbor-joining phylogenetic tree (Fig. 6.1)

<table>
<thead>
<tr>
<th>Species</th>
<th>TPS</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abies grandis</td>
<td>pinene synthase</td>
<td>024475</td>
</tr>
<tr>
<td>Abies grandis</td>
<td>limonene synthase</td>
<td>O22340</td>
</tr>
<tr>
<td>Abies grandis</td>
<td>β-phellandrene synthase</td>
<td>AAF61453</td>
</tr>
<tr>
<td>Actinidia arguta</td>
<td>linalool synthase</td>
<td>ADD81294</td>
</tr>
<tr>
<td>Actinidia chinensis</td>
<td>nerolidol synthase</td>
<td>AER36088</td>
</tr>
<tr>
<td>Actinidia deliciosa</td>
<td>α-farnesene synthase</td>
<td>ACO40485</td>
</tr>
<tr>
<td>Actinidia deliciosa</td>
<td>germacrene D synthase</td>
<td>AAX16121</td>
</tr>
<tr>
<td>Antirrhinum majus</td>
<td>nerolidol/linalool 1 synthase</td>
<td>ABR24417</td>
</tr>
<tr>
<td>Antirrhinum majus</td>
<td>myrcene synthase</td>
<td>AAO41727</td>
</tr>
<tr>
<td>Cinnamomum tenuipilum</td>
<td>geraniol synthase</td>
<td>Q8GUE4</td>
</tr>
<tr>
<td>Citrus unshiu</td>
<td>d-limonene synthase</td>
<td>BAD27257</td>
</tr>
<tr>
<td>Clarkia breweri</td>
<td>linalool synthase</td>
<td>AAD19840</td>
</tr>
<tr>
<td>Gossypium arboreum</td>
<td>γ-cadinene synthase</td>
<td>Q43714</td>
</tr>
<tr>
<td>Ocimum basilicum</td>
<td>geraniol synthase</td>
<td>AAR11765</td>
</tr>
<tr>
<td>Picea abies</td>
<td>α-farnesene synthase</td>
<td>AAS47697</td>
</tr>
<tr>
<td>Picea sitchensis</td>
<td>(+)-3-carene synthase</td>
<td>ADU85928</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>α-farnesene synthase</td>
<td>Q84KL5</td>
</tr>
<tr>
<td>Quercus ilex</td>
<td>myrcene synthase</td>
<td>Q93X23</td>
</tr>
<tr>
<td>Santalum album</td>
<td>santalene synthase</td>
<td>E3W202</td>
</tr>
<tr>
<td>Santalum album</td>
<td>β-bisabolene synthase</td>
<td>B5A434</td>
</tr>
<tr>
<td>Santalum album</td>
<td>sesquiterpene synthase</td>
<td>B5A435</td>
</tr>
<tr>
<td>Santalum album</td>
<td>sesquisabinene B synthase</td>
<td>ADP37190</td>
</tr>
<tr>
<td>Santalum austrocaledonicum</td>
<td>santalene synthase</td>
<td>E3W203</td>
</tr>
<tr>
<td>Santalum austrocaledonicum</td>
<td>β-bisabolene synthase</td>
<td>ADO87003</td>
</tr>
<tr>
<td>Santalum austrocaledonicum</td>
<td>α-humulene/γ-cadinene synthase</td>
<td>E3W207</td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>santalene synthase</td>
<td>E3W204</td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>α-bisabolol synthase</td>
<td>E3W206</td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>sesquisabinene B (TPS4) synthase</td>
<td></td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>hedyacaryol/α-bulnesene synthase</td>
<td>E3W208</td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>TPS5</td>
<td></td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>TPS6</td>
<td></td>
</tr>
<tr>
<td>Santalum spicatum</td>
<td>TPS7</td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Activity</td>
<td>Accession Number</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Solanum habrochaites</td>
<td>Santalene/bergamotene synthase</td>
<td>ACJ38409</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>(-)-alpha-terpineol synthase</td>
<td>AAS79352</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>Nerolidol/linalool synthase</td>
<td>ADR74212</td>
</tr>
<tr>
<td>Zea mays</td>
<td>Sesquithujene A synthase</td>
<td>Q6JD70</td>
</tr>
</tbody>
</table>
Figure 6.51 - Metabolic pathway for terpene biosynthesis according to Tholl and Lee (2011). Abbreviations: AACT, Acetyl-CoA thiolase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; MK, mevalonate kinase; PMK, phosphomevalonate kinase; MDC, pyrophosphomevalonate decarboxylase; IPPI, isopentenylpyrophosphate isomerase; DMAPP, dimethylallyl pyrophosphate; FPPS, farnesyl pyrophosphate synthase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 2-C-methylerythritol 4-phosphate reductase; MCT, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 1-hydroxy-2-methyl-butenyl 4-diphosphate reductase; GPS, geranyl pyrophosphate synthase; GGPP, geranylgeranyl pyrophosphate synthase; TPS, terpene synthase; P450, cytochrome P450.
Figure 6.S2 - Functional classification of *S. spicatum* xylem transcriptome
N terminal

NcDNA MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N103 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N106 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N116 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N114 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N125 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60
N151 MDLCQIPPSPIQPSNFGPDDSSVRRSANYPANLWDYDFLSQSLHIIPTVEQHQVGLAE 60

NcDNA KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N103 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N106 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N116 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N161 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N114 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N125 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120
N151 KLKGEVKSLITGPMEPLAKLEFIDNVRRLGLKYQFETEIKEALVNISRDGYESWWVDNLR 120

NcDNA ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N103 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N106 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N116 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N161 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N114 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N125 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180
N151 ATALRFRLLRENGIFVPQDVFERFQNKEFGKFNCELQDVKGLINLYEASFLGWEGEDIL 180

C terminal

NcDNA DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N103 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N106 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N116 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N161 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N114 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N125 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240
N151 DEARTFSTAQLKVEKISPLNKLIVHHLPLHLHRAIYEARWFIDYDDADEMNTPT 240

NcDNA LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N103 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N106 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N116 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N161 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N114 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N125 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300
N151 LLKYAKLDNFIVSQFHQAERILARWWVTVGTLKLPFARNGLIQSYMYAIGMFLEPYLG 300

DDxxD

NcDNA VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N103 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N106 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N116 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N161 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N114 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N125 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360
N151 VREMAEKVLALITDDYDYVDYGMEELFDTIDERWISKVDQLPRTRMLPLMTMTN 360

NcDNA SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N103 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N106 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N116 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N114 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N125 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420
N151 SNNIGYWALKEGFGNPITATKVVADOLKSYTKEAKWFHHEKPTLEELYENALVSIGFP 420

NcDNA NLLVTSYLLTVNPTKKLDYDSPLFVRSACLCHINDLGTSPDEMERGDNLKLSIQG 480
N103 NLLVTSYLLTVNPTKKLDYDSPLFVRSACLCHINDLGTSPDEMERGDNLKLSIQG 480
N106 NLLVTSYLLTVNPTKKLDYDSPLFVRSACLCHINDLGTSPDEMERGDNLKLSIQG 480

192
Figure 6.53 - Comparison of the deduced amino acid sequence of SspiTPS4 from eight individuals of S. spicatum. Amino acid differences are marked by black boxes. The aspartate-rich DDxxD region is shown. The N terminal domain is labelled with a bar.
Figure 6.S4 - Genomic structure of Ssp/TPS4 showing the intron/exon boundaries. $R(R/P)X_8$ motif in the N-terminal region implicated in prenyl diphosphate ionization and the aspartate rich metal ion binding domain (DDxxD) are highlighted (Aubourg et al., 2002, Lesburg et al., 1998, Starks et al., 1997, Williams et al., 1998)
Figure 6.55 - Mass spectral data of in vitro assay products of recombinant Ssp/TPS4 using FPP as a substrate and Mg$^{2+}$. Numbering according to Figure 6.2: 1 unknown, 2 sesquisabinene B, 3 α-acoradiene, 4 γ-curcumene, 5 β-bisabolene, 6 β-sesquiphellandrene.
Figure 6.56 - Mass spectral data of *in vitro* assay products of recombinant SspTPS4 using GPP as a substrate and Mg²⁺. Numbering according to Figure 6.3: 1 α-pinene, 2 β-pinene, 3 sabinene, 4 myrcene, 5 linalool, 6 α-terpineol

1. α-pinene
2. β-pinene
3. sabinene
4. myrcene
5. linalool
6. α-terpineol
6.8. References


Hammer, K., Carson, C., Riley, T., 1998. In-vitro activity of essential oils, in particular *Melaleuca alternifolia* (tea tree) oil and tea tree oil products, against


Moretta, P., 2001. Extraction and variation of the essential oil from Western Australian sandalwood (Santalum spicatum). Doctor of Philosophy (PhD) thesis, The University of Western Australia, Perth, Australia.


1815-1820.


CHAPTER SEVEN

Genetic and environmental influences on essential oil composition in West Australian sandalwood (*Santalum spicatum*)

Jessie Moniodis, Michael Renton, Liz Barbour, and Margaret Byrne

1. School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley WA 6009 Australia

2. School of Chemistry and Biochemistry (M310), University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

3. The Department of Parks and Wildlife, 17 Dick Perry Avenue Technology Park, Western Precinct, Kensington, WA 6151 Australia

* corresponding author
7.1. Abstract

West Australian sandalwood, *Santalum spicatum* (R.Br.) A.DC., contains a highly valuable essential oil in the heartwood of mature trees which is also variable across its natural distribution. *S. spicatum* is currently harvested from natural stands in Western Australia (WA) which has resulted in the depletion of many natural resources. In this study, microsatellite markers were used to compare patterns of genetic variation genetic with the variability of essential oil composition in 186 trees distributed throughout natural stands in Western Australia, focussing on the semi-arid (wheatbelt) trees where current harvesting practises occur and plantations have been established. A small number of trees from two areas in the arid region (north and goldfields) were also sampled for genetic and chemical comparisons. Considerable variation was reported in the growing environment of *S. spicatum* and tree morphology was also diverse. Mantel tests revealed a significant association between oil composition and genetic distance across all trees (*r* = 0.129, *P* = 0.02), but not when individual regions were assessed separately. An Adonis test showed a significant association between oil composition and provenance (*P* = 0.001). The combined results indicate that variation in *S. spicatum* chemotypes is probably subject in large part to environmental control, associated in particular with the habitat where it grows. The ability to study this in detail is complicated by numerous factors, given the parasitic nature of sandalwood and the long development time of heartwood oil. Investigation of population structure using several approaches all supported the evolutionary divergence of the arid (north, goldfields) and the semi-arid (wheatbelt) populations in Western Australia. This study provides the basis for managing plantations of this species to improve oil quality, as well as improving knowledge of natural stands to aid conservation efforts.
7.1. Introduction

West Australian sandalwood, *Santalum spicatum* (R.Br.) A.DC., contains a highly valuable essential oil in the heartwood of mature trees. Sandalwood oil has widespread applications in perfumery, aromatherapy, medicines and in the manufacture of incense sticks (Clarke, 2006, Loneragan, 1990, Brennan & Merlin, 1993). *Santalum spicatum* is native to WA and contributed significantly to the economic development of this state in its early years. Currently, this species is harvested from natural stands, and many natural resources have been depleted due to high demand and unsustainable exploitation, combined with clearing for agriculture, habitat loss, slow regeneration, grazing by feral goats and illegal poaching (Rai, 1990, Loneragan, 1990). Thus, the past 15 years has seen a movement towards establishing plantations in the agricultural (wheatbelt) regions of WA to ensure a continued future supply, especially as worldwide demand is growing. Sandalwood plantations will alleviate the pressure on natural stands and provide the opportunity to improve plant stock through selection based on knowledge of genetic and environmental factors that may influence heartwood oil production.

The essential oil of *S. spicatum* is extracted from the wood and roots of mature trees. Oil composition is variable across its natural range, and this variability contributes to the lower value of wood compared to Indian sandalwood (*S. album*) (Moretta, 2001, Howes et al., 2004). Oil quality is largely defined by high levels of α- and β-santalol, and low levels of the allergenic E,E-farnesol (Howes et al., 2004, Verghese et al., 1990). The most valuable sandalwood source, *S. album*, displays little variability in oil content across its natural range with little or no farnesol present, α- and β-santalol amounts
consistently greater than 80 % and oil content as high as 9 % (Jones et al., 2007, Verghese et al., 1990). Other oil-bearing sandalwoods, *S. australaledonicum* endemic to the archipelago of New Caledonia, and *S. insulare*, found on 13 islands of Eastern Polynesia, also display variation in α- and β-santalol content and total oil yields in natural stands (Braun et al., 2007, Page et al., 2010).

Investigation of genetic diversity in an Australian germplasm collection of *S. album* in a nuclear RFLP study reported very low diversity for a tree species (Jones et al., 2009). In comparison, high levels of genetic diversity have been reported in *S. australaledonicum* and *S. insulare* using microsatellite markers. The greater genetic diversity in *S. australaledonicum* and *S. insulare* is also reflected in higher levels of chemical diversity across their natural range. If variability in the heartwood oil is heritable across *Santalum*, causes underlying this variation can be identified and the opportunity to select superior trees with desired heartwood properties for future domestication becomes possible. A considerable amount of research into the causes of chemical variability is still needed to improve the current management of plantations and natural stands across the genus.

Chemical variation in *S. spicatum* might be explained by genetic differences amongst provenances, while the species also grows across a range of diverse environments. A previous investigation into the genetic structure of *S. spicatum* trees using nuclear and chloroplast RFLPs, revealed genetic differences in arid (northern) and semi arid (southern) areas of WA, and moderate levels of genetic diversity compared to other native tree species (Byrne et al., 2003b). There are also known differences in heartwood
oil composition across natural populations (Moretta, 2001), although the relationship between genetic structure and oil content is currently unknown. Some studies have focused on associations between neutral markers and chemical composition in order to identify any association with broad genetic structure that may represent an influence of historical and environmental conditions on oil components. For example, a study on *S. austrocaledonicum* populations in New Caledonia found no significant correlation between genetic and chemical distance matrices using data from microsatellite markers and heartwood oil composition (Bottin et al., 2007). In a study on Turkish oregano (*Origanum onites* L.), the authors reported no significant relationship between terpenoid composition in the leaves of plants and genetic structure using random amplified polymorphic DNA (RAPD) markers (Tonk et al., 2010) and a study on tree basil (*Ocimum gratissi*) identified several chemotypes based on chemical and morphological data and found strong correlation between thymol, eugenol and geraniol chemotypes and RAPD markers (Vieira et al., 2001). In *S. spicatum*, oil biosynthesis may have a strong genetic influence; however, given its likely role in long-term defence, total oil production is likely to also be influenced by environmental features.

A number of environmental factors may drive selection of a specific oil phenotype including abiotic factors (temperature, rainfall, soil composition) and biotic factors (such as age of the trees, vegetation structure, host species and the presence of pathogens). *Santalum spicatum*, like almost all members of Santalaecae, is a hemi-parasite, meaning it obtains some nourishment from a host species such as *Acacia* or *Allocasuarina* for growth and development (Applegate et al., 1990, Loneragan, 1990b, Hewson & George, 1984). *Santalum spicatum* grows in variable climatic conditions that range from the semi-arid south (~ 300–600mm rainfall) to the arid northern areas of
Western Australia (~ 150–300mm rainfall) (Loneragan, 1990, Fox & Brand, 1993). Sandalwood trees also display considerable variation in height, form, leaf shape, colour and fruiting patterns across its geographic range. Such variations are a likely result of both genetic and environmental factors that may contribute to total heartwood oil production. Several studies have reported correlations between environmental factors and oil variability (Duarte et al., 2010, Vokou et al., 1993, Maffei et al., 1993, Tommasi et al., 2007, Boira & Blanquer, 1998). For example, chemical variability in Thymbra capitata L., and Thymus piperella L. have displayed positive correlations between certain chemotypes and climatic conditions (Boira & Blanquer, 1998, Tommasi et al., 2007). Additionally, some chemotypes of Thymus piperella L. were influenced by altitude and bioclimatic factors affecting water balance in soils (Boira & Blanquer, 1998). In S. spicatum, production of terpenoids in the heartwood of mature trees (> 10 years of age) is likely to be subject to the interactions of both genetic drift and selection pressures in the environment (Bottin et al., 2007).

The aim of the present study was to test the predictions that i) genetically similar individuals will share more similarities in essential oil composition, ii) oil composition is influenced by some environmental features, and iii) certain morphological features can be linked to essential oil composition or content. Since the RFLP study on genetic structure in S. spicatum, (Byrne et al., 2003a, Byrne et al., 2003b), microsatellites have been developed for this species (Millar et al., 2011). Microsatellites are expected to display higher levels of polymorphisms and expected heterozygosity than RFLP markers (Powell et al., 1996). Therefore, microsatellite markers were used to compare patterns of genetic variation genetic with the variability of essential oil composition in 186 trees distributed throughout natural stands in Western Australia. Focus was on the
semi-arid (wheatbelt) trees where current harvesting practises occur and plantations are being established; however, the study was extended to a small number of trees from two areas in the arid region (north and goldfields) for genetic and chemical comparisons.

7.3. Experimental

7.3.1. Oil extraction and analysis

For all individuals, wood shavings were collected for analysis of oil components by drilling into the stems of trees 30 cm from the base of the tree with a 25 mm manual hand drill. Oil was extracted from air-dried wood shavings (3-5 g) of the core and placed into ethanol for 7 – 14 days with 10 mM isobutyl benzene (IBB) as an internal standard. Essential oil composition was separated by gas chromatography, quantified by flame ionization detection (GC-FID) and identification of components was achieved using mass spectrometry (GC-MS) (Shimadzu). An external standard plot using a dilution series of authentic *S. spicatum* oil (Mt. Romance, Albany WA) was generated to determine total oil yield from extracts. Gas chromatography conditions were as follows: the stationary phase was a DB-Wax column, 30 m x 0.25 mm ID x 0.25 µM film thickness (Agilent); carrier gas was helium at 1 ml per minute; one microliter of extract was injected using a split ratio of 10:1; injector temperature was 200 °C; detector was set to 250 °C; oven temperature started at 40 °C and was raised in 4 °C increments to 220 °C where it was held for a further 25 min.; scan mode was used over the range of 41 to 250 m/z.
7.3.2. DNA extraction and genotyping

Leaves of *S. spicatum* trees were collected from a total of 186 individuals growing in natural stands throughout Western Australia; including the semi-arid, south wheatbelt (*n* = 144), the arid north Carnarvon and Shark Bay regions (*n* = 23) and the goldfields (*n* = 19) (Figure 7.1). Total genomic DNA was extracted from freeze-dried leaf tissue using a modified version of the CTAB/PVP method that incorporates sodium sulphite into the extraction buffer (Byrne & Moran, 1994, Byrne et al., 2001). Poor quality DNA was obtained for some samples, probably owing to poor health of individual trees, and a QIAGEN kit (DNeasy Plant Mini Kit) was used to extract DNA from these samples.

![Sampled Regions Map](image)

*Figure 7.6 - Map of Western Australia showing the approximate distribution of the 186 individuals sampled from the arid north regions (Northern and Goldfields) and the semi-arid (Wheatbelt) region in the south-west*
Genetic analysis was performed using twelve pairs of microsatellite primers: SsA103, SsA105, SsA109, SsA106, SsB122, SsB126, SsA113, SsC011, SsA001, SsA119, SsB011 and SsB128, recently developed for *S. spicatum* (Millar et al., 2011).

Microsatellite loci were amplified in a total volume of 15 µL per reaction containing 20 ng template DNA, PCR buffer (composed of 0.2 mM dNTPs, 50 mM KCl, 20 mM Tris-HCl pH 8.4), 0.3 µM forward and reverse primers (forward primers fluorescently labelled with FAM, NED, PET and VIC dyes) and 0.5 units of *Taq* polymerase (QIAGEN). Magnesium chloride concentration for most primer pairs was optimized at 1.5 mM with the exception of SsB126 and SsB128, which were lowered to a final concentration of 1.25 mM and 1.0 mM respectively to increase specificity. Two primer pairs, SsB126 and SsB011, also required addition of 1 M Betaine to the PCR mixture to improve amplification (Raj Chakrabarti & Schutt., 2002). PCR cycling conditions for all loci consisted of an initial denaturation at 96°C for 2 minute followed by 30 cycles of 30 seconds at 95 °C, 30 s at 56 °C annealing temperature, 30 s at 72 °C and a final extension at 72 °C for 5 minutes. One microliter of PCR product was added to 13 µL of LIZ500 /formamide solution. Fragment analysis was carried out by automated fluorescent scanning detection using an Applied Biosystems 3730 DNA Analyser (Applied Biosystems) and GenemapperTM v3.7 analysis software (Applied Biosystems).

7.3.3. **Genetic data analysis**

A genetic distance-based analysis was conducted in GenAlex V6.41 (Peakall & Smouse 2006), which converted the genetic data into a pairwise individual by individual genetic
distance matrix, followed by an Analysis of Molecular Variance (AMOVA). Genetic distance was calculated using the CS-chord distance (Cavalli-Sforza and Edwards, 1967). GenAlex V6.41 was also used to calculate the number of private alleles and group-specific alleles. Phylogenetic construction was based on the neighbour-joining method implemented in Powermarker V3.25 (Lui and Muse, 2004; http://www.powermarker.net). The program STRUCTURE V2.3.4 (Prichard et al., 2000; Falush et al., 2003) was used to infer population structure using a burn-in of 100 000, run length of 100 000, and a model which allowed for admixture and correlated allele frequencies. Structure analysis was undertaken for the samples overall and for each geographic region separately. The graphical display of STRUCTURE results was generated using DISTRACT software (Rosenberg 2002; http://www.cmb.usc.edu/noahr/distruct.html). Genetic diversity parameters, average number of alleles, gene diversity and polymorphism information content (PIC), were calculated using Powermarker. The presence of null alleles scoring errors and large allele drop-out was determined using Microchecker (Van Der Hoeven et al., 2000).

7.3.4. Environmental and morphological analysis

The GPS coordinates of each sampled tree were recorded along with tree height, tree diameter at 15 cm and 30 cm, and bole length. Environmental observations were made for each individual, including soil type and classification, soil pH, number and species of surrounding trees (potential hosts), distance to these trees from the sandalwood tree within a 15 m radius and their spatial arrangement relative to the sandalwood tree. Extensive ground cover was excluded from the analysis to simplify the dataset. Potential relationships between these observations and heartwood composition and yield
were investigated using a computer program script written in R. Individuals were classified based on broad abiotic categories of soil type that included 15 simplified groups: gravel, granite, sand, loam, sand over loam, sand over clay, red loam over clay, loam sand over gravel, loam sand over granite, loam over clay, loam over gravel, sand over granite, sand over gravel, loam over granite and loam over clay granite. For statistical analyses involving oil composition, percentage data was used.

7.4. Results

7.4.1. Oil composition

*Santalum spicatum* trees collected from natural stands in Western Australia that had been analysed previously for oil content and yield, displayed quantitative variation across the sampling region. The oil composition was based on 16 main sesquiterpenes including \( \alpha \)- and \( \beta \)-santalol and \( E,E \)-farnesol, which constitute the bulk of the essential oil (50-80\%). Total oil content was similar across all individuals; however sesquiterpene composition was highly variable. Trees from the arid northern region contained the highest levels of desirable fragrance compounds, \( \alpha \)- and \( \beta \)-santalol, as did individuals from the south-west. South-eastern populations contained greater amounts of \( E,E \)-farnesol, and were low in santalols. These trees generally also contained higher levels of \( \alpha \)-bisabolol. (Moniodis et al., submitted 2014).
### 7.4.2. Genetic Diversity and Structure

When individuals were grouped into small populations, tests of Hardy-Weinberg equilibrium showed that one locus SsB128 deviated significantly from HW expectations and was thus excluded from all analyses. The remaining 11 loci conformed to HW equilibrium except at one or two sites for SsA119, SsA105, SsB126 and SsA113, due to small population sizes. These loci were included since the aim was to determine the major differences between the three geographic regions, rather than conduct a detailed population study. Tests for linkage disequilibrium showed that no loci were significantly linked; therefore the 11 microsatellite primers were used for the remaining analyses. There was no consistent evidence of scoring errors, stuttering or large allele drop out across the 11 loci. There was evidence of homozygous excess for SsA119 in two populations, SsB126 in three populations, and SsC011 in one population, which could be due to null alleles or to sampling effects. The percentage of total missing data across all populations was less than 5%.

The mean observed heterozygosity over all loci was 0.65 (Table 7.1). The average number of alleles across the 11 loci was 16 (± 5.8) for pooled individuals; 13 (± 4.3) within the wheatbelt; 10 (± 4.6) within the goldfields and 11 (± 4.3) within the north-sampled trees. The arid north and goldfields trees had a higher proportion of rare alleles (56%) than south-wheatbelt trees, despite the small representation of trees from these regions. Mean observed heterozygosity across loci was almost identical between north (0.79) and goldfields (0.76) and lower in the wheatbelt region (0.66). An AMOVA revealed that there was minimal hierarchal genetic structure, with most of the molecular
variance occurring within individuals (78 %) or among individuals within sites (18 %), rather than being partitioned between the three regions (4 %).

Table 7.1 - Summary of genetic diversity estimates obtained with eleven nuclear microsatellites for *Santalum spicatum*. Results include combined individuals

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele Size</th>
<th>Major Allele</th>
<th>Number of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Frequency</td>
<td>Alleles</td>
</tr>
<tr>
<td>ssA103</td>
<td>188 - 238</td>
<td>0.16</td>
<td>23</td>
</tr>
<tr>
<td>ssA105</td>
<td>282 - 314</td>
<td>0.21</td>
<td>16</td>
</tr>
<tr>
<td>ssA109</td>
<td>152 - 194</td>
<td>0.17</td>
<td>22</td>
</tr>
<tr>
<td>ssA106</td>
<td>203 - 211</td>
<td>0.96</td>
<td>5</td>
</tr>
<tr>
<td>ssB122</td>
<td>259 - 319</td>
<td>0.24</td>
<td>26</td>
</tr>
<tr>
<td>ssB126</td>
<td>111 - 147</td>
<td>0.29</td>
<td>15</td>
</tr>
<tr>
<td>ssA001</td>
<td>145 - 183</td>
<td>0.19</td>
<td>16</td>
</tr>
<tr>
<td>ssA113</td>
<td>203 - 237</td>
<td>0.18</td>
<td>18</td>
</tr>
<tr>
<td>ssC011</td>
<td>239 - 255</td>
<td>0.49</td>
<td>8</td>
</tr>
<tr>
<td>ssA119</td>
<td>190 - 228</td>
<td>0.32</td>
<td>18</td>
</tr>
<tr>
<td>ssB011</td>
<td>240 - 268</td>
<td>0.20</td>
<td>14</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.31</td>
<td>16.45</td>
</tr>
</tbody>
</table>

There was little differentiation between north and goldfields trees ($F_{ST} = 0.001$), in comparison to a greater differentiation between wheatbelt and both north ($F_{ST} = 0.039$) and goldfields trees ($F_{ST} = 0.032$).

Analysis of all individuals in STRUCTURE revealed two major genetic clusters (K=2), comprising the arid (north and goldfields) and the semi-arid south (wheatbelt) region (Figure 7.2). Mean assignment probabilities within each cluster were between 0.05 and
0.70 in 85 – 95 % of individuals (mean probabilities: Cluster 1: 0.0569 +/- 0.001; Cluster 2: 0.8207 +/- 0.008). An increase in genetic differentiation was evident with increasing inter-population distance, with the most south-west populations showing high divergence to the north-trees compared to increasing admixture in populations geographically closer.

![Figure 7.2 - Assignment probability of individuals from the arid (north and goldfields) and semi-arid south (wheatbelt) region using STRUCTURE](image)

Figure 7.2 - Assignment probability of individuals from the arid (north and goldfields) and semi-arid south (wheatbelt) region using STRUCTURE

The STRUCTURE analysis was repeated within the two main clusters separately (Figure 7.3). Analysis of substructure of individuals in the arid (north and goldfields) regions showed support for two genetic clusters (K = 2), which produced the highest log likelihood scores (Evanno et al. 2005) and a clear separation of north and goldfields trees, which are geographically separate (Figure 7.3(a)). Mean assignment probabilities to each cluster were > 0.8 in 95 % of individuals (mean probabilities: Cluster 1: 0.9677 +/- 0.001; Cluster 2: 0.8738 +/- 0.008). Analysis of substructure of individuals from the semi-arid wheatbelt region also showed highest log likelihood support for two genetic clusters (K = 2); however there was no clear structure within the wheatbelt with all populations showing admixture for the two clusters (Figure 7.3b). Mean assignment
probabilities to each cluster were > 0.5 (mean probabilities: Cluster 1: 0.5232 +/- 0.001; Cluster 2: 0.5286 +/- 0.001).

Figure 7.3 - Assignment probability of individuals to each genetic cluster from (a) arid region including the north and goldfields (two distinct clusters) and (b) the semi-arid south west wheatbelt (two distinct clusters) using STRUCTURE

A two dimensional MDS ordination plot shows a distinct gradient with increasing geographic distance (Figure 7.4). There is no overlap between north and wheatbelt trees, which largely occupy a separate ordination space, but there is some overlap in the arid goldfields and semi-arid wheatbelt regions which are closer geographically. The MDS plot produced a stress value of 0.136, indicating a reasonable fit between ordination and data.
Individuals were grouped into small populations according to their geographic location. The population based phylogeny provides some evidence for genetic structure in *S. spicatum* (Figure 7.5a), showing separate clusters for north (‘santalol’ chemotypes), goldfields (‘bisabolol’ chemotypes) and wheatbelt trees, with support for nodes for the three regions greater than 80%. Wheatbelt populations were less resolved and no nodes had significant support greater than 50%. A dendogram based on chemical data shows a similar separation of the three regions (Figure 7.5b).
7.4.3. **Evaluation of heartwood-oil characteristics and genetic structure**

Mantel tests revealed a significant association between oil composition and genetic distance when all individuals were combined ($r = 0.129$, $P = 0.02$, Table 7.2).

Significant associations ($P < 0.05$) were observed for $\alpha$- and $\beta$-santalol content, the santalol to farnesol ratio and the sum of the 16 sesquiterpenes with both distance matrices, although $E,E$-farnesol did not show a significant correlation to either matrix (Table 7.2) suggesting a continuous distribution. An Adonis test revealed a significant difference in oil composition when regions were analysed separately ($P = 0.001$), which suggests differences in total oil composition across the three regions.

When only south-wheatbelt trees were analysed, mantel tests revealed no significant association between oil composition and genetic distance ($r = 0.003$, $P = 0.423$);
however there was a significant relationship to geographic distance ($r = 0.158$, $P = 0.001$). There was no association between individual sesquiterpenes and genetic distance, although there was a significant association between geographic distance and total santalol, $E,E$-farnesol content and the sum of the 16 sesquiterpenes (Table 7.2).

Table 7.2 - Mantel test results for oil composition and genetic and geographic distances for (a) all 186 individuals of *S. spicatum* from semi-arid wheatbelt and arid north and goldfields and (b) south wheatbelt

<table>
<thead>
<tr>
<th></th>
<th>Total α- and β- santalol</th>
<th>E,E- farnesol</th>
<th>Sum of 16 sesquiterpenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) All trees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oil Composition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r $\quad$ P</td>
<td>r $\quad$ P</td>
<td>r $\quad$ P</td>
</tr>
<tr>
<td>Geographic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Distance</td>
<td>0.275 $\quad$ 0.001**</td>
<td>0.366 $\quad$ 0.001**</td>
<td>n.s $\quad$ n.s</td>
</tr>
<tr>
<td>Genetic Distance</td>
<td>0.129 $\quad$ 0.02*</td>
<td>0.201 $\quad$ 0.02*</td>
<td>n.s $\quad$ n.s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.311 $\quad$ 0.001**</td>
</tr>
<tr>
<td>(b) South-Wheatbelt trees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Distance</td>
<td>0.158 $\quad$ 0.001**</td>
<td>0.121 $\quad$ 0.003**</td>
<td>0.08 $\quad$ 0.01*</td>
</tr>
<tr>
<td>Genetic Distance</td>
<td>n.s $\quad$ n.s</td>
<td>n.s $\quad$ n.s</td>
<td>n.s $\quad$ n.s</td>
</tr>
</tbody>
</table>
7.4.4. Evaluation of environmental and morphological features

There was considerable variation in the growing environment of *S. spicatum* across the study site. *Santalum spicatum* is found in woodland and low-open woodland forests where there are variations in host species and their density, soil composition, soil pH, rainfall and elevation. Sandalwood trees grew on a variety of soils from calcareous red earth soils found largely in the arid (north and goldfields regions) to loam, gravel, granite and sandy soils, and various combinations of these. The soil pH across ranged from 4.25 to 8.5, and the average soil pH from the semi-arid south-wheatbelt was lower (5.6 ± 0.6) than for arid trees (6.7 ± 0.6). Soil type was significantly associated to pH (P < 0.05). The number of trees in the arid region was much lower than the wheatbelt trees, although the sampling was random and across a large area and the same values were obtained for the standard deviations. Altitude varied considerably across the study site (from ~0 to 500 m). Each sandalwood tree had a unique spatial-arrangement of surrounding vegetation and potential suite of hosts. The host species from this study included a variety of *Acacia*, *Casuarina* and *Eucalypt* species, as well as ground cover shrubs and grasses. *Acacia* species are known to be short-lived so it is expected that this surrounding vegetation will change during the expected 100-year life-span of a sandalwood tree. Host species were more similar within regions (wheatbelt, goldfields, north) than between regions and provenances. In the wheatbelt particularly, distribution of *S. spicatum* trees was fragmented in regions where there was significant clearing for agriculture. Thus, many individuals appeared isolated from larger population groups, which occurred in more open woodland areas.
When all trees were combined, an Adonis test showed a significant association between oil composition and provenance ($P = 0.001$, Table 7.3), which was characterised based on geography as well as similarities in soil and climatic factors (rainfall and temperature). Significant associations were also found between provenance and santalol, $E,E$-farnesol content, the santalol to farnesol ratio and the sum of the 16 sesquiterpenes (Table 7.3). When trees from all regions were pooled, there was a significant association of oil composition to soil type ($P = 0.002$) but not to other environmental characters even though rainfall and altitude varied across the provenances (Table 7.3). Soil type was also significantly related to total santalol and $E,E$-farnesol content, and the santalol to farnesol ratio. There was no significant association of total oil profile to soil pH (Table 7.3), although a significant $P$ value was obtained when soil pH was tested against total santalol and $E,E$-farnesol content ($P = 0.04$). When only south-wheatbelt trees were analysed, significant associations to provenance were also detected. An Adonis test showed a significant association between provenance and oil composition ($P = 0.008$), but not to individual oil components (Table 7.3). Soil type was significantly associated to oil composition across the wheatbelt trees ($P = 0.03$), although not when individual components were tested. There was no significant association of total oil profile to soil pH (Table 7.3), although a marginally significant $P$ value was attained when soil pH was tested against $E,E$-farnesol content ($P = 0.05$). A slightly significant association was found between rainfall and oil composition as well as $E,E$-farnesol content (Table 7.3). Across the wheatbelt, there was no significant associations between altitude to oil composition or individual sesquiterpenes. However, a significant correlation was detected between the potential number of hosts and oil profile as well as the total $\alpha$- and $\beta$-santalol content.
Table 7.3 - Adonis test results showing P-values for oil composition and environmental and morphological features for (a) all 186 individuals of *S. spicatum* from semi-arid wheatbelt and arid north and goldfield and (b) south wheatbelt

(a) All trees

<table>
<thead>
<tr>
<th>Environmental</th>
<th>Oil Composition</th>
<th>Total α- and β-santalol</th>
<th>E,E-farnesol</th>
<th>Santalol:Farnesol</th>
<th>Sum of 16 sesquiterpene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provenance</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.01*</td>
<td>0.03*</td>
<td>0.001***</td>
</tr>
<tr>
<td>Potential host no.</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Main host species</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Altitude</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Rainfall</td>
<td>n.s</td>
<td>n.s</td>
<td>0.04*</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Soil type</td>
<td>0.002**</td>
<td>0.03*</td>
<td>0.01*</td>
<td>0.006***</td>
<td>n.s</td>
</tr>
<tr>
<td>Soil pH</td>
<td>n.s</td>
<td>0.04*</td>
<td>0.04*</td>
<td>n.s</td>
<td>n.s</td>
</tr>
</tbody>
</table>

(b) South wheatbelt trees

<table>
<thead>
<tr>
<th>Environmental</th>
<th>Provenance</th>
<th>Potential number of hosts</th>
<th>Main host species</th>
<th>Altitude</th>
<th>Rainfall</th>
<th>Soil type</th>
<th>Soil pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provenance</td>
<td>0.008**</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>0.01*</td>
<td>n.s</td>
<td>0.03*</td>
</tr>
<tr>
<td>Potential number of hosts</td>
<td>0.003**</td>
<td>0.002**</td>
<td>n.s</td>
<td>n.s</td>
<td>0.01*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main host species</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.05</td>
<td>n.s</td>
<td>0.02*</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Soil type</td>
<td>0.03*</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Soil pH</td>
<td>n.s</td>
<td>n.s</td>
<td>0.05</td>
<td>n.s</td>
<td>0.03*</td>
<td>n.s</td>
<td>n.s</td>
</tr>
</tbody>
</table>

Morphological

| Tree height   | n.s        | n.s                        | n.s             | n.s      | n.s      |
| Stem diameter (at 15 cm) | n.s | n.s | n.s | n.s | n.s |
| Stem diameter (at 30 cm) | n.s | n.s | n.s | n.s | n.s |
7.4.5. Evaluation of environmental and morphological features

Trees morphology was diverse across the study sites, which included differences in tree height and stem diameter. Tree height ranged from 170 to 1200 cm (mean = 440 ± 260 cm) with a stem diameter at 15 cm from the base of the tree between 10 and 43 cm (mean = 22 ± 6.7 cm) and stem diameter at 30 cm from the base ranged from 9 - 39 cm (mean = 18 ± 5.7 cm). Large standard deviations are indicative of the large amount of variation in the natural habitat and difficulties in reliably determining tree age. No significant association between oil composition and morphological features were identified in this study when either trees were combined or when south-wheatbelt trees were analysed separately (Table 7.3).

7.5. Discussion

This study combined data from heartwood-oil composition with neutral genetic markers and information on the local habitat to investigate whether oil type was associated with genetic structure and/or environmental variables. The pattern observed in essential oil profiles across the three regions was broadly reflected in the genetic profiles and with environmental variables. Oil composition in *S. spicatum* showed variation across the individuals and regions sampled. The ‘santalol’ chemotype was found in northern trees and a few wheatbelt populations that yielded relatively high levels of santalol, whereas the goldfields trees were differentiated by production of larger levels of α-bisabolol and lower santalol content. The previous chemotaxonomic investigation of *S. spicatum*
heartwood oil samples showed a continuous distribution of $E,E$-farnesol levels, whereas other components were more prevalent in certain regions (Moniodis et al., 2014, submitted).

### 7.5.1. Genetic structure and oil composition

Investigation of population structure using several approaches all supported the evolutionary divergence of the arid (north, goldfields) and the semi-arid (wheatbelt) populations in Western Australia. The North and goldfields trees showed similarities and were differentiated than the wheatbelt trees, despite the goldfields being geographically closer to the wheatbelt than to trees from the North. This would suggest that historical patterns have influenced the current genetic structure of *S. spicatum* rather than differences arising from spatial separation. The degree of differentiations between regions is influenced by both drift that increases genetic differences, and gene flow that reduces it. Greater similarity and some admixture also indicate some gene flow between the north-east wheatbelt trees and the goldfields, which is possible given their closer geographic range. *Santalum spicatum* is predominantly insect pollinated thus gene flow is likely to be low; however the woylie (*Bettongia penicillata*), a small mammal, and emus (*Dromaius novaehollandiae*) have been reported to aid in seed dispersal (Murphy et al. 2005). The pattern of genetic structuring is consistent with previous studies that revealed two main genetic clusters in the semi-arid south and arid north regions in Western Australia (Byrne et al., 2003a,b). In these two studies, which used nuclear and chloroplast RFLP markers, different genetic influences on arid and semi-arid regions led to suggestions of a more recent expansion of sandalwood in the
south wheatbelt, possibly from populations in the north that displayed a drift-gene flow equilibrium in the nuclear genome. The genetic differentiation among regions also supports the presence of two ecotypes based on morphological variation (Fox and Brand 1993).

Further analysis of north and goldfields trees using Bayesian clustering methods revealed genetic differences, suggesting that a further level of genetic structuring occurs within the arid region, although this may be confounded by the disjunct sampling in this study. While substructure was revealed within the arid trees, despite the small sample size, there was no detectable structure across the wheatbelt region, although populations were positioned based on their geographic position, reflecting a transition from the most northwest trees, through the goldfields to the most south-west area of the WA wheatbelt. The lack of structure in the wheatbelt is in agreement with the results of Byrne et al. (2003a, b) and supports the conclusion of recent expansion into this region.

Genetic diversity *S. spicatum* was moderate, although higher (both in heterozygosity and rare alleles) in the arid trees (north and goldfields) compared to the wheatbelt, even though sample sizes were much smaller for those regions. On average, diversity values are higher than a study on populations of *S. austrocaledonicum*, which had a similar range but lower overall average diversity (Bottin et al. 2005) even though that study sampled a larger number of trees across the New Caledonian archipelago where it is native. Unlike *S. spicatum*, *S. austrocaledonicum* is capable of clonal reproduction (Bottin et al. 2005) and this would be expected to reduce genetic variation.
The pattern observed in essential oil profiles across the three regions showed an association with the pattern of genetic structure for total oil composition and for the valuable sesquiterpenes, α- and β-santalol, and the santol:farnesol ratio. However, the lack of significant association for \( E,E \)-farnesol is surprising given its abundance in certain trees. The oil composition of trees from the north were more similar to some south-western wheatbelt trees despite being more similar genetically to trees in the goldfields. Oil composition of north and goldfields trees were different in their composition which is likely reflected in differences in their growing environments. When only wheatbelt trees were analysed, there was no significant association of oil composition with genetic distance, indicating the genetic drift is not influencing the chemotype profiles within this region.

The significant correlations between oil composition and genetic distance at the regional level but not a local level suggest a broad influence of environment rather than may be a result of drift or adaptation of oil phenotypes to local microenvironments, and further work is needed to decipher these differences. A study of \( S. australasianum \) that combined microsatellite data and heartwood oil composition, detected low or non-significant correlations between oil composition and genetic distance matrices, and concluded that drift is not likely to be the primary evolutionary force that has created differences in heartwood chemistry among and within island populations (Bottin et al. 2007). Another study on \( Teucrium polium \) L. also found that essential oil variation depends more on genetic makeup than the environment. (Djabou et al. 2012). Similarly, in \( S. spicatum \), our results support a stronger genetic control on oil production than drift, and as such requires more explicit investigation than neutral markers.
7.5.2. **Oil composition and environmental variables**

The pattern of essential oil variability in *S. spicatum* was significantly associated with provenance, indicating shared similarities in the environment (such as climate, temperature, and possibly some pathogens). Additionally, there was a significant correlation between the oil profile and geographic distance matrix when considering either the whole sample set or just wheatbelt trees, indicating terpenoid profiles are associated with geographic location. When considering the north, goldfields and wheatbelt trees, oil composition (α- and β-santalol, *E,E*-farnesol and the sum of the 16 sesquiterpenes) was significantly associated with provenance; however only total oil composition and the sum of the 16 components were significantly related when only wheatbelt trees were considered. This suggests more uniformity within regions and that some chemotypes might be geographically restricted. For example, pathogenic fungi or phytophagous organisms (termites, insects) may occupy certain habitats. If essential oil production constitutes a defence mechanism, the presence of certain species may trigger the expression of a set of genes that is different to neighbouring environments. Unlike *S. spicatum*, a genetic and chemical study on *S. austrocaledonicum* did not reveal congruence of oil profiles based on their geographic positions (Bottin et al. 2007). Other species have documented strong influences of the environment. For example in *Coridothymus capitatus* and *Satureja thymbra*, the terpenoid profile was significantly linked to the growing environment (Karousou et al. 2005). The authors found that plants of a single species growing in the same environment were more similar in their oil composition and could express a different terpenoid profile if grown in a different environment (Karousou et al. 2005). The combined results observed here seem to indicate that variation in *S. spicatum* chemotypes is probably subject in large part to
environmental control, associated in particular with the habitat where it grows. The ability to study this in detail is complicated by numerous factors, such as the complex host-parasitic interactions, the number of possible haustorial connections in the lifetime of a single tree (> 100 years), the age of the tree, the long time required for onset of oil production (> 10 years), the lack of knowledge of the precise role of terpenoid components in the heartwood, and the associated pathogens in each provenance. In the latter case, it would be extremely difficult to catalogue all possible triggers of oil production in the natural environment. If oil composition is linked to provenance or growing environment, then these provenances could be targeted to identify trees with superior oil composition to be used as seed source for improving plantation stock. Several trees with high quality oil have been identified in several provenances. Ideally, analysis of oil composition prior to obtaining seeds for future plantations would be desirable and given the difficulty of conducting field experiments, this would provide some information on heritability and environmental influences on essential-oil composition.

Additional environmental factors may drive selection of a particular oil phenotype such as climate (including temperature and rainfall availability), vegetation structure (host species and haustorial connections) and abiotic stressors. Results of this study also suggest other environmental features may be important influencers of oil composition. Australian soils are among the oldest, most nutrient poor soils with a variable composition across the vast landscape (Steffen et al. 2009). The significant association between soil type, classified based on physical characteristics, and oil composition indicates soil composition and associated variables may influence oil composition. This occurred at both the regional level and the local level as oil composition in wheatbelt
trees was also found to have a significant association with soil type, although the strong association of oil components at the regional level was not present within the wheatbelt region. In particular, soil pH showed an association with particular oil components, at the regional level and for the wheatbelt trees. Soil pH and the availability of divalent metal ions such as Mg$^{2+}$, Mn$^{2+}$, Ni$^{2+}$ and Co$^{2+}$ have been reported to affect the production of sesquiterpenes in plants (Duarte et al. 2010), and could thus be influencing chemical variability in *S. spicatum*. In contrast, in a study on *S. austrocaledonicum*, no significant association was found between soil composition and oil profile (Bottin et al. 2007). Elemental examination of soil composition would allow further inferences about the association of soils and sesquiterpene production to be made.

The sandalwood-host parasitic relationship could be a source of additional selection pressure. The essential oil profile showed a significant association with the number of potential hosts or vegetation density surrounding each sandalwood tree within 15m, although, this does not account for all host species, which can reach 30 m (Woodall and Robinson 2003). Sandalwood establishment, survival and development depends on the species of host and the nitrogen fixing capacity (Radomiljac et al. 1999). Many nitrogen fixing plants, such as *Acacia* and *Casuarina* species that are the main sandalwood hosts in the wheatbelt region, constitute a major part of vegetation forms in Australia (Steffen et al. 2009). In *S. album*, nitrogen fixing hosts have been reported to influence foliar nitrogen concentrations, which were significantly higher when parasitised to nitrogen-fixing hosts than on eucalypt, or with no host. Strong positive relationships were detected between foliar N concentration, rates of net photosynthesis and water use efficiencies (Radomiljac et al. 1999). If host species can influence foliar constituents, it
is not unreasonable to assume that the species of host and the extent of haustorial connections could influence the biosynthesis of terpenoids in the heartwood. The contribution of hosts to oil production would be very difficult to test. Even in a controlled plantation experiment, variation would also exist in the haustorial connectivity and in the genetic composition of host species. Other variables in this study, such as rainfall, were only marginally significant in wheatbelt trees.

7.5.3. Oil composition and morphological variables

No morphological measures, including stem diameter or tree height, showed any association with oil composition. These measures would also relate to tree age, which is difficult to estimate in wild *S. spicatum* trees, but may also contribute to total variability in terpenoid profiles. The source of chemotype variation in *S. spicatum* is difficult to ascertain due to variation in biotic and abiotic factors. In addition to any genetic influences, there appears to be a selection pressure that strongly relates to the environment (soil, hosts, pathogens, phenotype plasticity), and thus requires further investigation.

7.5.4. Management of Sandalwood

*Santalum spicatum* is currently harvested from natural stands. The establishment of plantations will ensure a continued supply into the future as worldwide demands continue to grow and enable conservation of the natural resource. Plantations alleviate
the need to harvest from natural populations and conservation efforts are needed as natural resources continue to decline. Plantations in Western Australia provide additional environmental benefits. Clearing for agriculture has significantly contributed to problems of soil salinisation, erosion and loss of species biodiversity, which has impacted local flora and fauna communities (Tonts and Selwood 2003; Woodall and Robinson 2003). Plantations of sandalwood and host species therefore provide a diverse habitat for wildlife, and improve hydrological balance and ameliorate soil salinity through establishment of deep-rooted perennial species (Woodall and Robinson 2003). Additional information on the environmental and genetic contributions to essential oil variability warrants further investigation in order to improve product quality during the plantation rotation of 25 years.

Currently, little is known about the heritability of oil phenotypes and how to predict responses to selection pressures. A better knowledge of the heritability of oil phenotypes will be a long-term process that will require comparisons of trees grown in a uniform environment over multiple generations (Falconer 1981). These requirements will make studying chemical variability in the field very difficult for sandalwood, given its hemi-parasitic nature and the long time for the onset of heartwood-oil (> 10 years). Alternatively, sequence information on the biosynthetic genes involved in oil production and regulation may be a more realistic approach in determining the genetic contributions to oil variation (Jones et al. 2008, Jones et al. 2011, Diaz-Chavez et al. 2013). Neutral markers have allowed detection of genetic structure in S. spicatum and thus provided information for cultivation and conservation management. In the future technologies such as restriction site-associated DNA (RAD) sequencing (Baxter et al. 2011), or marker-based quantitative genetics and linkage studies (Andrew et al. 2005),
which aim to improve the understanding of adaptive evolution and identify key regions in the genome that control known phenotypes, could be used to study the heritability and genetic relationship of essential oil variability in *Santalum*. In practical terms, this study provides the basis for managing plantations of *S. spicatum* to improve their productivity, as well as improving knowledge of natural stands to aid conservation efforts.

7.6. Acknowledgements

The authors are grateful for financial assistance from the Australian Research Council and the Forest Products Commission of Western Australia through linkage project LP0882690. The authors also wish to thank Len Norris for his assistance on field trips and Bronwyn McDonald and Shelley McArthur for technical support for genetic analyses.
7.7. References


Moretta P (2001) Extraction and variation of the essential oil from Western Australian sandalwood (*Santalum spicatum*). Perth: University of Western Australia.


R, the Core R Team (2013) A language and environment for statistical computing. R
Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-
project.org/.

Radomiljac A, McComb JA, McGrath J (1999) Intermediate host influences on the root
hemi-parasite *Santalum album* L. biomass partitioning. For Ecol Manage 113:
143-153.

Radomiljac A, McComb JA, Pate JS (1999) Gas exchange and water relations of the
root hemi-parasite *Santalum album* L. in association with legume and non-

CE, editors. Proceedings of the symposium on sandalwood in the Pacific April


Rosenberg NA (2004) DISTRACT: a program for the graphical display of population

MS, Werner PA (2009) Australia's biodiversity and climate change. Australia:
CSIRO publishing.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees
from microsatellite DNA. Genetics 144: 389-399.

Tommasi L, Negro C, Cerfeda A, Nutricati E, Zuccarello V, De Bellis L, Miceli A
(2007) Influence of Environmental Factors on Essential Oil Variability in
*Thymbra capitata* (L.) Cav. Growing Wild in Southern Puglia (Italy). J Essent


CHAPTER EIGHT

Final Conclusions

8.1. General Conclusions

The focus of this thesis was to provide a more detailed investigation into sesquiterpene biosynthesis in sandalwood as well as investigate the natural chemical diversity, genetic structure and origins of essential oil variation in *S. spicatum* across its' natural distribution. *Santalum spicatum* is a small but significant industry in Western Australia which contributes to the economy of the state, benefiting rural farmers as well as providing environmental benefits for vegetative regions suffering from problems of soil salinity and loss of biodiversity. Plantations of *S. album* in the North of Western Australia will also be highly valuable in the future. There is potential to further develop the sandalwood industry. The timber and oil is highly sought after for its' unique aroma, with additional medicinal properties which can be used to expand the sandalwood industry. The opportunity to diversify the market for sandalwood products, address environmental concerns in the agricultural (wheatbelt) regions of WA and provide extra income for rural farmers, highlights the importance of furthering research into tree improvement programs for sandalwood species.
The work presented in this thesis has improved the current understanding of sesquiterpene production in sandalwood. A functional and genomics approach has identified key enzymes in the biosynthetic pathway of sesquiterpene production. Four key terpene synthases have been characterised including a santalene synthase, bisabolene synthase, bulnesol/guaiol synthase [1] and a sesquisabinene synthase (This thesis, Chapter 6). Products from the four enzymes may explain up to one quarter of the total oil content in *S. spicatum*, with quantitative variation expected to exist across the natural range (This thesis, Chapter 4). Sequences for additional enzyme classes which may contribute to oil production and regulation in *S. spicatum* were also identified including cytochrome P450, phosphatases and genes in the MEV and MEP pathways. Future exploitation of *S. spicatum* should include identification of *in situ* farnesol biosynthesis, as this component is still poorly understood and greatly influences overall oil-quality. Promoter sequences, transcription factors and differential expression patterns may provide additional clues into oil production and regulation in *S. spicatum*.

Additional progress was made towards understanding oil production in *Santalum* through functional characterisation of recombinant P450s in the CYP76F subfamily producing the valuable Santalols and bergamotol [2]. A full length sequence which was 99% identical (at the amino acid level) to the CYP76F39v1 from *S. album* was also isolated from the *S. spicatum* 454 library (Thish Thesis, Chapter 6), indicating the same oxidative role. Both the santalene synthase and santalene hydroxylase appear are highly conserved in sequence across the genus, which is in contrast to the guaiol/sesquiterpene synthases [1]. The functional conservation of santalol biosynthesis may suggest a positive selection for santalene/santalol production in *Santalum* and possibly in the Santalaceae family. Further work will focus on the characterisation of additional sequences in the CYP76F family which may be more
stereoselective in producing the Z-isomers of α- and β-santalol rather than E-isomers which are not abundant in the native oil.

The work demonstrated in the chapters of this thesis highlights the potential for metabolically engineering microbes to produce the valuable fragrance ingredients of limited availability in the plant. The santalene synthase, CYP76 and CPR cDNAs provide a biotechnology opportunity for in vitro production of the valuable oil-components to supplement the growing sandalwood industry. This has been successfully demonstrated on a small-scale in yeast and bacterial systems using sandalwood genes (This thesis, Chapters 3 and 5) [1,2]. The terpene synthase and P450 sequences from Santalum can also be used to explore potential biomarkers for the onset of oil formation. Although substantial progress has been made towards understanding terpenoid biosynthesis in Santalum, very little is known about specific processes involved in the onset and regulation of oil production, which is hampered by the late commencement of oil formation in the heartwood (> 7 years). These results are of great interest to the flavour and fragrance market and the growing plantation industry of Santalum. The results presented in this thesis provide a platform for further metabolic engineering opportunities to improve yields and alter product specificity across the genus (through site-directed mutagenesis for example).
8.2. Chemodiversity

There is good potential to improve oil quality of *S. spicatum* trees in order to introduction better plant material into plantations and breeding programs. Currently, the variable nature of heartwood-characteristics makes it difficult to predict quality prior to harvesting. In order to better understand the basis of compositional variation in wild *S. spicatum* trees, a chemotaxonomic study was conducted across its natural distribution.

Trees from the Northern, South-Eastern (goldfields) and South-Western (wheatbelt) regions were investigated for oil quality, as well as any co-occurrence patterns between sesquiterpenes. Results presented in this thesis indicated that no significant difference in the total overall oil was evident across the three regions. Trees from the north were found to contain the highest levels of the desired α- and β-santalol as did several individuals from the wheatbelt where harvesting practices still occur. Trees from the arid (goldfields) region had the highest amount of the allergenic *E,E*-farnesol, and less total santalol content. These trees generally also contained higher levels of α-bisabolol. The existence of α-santalol, *E,E*-farnesol and α-bisabolol chemotypes highlight the possibility for selection programs for tree improvement. Also apparent from the oil analysis was the co-occurrence of *E,E*-farnesol and dendrolasin (and to a lesser extent the acyclic nerolidol and farnesene), which suggests a shared biogenesis. Cluster analysis indicated that at least four separate enzymes were involved in production of major constituents in sandalwood oil. Future work will focus on specific genetic and environmental factors which influence the onset of oil to know whether this can be manipulated *in planta* to increase expression of certain genes or initiate the early production of heartwood oils which may ultimately lead to reduced rotation times and ultimately increase the market value of *S. spicatum* oil.
8.3. Genetic and environmental influence on oil quality

To help determine whether oil profiles were genetically or environmentally influenced, variability in heartwood oil composition was compared to variation patterns in molecular genetic data, morphological differences and environmental features. Results from this thesis have shown that essential oil composition in *S. spicatum* is influenced by both genetics (probably to a lesser extent) and by selection pressure arising from the growing environment. Most chemical variation was seen across rather than within provenances. Essential oil composition is likely to be influenced by factors such as tree-age, climatic differences, specific pathogens or insects which may trigger production of certain oil components. Improved management of plantations and natural stands is enabled through continued research into the genetic and environmental factors which control total oil production. Future research will focus on the contribution of the host species to the onset of oil, physiological processes required for heartwood initiation and development, stimulants which can trigger oil production, and employing new technologies that will further link oil phenotypes to the natural habitat.
8.4. References
