J Chem Ecol (2010) 36:652–661
DOI 10.1007/s10886-010-9798-y

A Biochemical Interpretation of Terpene Chemotypes in *Melaleuca alternifolia*

Andras Keszei · Yasmin Hassan · William J. Foley

Received: 19 November 2009 / Revised: 24 March 2010 / Accepted: 29 April 2010 / Published online: 29 May 2010

Abstract The variation of foliar monoterpenes in the Australian Tea Tree (*Melaleuca alternifolia*) has been of significant interest both to the essential oil industry as well as to ecologists. The majority of studies on leaf chemistry have been aimed directly towards obtaining oil of higher quality or quantity. In the current study, we aimed to understand how molecular mechanisms contribute to the chemical variability of this species, based on chemical analysis of the leaf oils from a biochemical perspective. Correlations between monoterpenes across the species as well as within chemotypes show strong, persistent patterns, which enable us to establish groups based on possible common biosynthetic origins. We found that three distinct enzymes corresponding to these groups: a sabinene-hydrate synthase, a 1,8-cineole synthase, and a terpinolene synthase may be sufficient to explain all six chemotypes in *M. alternifolia*.

Key Words *Melaleuca alternifolia* · Australian Tea Tree · Chemotypes · Terpene synthase (TPS) · Terpinen-4-ol · Cineole · Terpinolene · Barycentric plot · Viridiflorene · Aromadendrene · Calamenene

Introduction

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

Distinct intra-specific chemical variation in foliar terpenes can occur in natural populations (Vernet et al. 1986) and is widespread in many Australian plants of the family Myrtaceae (Keszei et al. 2008). *Melaleuca alternifolia* is a Myrtaceous tree that shows chemotypic variation of foliar terpenes throughout its range (Butcher et al. 1994). The chemotype rich in terpinen-4-ol is widely used medicinally (Cox et al. 2001; Hammer et al. 2006), and therefore, a lot of research in the species has been restricted to the one chemotype. The presence of chemotypes in Tea-Tree indicates that the foliar leaf oil traits may be the result of adaptation to complex biotic interactions, as those found to influence chemical variability in thyme (Linhart and Thompson 1999; Linhart et al. 2005). To gain a similar understanding of the well-studied system that Australian Tea-Tree represents, we need to focus on leaf oil in light of its variability and biosynthetic origin, and not as a raw product.

Melaleuca alternifolia contains foliar oils that are a complex mixture of mono- and sesquiterpenes (Brophy et al. 1989). Terpene synthases are responsible for the direct production of the majority of the terpenes found in essential oils. They are capable of producing single or multiple compounds from the same prenyl-diphosphate substrate (Degenhardt et al. 2009). Examining correlations between groups of terpenes and working under the assumption that

A. Keszei (✉) · Y. Hassan · W. J. Foley
Research School of Biology, Australian National University, Canberra 0200, Australia
e-mail: andras.keszei@anu.edu.au

strong positive correlation indicates a common biogenetic origin, Zavarin (1970) predicted the presence of a single terpene synthase in *Picea* that catalyzes the formation of sabinene, terpinolene, and γ-terpinene in a set ratio. An enzyme was isolated later that encodes a terpene synthase that showed the predicted catalytic profile (Gijzen et al. 1991). Qualitative variation in plant terpenes likely is due to variation in terpene synthase genes, and may be deduced from identifying the correlations among compounds that potentially have the same biosynthetic origin.

In *M. alternifolia*, previous work has established chemotypes based on variation in the concentration of 1,8-cineole, terpinen-4-ol, and terpinolene (Butcher et al. 1992). *Melaleuca alternifolia* is restricted to coastal flood plains of north-eastern New South Wales (NSW) with neighboring populations in the 'granite belt' in southeast Queensland, Australia. The geographic distribution of the three major monoterpenes reveals terpinen-4-ol in the majority in populations in the center of the natural range, while terpinolene dominates in the northwest, and 1,8-cineole in the south. Butcher et al. (1994) have shown that this corresponds to co-occurrence with the sister species *M. linariifolia* in the south and *M. trichostachya* in the north of NSW. *Melaleuca linariifolia* also is characterized by having both terpinen-4-ol and 1,8-cineole rich chemotypes, while *M. trichostachya* has terpinolene and 1,8-cineole rich chemotypes. It has been proposed that the respective *M. alternifolia* chemotypes may be a result of past hybridization. However, it also is possible that the convergence of chemical traits is due to similar patterns of herbivore pressure where the species geographically co-occur.

The purpose of the current work was to identify correlations and chemical patterns that allow us to predict the biosynthetic origin of chemotypic variation in *M. alternifolia* terpenes. This in turn will facilitate a search for the molecular components involved, and ultimately an understanding of the forces and mechanisms that have brought about and maintain chemical variation in Australian Tea-Tree.

Methods and Materials

Population Sampling Mature leaf material (ca. 20 g wet mass) was collected from 20 mature trees at eight sites and 15 from a ninth site across the species' natural geographic range (Table 1). However, populations from the far south of the range (Port Macquarie) were excluded, because they have been proposed to be hybrids with southern lineages of *M. linariifolia* (Butcher et al. 1995). We chose trees that were at least 100 m apart to avoid collecting from related trees (Rossetto et al. 1999), and the location of each tree was recorded. Samples were refrigerated at 4°C within 2 hr of collection. The extraction of leaf oils was completed on the day of collection in order to minimize the possibility of evaporation and oxidative degradation.

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

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

[a] Sites in New South Wales (NSW), Australia; for further information and a map of the sites, see Rossetto et al. (1999)

Extraction and Analysis of Terpenes Mature leaf material was separated from stems, weighed, and placed in 10 ml of ethanol (Baker et al. 2000) containing 0.25 $g \cdot l^{-1}$ of *n*-tridecane as an internal standard. The samples were left in ethanol for 7 d to complete extraction. A further 4 g leaf material from each tree were oven-dried at 60°C and weighed to calculate oil concentration on a dry weight basis.

Gas chromatography was carried out on an Agilent 6890 GC using an Alltech AT-35 (35% phenyl, 65% dimethylpolyoxylane) column (Alltech, Wilmington, DE, USA). The column was 60 m long (internal diam. 0.25 mm) with a stationary phase film thickness of 0.25 μm. Helium was used as carrier gas. The ethanol extract was filtered through a 0.45 μm filter, and 1 μl was injected at 250°C at a 1:25 split ratio. The temperature program was as follows: 100°C for 5 min, ramping to 200°C at $20°C \cdot min^{-1}$, followed by a ramp to 250°C at $5°C \cdot min^{-1}$, and held at 250°C for 4 min. The total elution time was 25 min. The components of the solvent extract were detected by using an FID and an Agilent 5973 Mass Spectrometer dual setup through an SGE MS/FID splitter. Several peaks were identified tentatively by comparisons to authentic standards (α-pinene, β-pinene, β-myrcene, β-phellandrene, terpinen-4-ol, 1,8-cineole, α-terpineol, β-caryophyllene, kindly made available by Mike Lacey and Thomas Wallenius, CSIRO Canberra). The remaining peaks were identified by comparison to mass spectra and retention times of compounds published using the same AT35 chromatographic column and run parameters (Southwell and Lowe

1999; Russell and Southwell 2002). Quantities of compounds were calculated by recording peak areas of the FID trace in MSD Chemstation Data Analysis (Agilent Technologies, Deerfield, IL, USA) and relating them to the internal standard (tridecane) and leaf dry weight.

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

Results

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

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

Pearson's correlation values among concentrations of all of the monoterpenes within the full sample set established three groups showing strong intercorrelations ($r>0.8$): **Group A**: α-thujene, α-terpinene, γ-terpinene, terpinen-4-ol, **Group B**: α-pinene, β-pinene, myrcene, limonene, 1,8-cineole, α-terpineol, and **Group C**: α-phellandrene, terpinolene, linalool.

Common biosynthetic intermediates (Fig. 1) suggest that these three main groups may be the products of three distinct terpene synthases. Concentrations of *(Z)*-piperitol, *p*-cymene, β-phellandrene, sabinene, and *(Z)*-sabinene hydrate did not show strong ($r<0.8$) correlations to any other group, however, they showed the strongest correlation to group A. As each of these compounds readily undergoes post-biosynthetic conversion or is the product of such a process (Bohlmann et al. 1999; Keszei et al. 2008), their concentration may not directly reflect enzymatic activity. As such, the lack of a strong positive correlation need not rule out biosynthetic relatedness. Provided that there was either a strong statistical correlation or evidence of a common biosynthetic origin, monoterpenes were assigned to one of the three groups. The variation in the proportions of these three groups relative to the total monoterpene pool is discontinuous, and clearly resolves all chemotypes (Fig. 2).

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

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

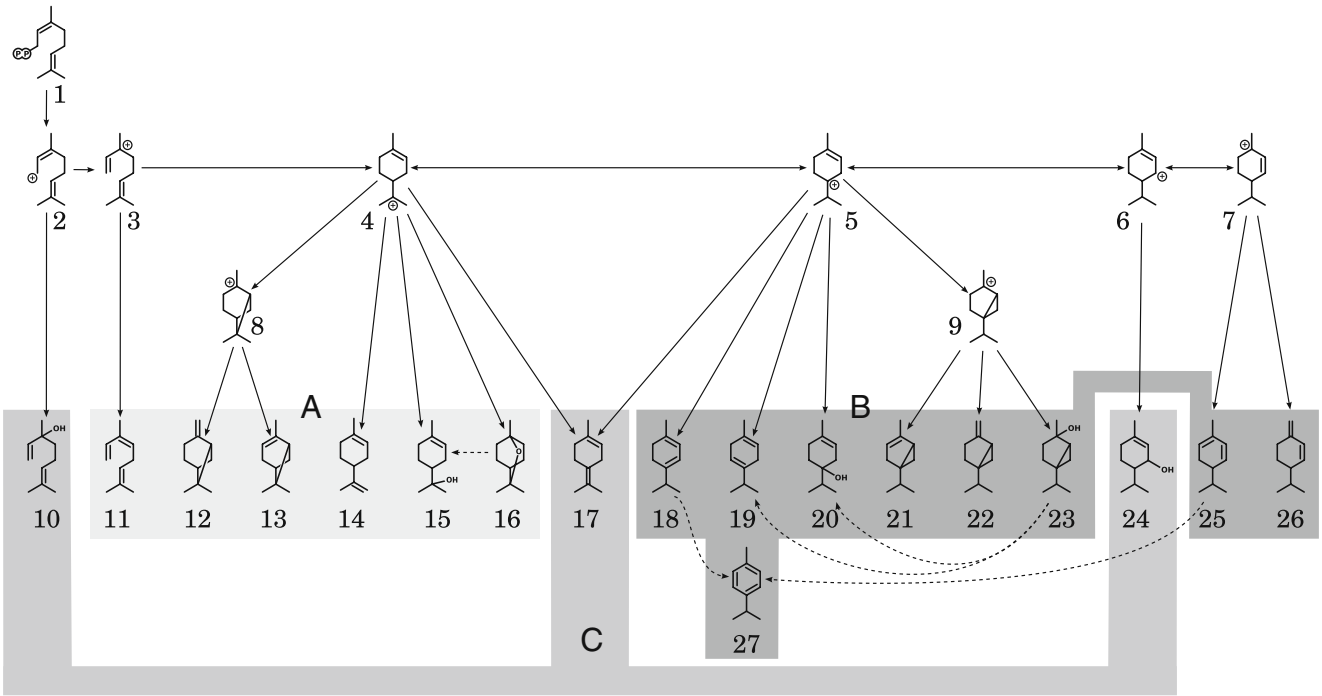


Fig. 1 Biosynthetic relationships between monoterpenes found in *Melaleuca alternifolia*. The products of the three proposed terpene synthases are indicated by shaded boxes. 1: geranyl diphosphate, 2: linalyl carbocation, 3: geranyl carbocation, 4: α-terpinyl carbocation, 5: terpinenyl carbocation, 6: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

correlation matrices to separate monoterpenes of different biosynthetic origins into separate groups.

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

In chemotype 1, dominated by terpinen-4-ol, two groups emerge: Group 1A: α-thujene, a-terpinene, γ-terpinene, terpinolene, terpinen-4-ol, *(Z)*-piperitol; and Group 1B: limonene, 1,8-cineole. Concentrations of myrcene, α-pinene, β-pinene, α-terpineol, α-phellandrene, and β-phellandrene show significant correlation to compounds in both groups, but they show somewhat stronger correlation to Group 1A. Concentration of linalool, *p*-cymene, sabinene and sabinene hydrate show no significant correlations to other compounds.

In the terpinolene dominated chemotype 2, two strong groups can be distinguished: Group 2C: α-thujene, α-phellandrene, α-pinene, α-terpinene, γ-terpinene, terpinolene, linalool; Group 2B: α-pinene, myrcene, limonene, β-phellandrene, 1,8-cineole, α-terpineol, γ-terpinene, terpinen-4-ol. Concentrations of β-pinene, *p*-cymene, sabinene, and sabinene hydrate show no significant correlation to other compounds.

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

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

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

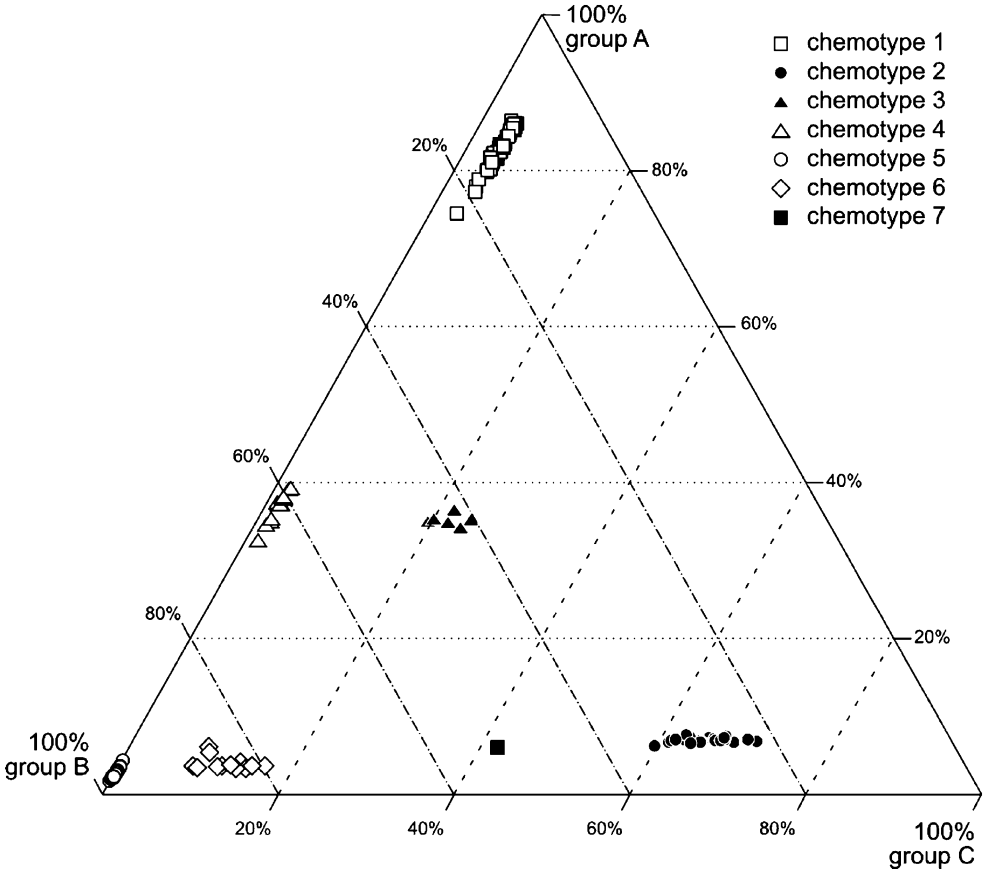


Fig. 2 A ternary barycentric plot showing the contributions of the three biochemical groups of compounds (as shown in Fig. 1) to the monoterpene fraction of the leaf oil profile of *Melaleuca alternifolia*. All six chemotypes are resolved and show discontinuous distributions. The intermediate chemotypes (3,4 and 6) are positioned between the cardinal chemotypes (1,2 and 5), which confirms a codominant relationship between the genetic components involved

allo-aromadendrene, viridiflorene, globulol, viridiflorol, isoaromadendrene epoxid e, unknown peak at 15.11 min, β-gurjunene; Group F: α-elemene, γ-muurolene; Group G: 13.79 min, δ-cadinene, β-cadinene, γ-cadinene, cubenol, (also showed strong correlation to α-gurjunene and γ-muurolene). None of the other sesquiterpenes showed significant correlation to any other compounds.

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

Geographic Distribution of Chemotypes Variation of chemotypes was observed both within and among sites. Two populations contained only a single chemotype (chemotype 1), while two sites (Chaffin Swamp and Bald Rock Creek) contained 5 of the 7 chemotypes (Table 1). Interestingly, these two populations also contained some of the extremes of oil yield. Populations were dominated by either chemotype 1, 2, or 5, while other chemotypes occurred at lower frequencies. There also was a distinct chemical separation of trees across the geographic range with respect to these former three chemotypes. Approximately 70% of trees sampled at both Queensland sites were chemotype 2, while in the Richmond River valley in northern NSW 95% of trees sampled chemotype 1, and more than 80% of trees along southern tributaries of the Clarence River system were chemotype 5 (Table 1).

Discussion

Monoterpene Chemotypes We identified six chemotypes in this study, all of which had been described previously (Butcher et al. 1994), and an individual tree that did not group with any of the recognized chemotypes. We grouped the monoterpenes on the basis of their likely biosynthetic

Table 2 Pearson's pairwise correlation coefficients between individual terpene concentrations (mg·g^{-1} dry leaf weight) from all individuals of *Melaleuca alternifolia*

	β-gurjunene	guaiene	12.85	12.92	α-gurjunene	β-caryophyllene	aromadendrene	13.59	α-elemene	13.78	alloaromadendrene	γ-muurolene	viridiflorene	bicyclogermacrene	δ-cadinene	β-cadinene	γ-cadinene	calamenene	15.11	globulol	isoaromadendrene epoxide	viridiflorol	α-eudesmol	cubenol	β-eudesmol	
elixene	0.19	0.41	0.42	0.01	0.59	0.61	0.17	0.17	0.76	0.67	0.73	0.66	0.54	0.91	0.73	0.61	0.71	-0.06	0.06	0.33	0.33	0.53	0.33	0.48	-0.05	elixene
β-gurjunene		0.74	-0.11	0.62	0.64	0.37	0.75	0.77	0.37	0.40	0.62	0.38	0.75	0.14	0.47	0.36	0.47	0.49	0.70	0.70	0.62	0.57	0.34	0.52	0.18	β-gurjunene
guaiene			0.09	0.60	0.92	0.66	0.87	0.82	0.65	0.63	0.79	0.69	0.90	0.34	0.77	0.73	0.80	0.61	0.77	0.84	0.73	0.65	0.41	0.73	0.18	guaiene
12.85				-0.07	0.15	0.26	-0.08	-0.04	0.33	0.37	0.21	0.31	0.08	0.55	0.24	0.23	0.24	-0.16	-0.08	0.00	-0.03	0.09	0.09	0.07	-0.01	12.85
12.92					0.42	0.18	0.66	0.67	0.16	0.22	0.32	0.15	0.52	-0.05	0.26	0.09	0.26	0.54	0.68	0.41	0.28	0.24	0.22	0.29	0.23	12.92
α-gurjunene						0.78	0.82	0.77	0.83	0.73	0.87	0.83	0.93	0.54	0.85	0.82	0.88	0.41	0.66	0.82	0.69	0.70	0.43	0.77	0.16	α-gurjunene
β-caryophyllene							0.49	0.46	0.81	0.79	0.67	0.77	0.66	0.59	0.79	0.77	0.80	0.18	0.38	0.54	0.48	0.54	0.32	0.63	0.03	β-caryophyllene
aromadendrene								0.94	0.46	0.40	0.69	0.47	0.86	0.11	0.48	0.46	0.53	0.60	0.86	0.89	0.71	0.57	0.38	0.57	0.31	aromadendrene
13.59									0.44	0.42	0.64	0.46	0.83	0.11	0.46	0.42	0.49	0.56	0.86	0.81	0.63	0.53	0.32	0.54	0.30	13.59
α-elemene										0.79	0.73	0.93	0.71	0.77	0.85	0.82	0.87	0.02	0.30	0.51	0.40	0.57	0.34	0.68	-0.05	α-elemene
13.78											0.69	0.75	0.66	0.68	0.82	0.72	0.81	0.19	0.32	0.44	0.37	0.51	0.27	0.60	0.04	13.78
alloaromadendrene												0.68	0.91	0.69	0.79	0.70	0.79	0.36	0.52	0.83	0.80	0.85	0.46	0.71	0.10	alloaromadendrene
γ-muurolene													0.70	0.65	0.84	0.86	0.88	0.04	0.31	0.52	0.43	0.53	0.29	0.69	-0.02	γ-muurolene
viridiflorene														0.50	0.77	0.66	0.78	0.49	0.71	0.87	0.77	0.76	0.47	0.75	0.23	viridiflorene
bicyclogermacrene															0.68	0.53	0.64	-0.17	0.00	0.27	0.27	0.52	0.28	0.43	-0.12	bcg
δ-cadinene																0.89	0.98	0.33	0.37	0.56	0.54	0.65	0.37	0.78	0.03	δ-cadinene
β-cadinene																	0.93	0.27	0.32	0.59	0.55	0.59	0.31	0.72	-0.02	β-cadinene
γ-cadinene																		0.34	0.41	0.60	0.55	0.64	0.39	0.79	0.06	γ-cadinene
calamenene																			0.63	0.52	0.51	0.35	0.30	0.41	0.28	calamenene
15.11																				0.72	0.54	0.44	0.31	0.49	0.35	15.11
globulol																					0.92	0.79	0.39	0.63	0.17	globulol
isoaromadendrene epoxide																						0.85	0.34	0.56	0.03	isoaromadendrene epoxide
viridiflorol																							0.44	0.63	-0.03	viridiflorol
α-eudesmol																								0.53	0.42	α-eudesmol
cubenol																									0.30	cubenol
β-eudesmol																										β-eudesmol
α-pinene	-0.51																									α-pinene
sabinene	0.26	0.08																								sabinene
β-pinene	-0.71	0.82	0.03																							β-pinene
myrcene	-0.63	0.94	0.12	0.93																						myrcene
α-phellandrene	0.35	-0.53	-0.14	-0.21	-0.48																					α-phellandrene
α-terpinene	0.77	-0.16	0.30	-0.60	-0.35	-0.30																				α-terpinene
limonene	-0.80	0.80	-0.06	0.97	0.91	-0.24	-0.67																			limonene
β-phellandrene	0.03	0.65	0.28	0.57	0.66	-0.13	0.11	0.45																		β-phellandrene
p-cymene	0.36	-0.12	-0.04	-0.46	-0.33	-0.24	0.63	-0.48	-0.12																	p-cymene
1,8-cineole	-0.82	0.77	-0.08	0.96	0.90	-0.22	-0.71	1.00	0.41	-0.51																1,8-cineole
γ-terpinene	0.74	-0.14	0.26	-0.58	-0.34	-0.31	0.99	-0.66	0.11	0.71	-0.69															γ-terpinene
terpinolene	0.32	-0.58	-0.16	-0.24	-0.52	0.99	-0.32	-0.27	-0.21	-0.24	-0.25	-0.33														terpinolene
linalool	0.26	-0.54	-0.18	-0.19	-0.48	0.98	-0.38	-0.22	-0.21	-0.27	-0.19	-0.39	0.99													linalool
(Z)-sabinene hydrate	0.25	-0.01	0.75	-0.08	-0.03	-0.06	0.25	-0.14	0.05	0.02	-0.15	0.22	-0.07	-0.08												(Z)-sabinene hydrate
terpinen-4-ol	0.71	-0.12	0.26	-0.58	-0.32	-0.35	0.99	-0.65	0.11	0.71	-0.68	1.00	-0.37	-0.43	0.21											terpinen-4-ol
α-terpineol	-0.70	0.90	0.04	0.94	0.97	-0.41	-0.46	0.96	0.55	-0.34	0.94	-0.44	-0.45	-0.41	-0.08	-0.43										α-terpineol
(Z)-piperitol	0.79	-0.20	0.45	-0.55	-0.36	-0.12	0.89	-0.65	0.13	0.53	-0.68	0.88	-0.13	-0.18	0.35	0.87	-0.46									
	α-thujene	α-pinene	sabinene	b-pinene	myrcene	α-phellandrene	α-terpinene	limonene	β-phellandrene	p-cymene	1,8-cineole	g-terpinene	terpinolene	linalool	(Z)-sabinene hydrate	terpinen-4-ol	α-terpineol									

Monoterpenes (bottom) and sesquiterpenes (top). Values above 0.7, 0.8, and 0.9 are emphasized using incremental shading. Strong positive correlations (>0.7) between monoterpenes correspond to grouping based on common carbocation intermediates as shown in Fig. 1

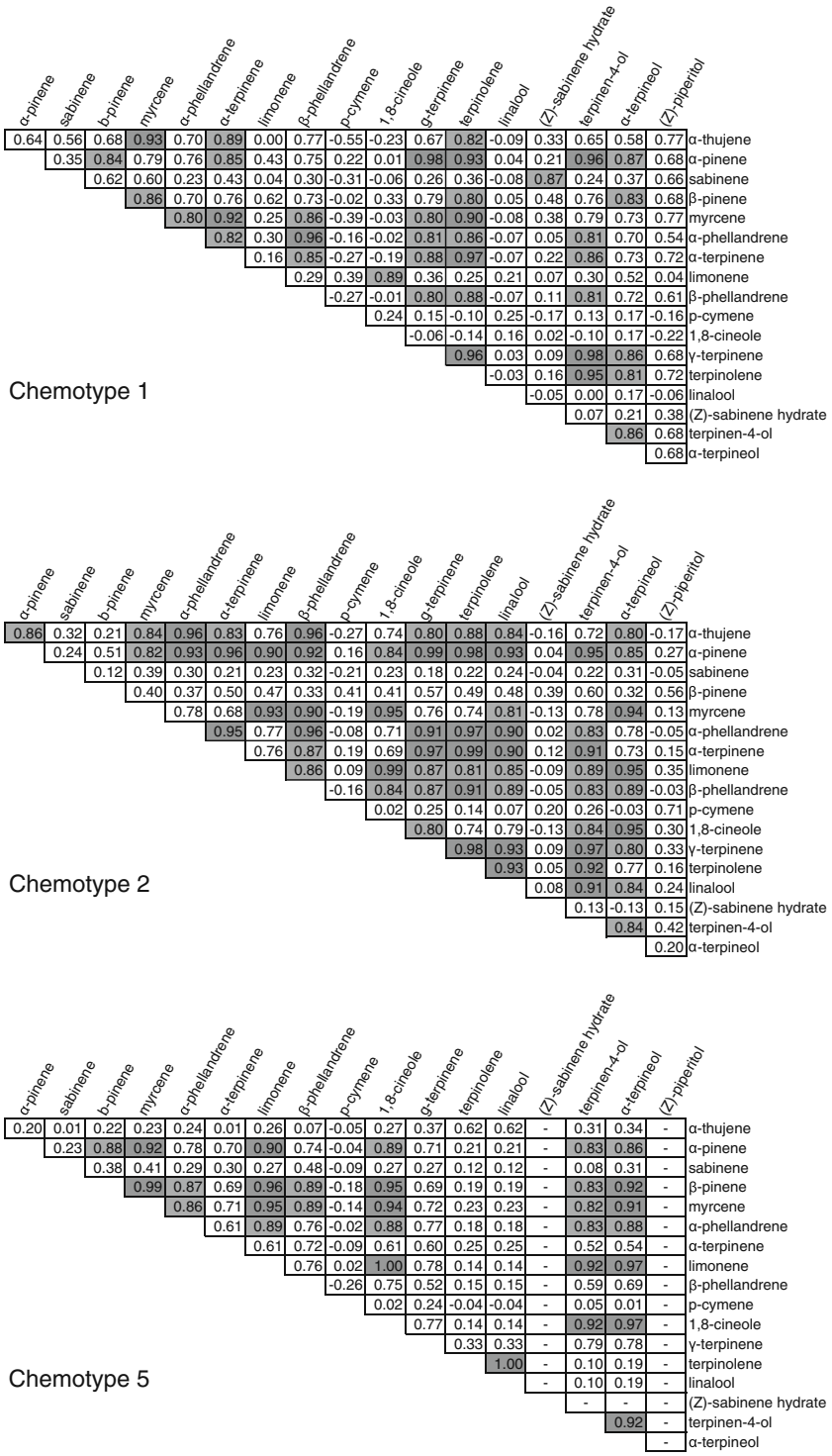
origins (Fig. 1) and calculated intercorrelations between them. Here we discuss the possible underlying biochemical factors that lead to these distinct patterns.

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

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

In chemotypes 2, 3, 6, and 7, however, terpinolene most likely is synthesized by a separate terpinolene synthase. In *M. alternifolia*, the concentration of terpinolene is positively

Table 3 Pearson's pairwise correlation coefficients between individual monoterpene concentrations shown for only the three cardinal chemotypes of *Melaleuca alternifolia*



Values above 0.8 and 0.9 are emphasized using incremental shading. The compounds with the highest levels of correlation in these groups are expected to be the products of a common terpene synthase

correlated with those of linalool and α-phellandrene, and together they comprise Group C terpenes. However, when just focusing on chemotype 2 individuals, it can be seen that α-terpinene, γ-terpinene, α-phellandrene, and β-phellandrene also show strong correlation with terpinolene. These compounds, therefore, may be synthesized by more than one of the enzymes that define the terpene composition of Tea Tree oil.

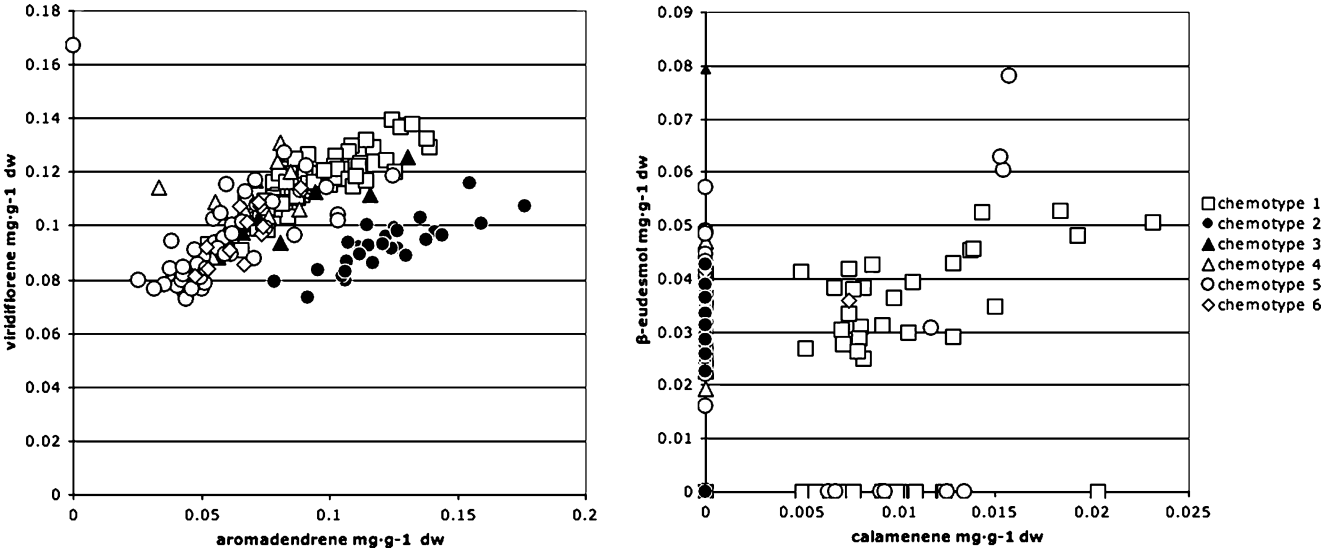


Fig. 3 Correlations of concentrations of the sesquiterpenes viridiflorene and aromadendrende, as well as of β-eudesmol and calamenene with respect to chemotypes

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

Three groups of monoterpenes describe the majority of variation in Australian Tea Tree. Pairwise correlation of concentrations of individual monoterpenes as a proportion of total monoterpenes shows that almost all monoterpenes show strong correlation to one of the major groups. The lowest correlation to any single group is shown by *p*-cymene. However, it has been shown that this compound may be an oxidation product that can be derived from multiple monoterpenes (Brophy et al. 1989), and its concentration may depend more on extrinsic factors that affect *in vivo* oxidation, rather than on individual enzymatic processes.

Sesquiterpene Chemotypes This is the first study to examine sesquiterpene patterns in *M. alternifolia*, and it is clear from the analyses that there are strong correlations among concentrations of individual compounds. Viridiflorene and aromadendrene showed two distinctly different ratios, and the higher viridiflorene ratio was exclusive to chemotype 2. As both share the same viridiflorane structure, it leads us to expect two aromadendrene synthase variants, both responsible for synthesizing the same suite of compounds in slightly different, but nevertheless characteristic ratios. The only other pattern linked to a monoterpene chemotype was the concentration of calamenene, which was consistently absent from all chemotype 2 individuals. As calamenene has two less hydrogens ($C_{15}H_{22}$) than would result from the cyclization and dephosphorylation of FPP ($C_{15}H_{24}(P_2O_7H_3)$), it is likely that its synthesis is linked to the presence of a reductase from the P450 superfamily of enzymes, as in mint monoterpene biosynthesis (Bertea et al. 2001). Compared to monoterpenes, there were many more compounds that were not correlated to other constituents, and our sampling probably missed some combinations. It is interesting to note, however, that the only variation linked to monoterpene variation was found in chemotype 2. None of the other sesquiterpenes showing discontinuous distributions was correlated to other monoterpene chemotypes. This suggests that formation of sesquiterpenes is independent of the monoterpene chemotype in Tea Tree, and that the correlation in chemotype 2 may be at a molecular level.

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

The effect of the proposed individual terpene synthases appears to be codominant. This is supported by the intermediate position of chemotypes 3, and 4, and 6

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

The distribution of chemotypes across the natural range further supports the idea that chemotypes 3, 4, 6, and 7 are likely intermediate chemotypes, particularly because, unlike chemotypes 1, 2, and 5, neither our nor previous studies (Homer et al. 2000) have found them to occur in monochemotypic populations. Furthermore, if they are intermediates, then they should be found in populations that contain the contributing cardinal chemotypes. In one population, Dilkoon Creek, which occurs between the two chemically distinct regions of NSW, 35% of trees recorded were chemotype 4. Chemotypes 1 and 5 also are present in this population. These are the two chemotypes that were used in a controlled cross experiment that reinforced the intermediate nature of chemotype 4 (Shelton et al. 2002). Finally, chemotypes 3, 6, and 7 that contain group C compounds occur mostly in Queensland populations, where chemotype 2 is dominant.

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

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

References

BAKER, G. R., LOWE, R. F., and SOUTHWELL, I. A. 2000. Comparison of oil recovered from tea tree leaf by ethanol extraction and steam distillation. *J. Agric. Food Chem.* 48:4041–4043.

BERTEA, C. M., SCHALK, M., KARP, F., MAFFEI, M., and CROTEAU, R. 2001. Demonstration that menthofuran synthase of mint (*Mentha*) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Arch. Biochem. Biophys.* 390:279–286.

BOHLMANN, J., PHILLIPS, M., RAMACHANDIRAN, V., KATOH, S., and CROTEAU, R. 1999. cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the *Tpsd* gene family from grand fir (*Abies grandis)*. *Arch. Biochem. Biophys.* 368:232–243.

BROPHY, J. J. and DORAN, J. C. 1996. Essential Oils of Tropical *Asteromyrtus*, *Callistemon* and *Melaleuca* Species: In Search of Interesting Oils with Commercial Potential. ACIAR, Canberra.

BROPHY, J. J., DAVIES, N. W., SOUTHWELL, I. A., STIFF, I. A., and WILLIAMS, L. R. 1989. Gas chromatographic quality control for oil of *Melaleuca* terpinen-4-ol type (Australian Tea Tree). *J. Agric. Food Chem.* 37:1330–1335.

BUTCHER, P. A., BELL, J. C., and MORAN, G. F. 1992. Patterns of genetic diversity and nature of the breeding system in *Melaleuca alternifolia* (Myrtaceae). *Aust. J. Bot.* 40:365–375.

BUTCHER, P. A., DORAN, J. C., and SLEE, M. U. 1994. Intraspecific variation in leaf oils of *Melaleuca alternifolia* (Myrtaceae). *Biochem. Syst. Ecol.* 22:419–430.

BUTCHER, P. A., BYRNE, M., and MORAN, G. F. 1995. Variation within and among the chloroplast genomes of *Melaleuca alternifolia* and *Melaleuca linariifolia* (Myrtaceae). *Plant Syst. Evol.* 194:69–81.

CHEN, F., RO, D.-K., PETRI, J., GERSHENZON, J., BOHLMANN, J., PICHERSKY, E., and THOLL, D. 2004. Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant Physiol.* 135:1956–1966.

CORNWELL, C. P., LEACH, D. N., and WYLLIE, S. G. 1995. Incorporation of oxygen-18 into terpinen-4-ol from the $H2^{18}O$ steam distillates of *Melaleuca alternifolia* (tea tree). *J. Essent. Oil Res*. 7:613–620.

COX, S. D., MANN, C. M., and MARKHAM, J. L. 2001. Interactions between components of the essential oil of *Melaleuca alternifolia*. *J. Appl. Microbiol.* 91:492–497.

DEGENHARDT, J., KÖLLNER, T. G., and GERSHENZON, J. 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70:1621–1637.

FISCHER, N., NITZ, S., and DRAWERT, F. 1987. Original flavour compounds and the essential oil composition of marjoram (*Marjorana hortensis* Moench). *Flavour Fragr. J.* 2:55–61.

GIJZEN, M., LEWINSOHN, E., and CROTEAU, R. 1991. Characterization of the constitutive and wound-inducible monoterpene cyclases of Grand Fir (*Abies grandis*). *Arch. Biochem. Biophys.* 289:267–273.

HAMMER, K. A., CARSON, C. F., RILEY, T. V., and NIELSEN, J. B. 2006. A review of the toxicity of *Melaleuca alternifolia* (tea tree) oil. *Food Chem. Toxicol.* 44:616–625.

HOMER, L. E., LEACH, D. N., LEA, D., LEE, L. S., HENRY, R. J., and BAVERSTOCK, P. R. 2000. Natural variation in the essential oil content of *Melaleuca alternifolia* Cheel (Myrtaceae). *Biochem. Syst. Ecol.* 28:367–382.

HUBER, D. P. W., PHILIPPE, R. N., GODARD, K. A., STURROCK, R. N., and BOHLMANN, J. 2005. Characterization of four terpene synthase cDNAs from methyl jasmonate-induced Douglas-fir, *Pseudotsuga menziesii*. *Phytochemistry* 66:1427–1439.

KESZEI, A., BRUBAKER, C. L., and FOLEY, W. J. 2008. A molecular perspective on terpene variation in Australian Myrtaceae. *Aust. J. Bot.* 56:197–213.

LINHART, Y. B. and THOMPSON, J. D. 1999. Thyme is of the essence: Biochemical polymorphism and multi-species deterrence. *Evol. Ecol. Res.* 1:151–171.

LINHART, Y. B., KEEFOVER-RING, K., MOONEY, K. A., BRELAND, B., and THOMPSON, J. D. 2005. A chemical polymorphism in a multitrophic setting: Thyme monoterpene composition and food web structure. *Am. Nat.* 166:517–529.

ROSSETTO, M., SLADE, R. W., BAVERSTOCK, P. R., HENRY, R. J., and LEE, L. S. 1999. Microsatellite variation and assessment of genetic structure in tea tree (*Melaleuca alternifolia:* Myrtaceae). *Mol. Ecol.* 8:633–643.

RUSSELL, M. and SOUTHWELL, I. 2002. Monoterpenoid accumulation in *Melaleuca alternifolia* seedlings. *Phytochemistry* 59:709–716.

SHELTON, D., LEACH, D., BAVERSTOCK, P., and HENRY, R. 2002. Isolation of genes involved in secondary metabolism from *Melaleuca alternifolia* (Cheel) using expressed sequence tags (ESTs). *Plant Sci.* 162:9–15.

SOUTHWELL, I. A. and LOWE, R. F. 1999. Tea Tree. The Genus *Melaleuca*. Harwood Academic Publishers, Amsterdam.

SOUTHWELL, I. A. and STIFF, I. A. 1990. Differentiation between *Melaleuca allternifolia* and *Melaleuca linariifolia* by monoterpenoid comparison. *Phytochemistry* 29:3529–3533.

VERNET, P., GOUYON, P. H., and VALDEYRON, G. 1986. Genetic control of the oil Content in *Thymus vulgaris* L.—a case of polymorphism in a biosynthetic chain. *Genetica* 69:227–231.

WILDERMAN, P. R., XU, M., JIN, Y., COATES, R. M., and PETERS, R. J. 2004. Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol.* 135:2098–2105.

WISE, M. L., SAVAGE, T. J., KATAHIRA, E., and CROTEAU, R. 1998. Monoterpene synthases from common sage (*Salvia officinalis*)—cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *J. Biol. Chem.* 273:14891–14899.

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