Phytochemistry 71 (2010) 844–852

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Update in Bioinformatics

Functional and evolutionary relationships between terpene synthases from Australian Myrtaceae

Andras Keszei [a,*], Curt L. Brubaker [b,1], Richard Carter [a], Tobias Köllner [c], Jörg Degenhardt [c], William J. Foley [a]

[a] Research School of Biology, Australian National University, Canberra 0200, Australia
[b] CSIRO Division of Plant Industry, Black Mountain, Canberra ACT 2601, Australia
[c] Institut für Pharmazie, Martin-Luther-Universität Halle-Wittenberg, Halle 06120, Germany

ARTICLE INFO

Article history:
Received 21 January 2010
Accepted 16 March 2010

Keywords:
Monoterpene synthase
Sesquiterpene synthase
Functional phylogeny
Splicing variation
1,8-Cineole
Pinene
Sabinene hydrate
Citronellal
Geranial
Bicyclogermacrene
Caryophyllene
Alloaromadendrene
Eudesmol
Eucalyptus
Leptospermum
Melaleuca
Callistemon
Corymbia

ABSTRACT

Myrtaceae is one of the chemically most variable and most significant essential oil yielding plant families. Despite an abundance of chemical information, very little work has focussed on the biochemistry of terpene production in these plants. We describe 70 unique partial terpene synthase transcripts and eight full-length cDNA clones from 21 myrtaceous species, and compare phylogenetic relationships and leaf oil composition to reveal clades defined by common function. We provide further support for the correlation between function and phylogenetic relationships by the first functional characterisation of terpene synthases from Myrtaceae: a 1,8-cineole synthase from *Eucalyptus sideroxylon* and a caryophyllene synthase from *Eucalyptus dives*.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The family Myrtaceae is known for the high terpene concentration of the foliage and the considerable qualitative and quantitative variation in foliar terpenes at taxonomic, population and individual levels (for review see Keszei et al., 2008). This variation is important industrially as well as ecologically. Much effort has been devoted to cataloguing the composition of industrially important foliar oils of *Eucalyptus*, *Melaleuca*, *Leptospermum* and related genera (Boland et al., 1991; Brophy et al., 2000; Coppen, 2002) and in relating these variations to the feeding behaviour of herbivorous mammals and insects (Edwards et al., 1993; Lawler et al., 1998; Moore et al., 2004). In spite of this, there has been little effort to elucidate the underlying biochemical processes that lead to such profound variations in the chemical composition of the leaf oils.

Previous studies in *Eucalyptus* have reported that foliar oil concentrations are highly heritable suggesting that this aspect of leaf composition is under strong genetic control (Andrew et al., 2005; Butcher et al., 1996; Jones et al., 2002; Shepherd et al., 1999), but as yet, the specific genes involved in terpene biosynthesis in Myrtaceae are unknown. Although foliar oils in Myrtaceae are dominated by mono- and sesquiterpenes and many of the biochemical pathway elements leading to the production of these compounds have been well characterised in other plants (Chen et al., 2004; Martin et al., 2004; Lücker et al., 2002), similar approaches in *Melaleuca* yielded little result (Shelton et al., 2002, 2004). Genes responsible for quantitative variation in terpenes are most likely to be found in the 2-C-methyl-D-erythritol-4-phosphate (MEP) and mevalonic acid (MVA) pathways, which direct the flow of resources from primary metabolism to processes specialising in terpene biosynthesis (Wildung and Croteau, 2005). The vast

* Corresponding author.
E-mail address: andras.keszei@anu.edu.au (A. Keszei).
[1] Present address: Bayer BioScience N.V., Technologiepark 38, B-9052 Gent, Belgium.

0031-9422/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2010.03.013

majority of monoterpenes are synthesised in the plastids and sesquiterpenes in the cytosol, but there are exceptions (Nagegowda et al., 2008). Allocation of precursors into either of these compartments, and therefore the predominance of either mono- or sesquiterpenes in the oil is thought to be controlled by several steps. The first enzyme implicated is isopentenyl diphosphate isomerase (IDI), which is responsible for maintaining equilibrium between the two essential precursors isopentenyl diphosphate (IDP) and dimethyl-allyl diphosphate (DMADP). Transport of IDP between the plastids and the cytosol has been measured, but the process is unknown. Ultimately, the prenyl diphosphate synthases GDPS and FDPS, which are responsible for producing the direct substrates for terpene synthesis must also be taken into consideration for this aspect of leaf oil variability (Keszei et al., 2008). Ultimately, the diversity and variability of terpenes is due to the terpene synthases (TPS) (Gang, 2005), a family of enzymes which, unlike the majority of the enzymes involved in the biosynthesis of secondary metabolites, are renowned for being able to convert a single substrate into many different products (Schwab, 2003).

The ultimate aim in a molecular approach to terpene biosynthesis is to predict protein function from DNA sequence information. With terpene synthases, this is very difficult due to the tendency to arrive at the same catalytic function in different taxa via convergent evolution. Furthermore, functional enzymes may even arise from recombination between TPS subfamilies (Dudareva et al., 1996). However, assessing the degree of sequence homology across TPS sequences from closely related species can give an insight into evolutionary relationships across species.

The family Myrtaceae is an ideal system to study molecular differences relating to protein function of terpene synthases since many of its genera contain numerous closely related species with a variety of leaf oil profiles. In addition, intra-specific chemical variation is common in many species (Keszei et al., 2008). In such a system, there is a high likelihood of finding similar sequence variants with different functions, and such sequences are the most valuable in establishing relationships between DNA sequences and protein function. This paper provides an insight into the terpene synthases of Australian Myrtaceae.

2. Results

2.1. Chemical analysis

We collected leaf from 21 species that represent some of the most important members of Myrtaceae with regards to foliar oil production. *Eucalyptus polybractea* is one of the most important species for the production of 1,8-cineole in Australia, *Eucalyptus dives* leaf is harvested for its high piperitone content, *Corymbia citriodora* is planted worldwide for its citronellal-rich oil (Coppen 2002), and *Melaleuca alternifolia* yields a unique medicinal oil high in terpinen-4-ol (Butcher et al., 1996). Not only do these four species represent significant and markedly different chemistries, but they also represent taxonomically distinct clades within Myrtaceae (Steane et al., 2002; Wilson et al., 2005). *Corymbia*, *Eucalyptus*, and *Melaleuca* are separate genera, and *E. polybractea* and *E. dives* represent the two major subgenera of *Eucalyptus*: *Symphyomyrtus* (symphyomyrts) and *Eucalyptus* (monocalypts). The remaining species were chosen to better understand how terpene biochemistry affected, or was affected by the evolution of Myrtaceae. With this in mind, care was taken to choose several species with similar leaf oils, as well as species that are taxonomically close to each other. We used the data from recent syntheses of oil chemistry (Boland et al., 1991; Brophy et al., 2000; Coppen, 2002) as a guide to the likely chemistry of these different species but all samples were analysed by GC–MS as part of this work. Table 1 lists the species in the study.

Table 1
The list of species studied, indicating the CPGN abbreviations used in gene names, the number of positive clones obtained from 3′-RACE and the number of unique sequences identified in each of the Type III terpene synthase subfamilies.

Species	Abbreviation	Clones	TPSa	TPSb
Callistemon citrinus	CALci	5	2	2
Corymbia citriodora	CORci	8	4	–
E. viminalis ssp. *pryoriana*	EUCpr	6	1	1
Eucalyptus aggregata	EUCag	12	1	1
Eucalyptus bancroftii	EUCba	16	3	2
Eucalyptus camaldulensis	EUCca	23	6	1
Eucalyptus cinerea	EUCci	6	1	1
Eucalyptus dives	EUCdi	7	2	3
Eucalyptus globulus ssp. *globulus*	EUCgl	24	3	2
Eucalyptus grandis	EUCgr	4	1	2
Eucalyptus leucoxylon	EUCle	6	1	–
Eucalyptus melliodora	EUCme	4	1	1
Eucalyptus nicholli	EUCni	5	4	1
Eucalyptus pauciflora	EUCpf	6	1	1
Eucalyptus polybractea	EUCpb	8	2	2
Eucalyptus rossii	EUCro	9	3	2
Eucalyptus rubida	EUCru	13	3	3
Eucalyptus sideroxylon	EUCsi	18	2	4
Eucalyptus tricarpa	EUCtr	5	1	1
Leptospermum petersonii	LEPpe	4	2	–
Melaleuca alternifolia	MELal	10	2	2

As is characteristic of Myrtaceae, monoterpenes dominated the oils of most samples, except for that of *E. pauciflora* whose oil contains over 50% of the sesquiterpene, bicyclogermacrene. The hallmark terpene in eucalypts, 1,8-cineole was the dominant component in nine of the leaf oils. α-Pinene, α- and β-phellandrene, sabinene hydrate, β-citronellal, geranial, and piperitone were the dominant terpenes in the remainder of the samples. Chemically, both acyclic and cyclic monoterpenes, as well as monoterpene hydrocarbons, alcohols, ethers, aldehydes and ketones are represented. The dominant sesquiterpenes were β-caryophyllene, alloaromadendrene, bicyclogermacrene and β-eudesmol (Fig. 1).

Although leaf oils from most of our species had been previously analysed, (Boland et al., 1991), their compositions differed significantly from existing data. Most differences can be attributed to differences in sampling and sample preparation. As we collected young, expanding leaf directly into liquid nitrogen, there was little opportunity for spontaneous degradation of the biosynthesis products. The most conspicuous examples were the abundances of α-phellandrene and bicyclogermacrene, and corresponding low concentrations of *p*-cymene and viridiflorane type oxygenated sesquiterpenes which have been shown to be possible photochemical and thermal dehydrogenation products of the former (Spraul et al., 1991; Toyota et al., 1996).

To highlight how little is currently understood of the chemical variability in Myrtaceae, three of our 21 samples showed fundamental differences to known chemistries. Our samples from *Eucalyptus rubida* and *Eucalyptus tricarpa* were dominated by α- and β-phellandrene, respectively, whereas both species have been previously reported to contain high concentrations of 1,8-cineole (Boland et al., 1991). Likewise, the oil of our *Eucalyptus globulus ssp. globulus* contained 24% sesquiterpenes, mainly α- and β-eudesmols, while Boland et al. (1991) reported oil with a low sesquiterpene fraction high in globulol. Since the purpose of our study was to obtain chemical and expressed sequence data from the same sample, and from samples that most closely reflect the direct products of enzymatic catalysis as possible, these differences are immaterial.

2.2. Novel terpene synthases from Australian Myrtaceae

We sequenced over 200 individual clones of TPS fragments expressed in the young leaves. Based on BLAST search results, we

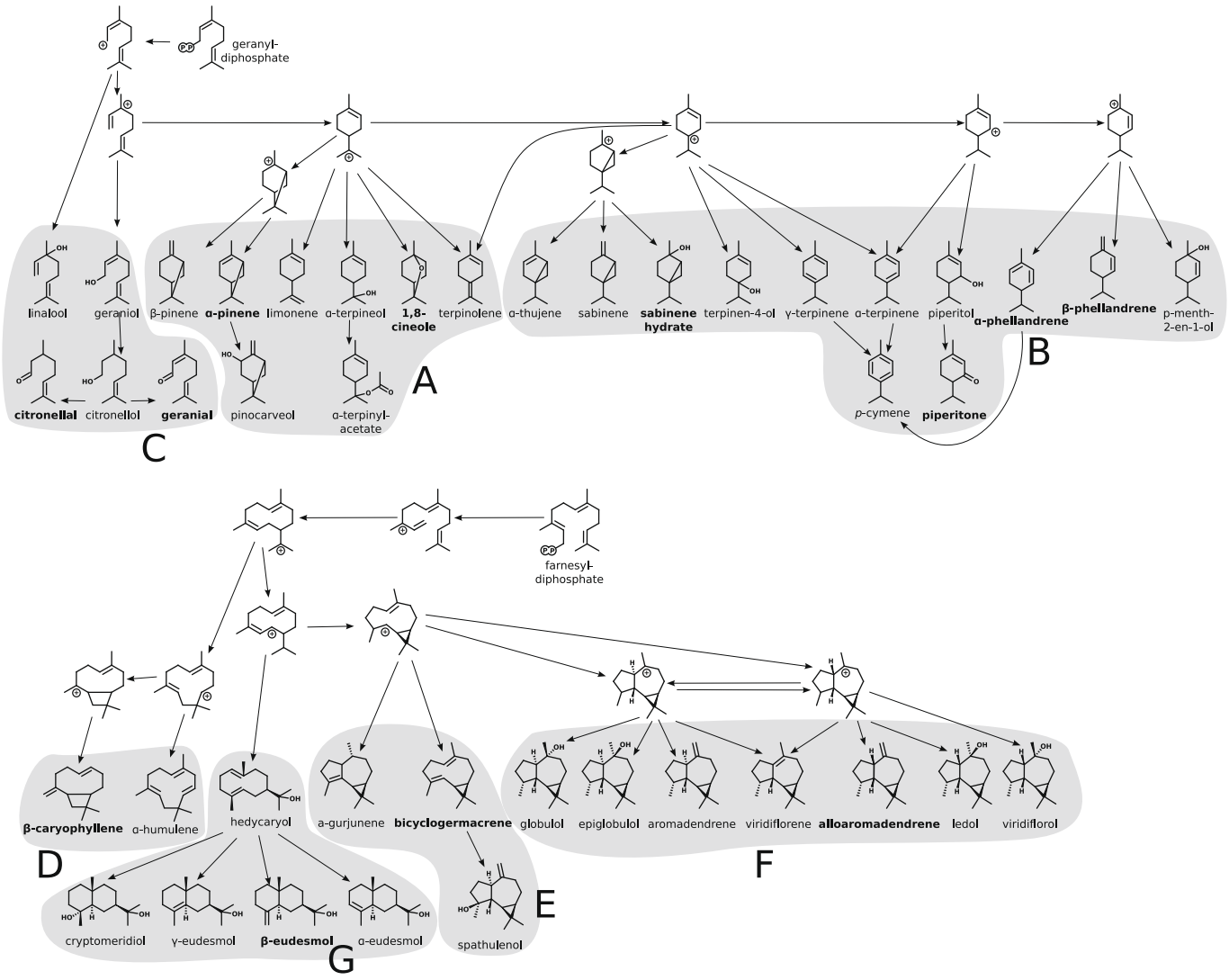

Fig. 1. The proposed biosynthetic routes for the major mono- and sesquiterpenes identified from the 21 species of Myrtaceae in the current study. Common carbocation precursors determine groupings A–G. Compounds which represent the major mono- or sesquiterpenes in individual oil samples are in bold.

identified 78 unique sequences that were similar to known sequences in the TPS gene family. Of these, 46 sequences were similar to TPSa, and 32 sequences were similar to the TPSb gene subfamily. Our fragments were, on average, 1000 bp in length and contained slightly less than half of the open reading frame (ORF). The fragments contained most of the catalytic pocket coding residues based on comparisons with the 3D protein structure of bornyl-diphosphate synthase from *Salvia officinalis* (AF051900) (Whittington et al., 2002).

We used *C. citriodora*, *E. dives*, and *E. sideroxylon* to obtain full-length sequences because of their taxonomic relationships and their diversity of oil profiles. The acyclic terpenes citronellal and citronellol dominate the oil of *C. citriodora*, whereas the major constituent of *E. dives* foliar oil was the cyclic monoterpene ketone, piperitone. In contrast the foliar oil of *E. sideroxylon* was composed predominantly of 1,8-cineole, a monoterpene cyclic ether.

We isolated three full-length TPS genes from *E. dives*, four from *E. sideroxylon*, and a single full-length TPS from both *C. citriodora* and *E. grandis*. Analysis of the 5′ termini using both the ChloroP and Protein Prowler servers (Emanuelsson et al., 1999; Bodén and Hawkins, 2005) predicted the presence of chloroplast targeting peptides in *E. sideroxylon* TPS1 and TPS5, and *E. dives* TPS3, suggesting that these were members of the TPSb monoterpene synthase subfamily. In contrast, *E. sideroxylon* TPS2 and TPS3, *C. citriodora* TPS4, *E. dives* TPS2 and *E. grandis* TPS2 showed no such domains, which agrees with the cytosolic localisation of TPSa sesquiterpene synthases.

2.3. Phylogenetic analysis of terpene synthases

We included the full-length TPS sequences found in *E. dives*. *E. sideroxylon*, *E. grandis* and *C. citriodora* in an amino acid alignment featuring representative sequences from all TPS subfamilies to confirm their phylogenetic placement (Fig. 3).

The relationships between the C-terminal TPS fragments were inferred from a nucleotide alignment, in which we included reference sequences from the full sequence phylogeny (Fig. 4). To correlate tree topology to functional outcomes, we mapped the abundances of terpenes with similar catalytic intermediates to the terminal branch nodes. The tree shows the same topology as the one obtained from full-length sequences, and provides further material for suggesting evolutionary trends.

2.4. Functional characterisation of full-length TPS

Transgenic expression of TPS1, the most abundant terpene synthase transcript from *E. sideroxylon* yielded a protein 553 amino acids in length, corresponding to the predicted size of the mature protein following the cleavage of the chloroplast targeting peptide. The protein showed no catalytic activity with FDP; however it produced a complex mixture of monoterpenes when incubated with GDP (Fig. 2). The major product was 1,8-cineole, the dominant monoterpene in *E. sideroxylon*, while most of the minor peaks were also present in the leaf oil.

Transgenic expression of *E. dives* TPS2 yielded a protein 566 amino acids in length, corresponding to the predicted size of the mature protein. The protein showed no catalytic activity with GDP; however it produced a mixture of caryophyllene and humulene and a small quantity of an unknown sesquiterpene when incubated with FDP (Fig. 2).

3. Discussion

3.1. Relationships between terpene synthases from Myrtaceae

TPS subfamilies are well defined, and coincide with substrate specificity and broad-term catalytic activity in Angiosperms (Bohlmann et al., 1998). There is strong differentiation between terpene synthases involved mainly in primary metabolic processes (TPSc and TPSe), also known as Type I and II terpene synthases, and the Type III terpene synthases (TPSa and TPSb) involved in the production of terpene secondary metabolites.

In the phylogram showing the relationships between known terpene synthases and those identified from Myrtaceae (Fig. 3), the transcripts from Myrtaceae (bold) clearly fall into clades representing Type III TPSs involved in secondary metabolism. All putative sesquiterpene synthase (TPSa) genes from Myrtaceae are monophyletic. Within TPSb, EUCsi;TPS5 is the only full-length sequence that is in a separate clade from other myrtaceous sequences, and instead shows highest similarity to the clade that represents characterised isoprene synthases (Sharkey et al., 2005) and the putative limonene synthase from *M. alternifolia* (Shelton et al., 2004). We concur with the conclusions of Sharkey et al. (2005) that these sequences are in fact isoprene synthases, as the most highly expressed monoterpene synthases from Myrtaceae form a separate phylogenetic group. This clade in TPSb is a sister clade to both the isoprene synthases and all other Angiosperm monoterpene synthases represented in the phylogeny. This confirms the hypothesis that isoprene biosynthesis and the ability to utilise the C_5 precursors in favour of larger prenyl diphosphates shares a common origin with monoterpene synthases, but evolved independently in Angiosperms.

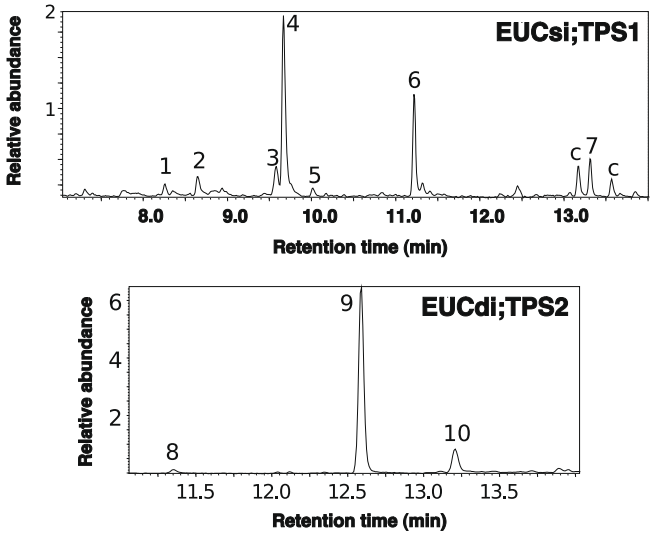

Fig. 2. Product spectrum of EUCsi;TPS1 and EUCdi;TPS2. The enzymes were expressed in *E. coli* and incubated with the substrates GPP and FPP, respectively, in the presence of 10 mM $MgCl_2$. Shown are the total ion traces of monoterpene products after GC–MS analysis. Compounds labelled with asterisks (*) were identified by comparison of mass spectra and retention times to those of authentic standards. All other products were identified using the Wiley mass spectra library. 1: sabinene, 2: myrcene, 3: limonene*, 4: 1,8-cineole*, 5: ocimene, 6: linalool*, 7: α-terpineol*, 8: unknown sesquiterpene, 9: β-caryophyllene, 10: α-humulene, c: contamination.

In Fig. 4, all TPS sequences isolated from Myrtaceae are represented including partial cDNA clones. With this wider coverage, we can see that both *Melaleuca* and *Eucalyptus* sequences occur in Clade 1, sequences from both the *Eucalyptus* and *Symphyomyrtus* subgenera occur in Clades 2 and 3, and all studied genera are represented in Clades 4 and 5. As sequences from Myrtaceae in both TPSa and TPSb show monophyly, and there is no further taxonomic signal corresponding to lower level phylogenetic grouping, we can assume that the split between the different clades of monoterpene synthases, and therefore the observed diversity of terpene synthases pre-dates the cladogenesis of myrtaceous genera.

As we could not find taxonomic relationships that would coincide with the terpene synthase phylogeny, we proceeded with a functional comparison. The major differences between terpene synthases within a subfamily are usually based on product specificity. During the conversion of prenyl diphosphates into terpenes, the substrate goes through intermediate carbocationic stages. We grouped both monoterpenes and sesquiterpenes found in the 21 samples according to shared carbocation precursors (Fig. 1), and mapped the abundance of terpenes belonging to each group to the species represented by terminal nodes of the phylogenetic tree.

The most abundant monoterpene synthases from both *E. dives* and *E. bancroftii*, two of the three samples whose oils contain high concentrations of α-phellandrene map to Clade 2. Furthermore, the most abundant transcripts from the samples with high concentrations of group A monoterpenes (mainly 1,8-cineole) all occur in Clade 3. Functional characterisation of the dominant monoterpene synthase from *E. sideroxylon* (EUCsi;TPS1) as a 1,8-cineole synthase confirms the link between functional and phylogenetic similarity in this group. Among the sesquiterpene synthase (TPSa) sequences, the main separation is between Clades 4 and 5. The most abundant transcripts from individuals containing high proportions of group D sesquiterpenes (β-caryophyllene and α-humulene) fall in Clade 4. Functional characterisation of EUCdi; TPS2 as a β-caryophyllene synthase once again strengthens the functional phylogeny.

Clade 5, containing the majority of the sesquiterpene synthase transcripts, shows no further phylogenetic separation or any clear pattern related to leaf oil profiles. All four types of sesquiterpene skeletons are represented in the leaf oils of these samples. We hypothesise that Clade 5 represents sesquiterpene synthases that catalyze conversions with a germacrenyl carbocation intermediate, of which the viridiflorane type compounds are the most abundant in Myrtaceae.

The strong bootstrap support for the described clades, and the high representation of sequences from species showing distinct chemical differences suggests that in Myrtaceae, and specifically in *Eucalyptus*, an early divergence and subsequent maintenance of function resulted in a close correlation between sequence homology and function not usually found among terpene synthases. Despite these correlations, caution should be exercised in interpreting functional patterns because very small sequence differences may profoundly affect the function of terpene synthases (Kampranis et al., 2007; Köllner et al., 2004). A well supported

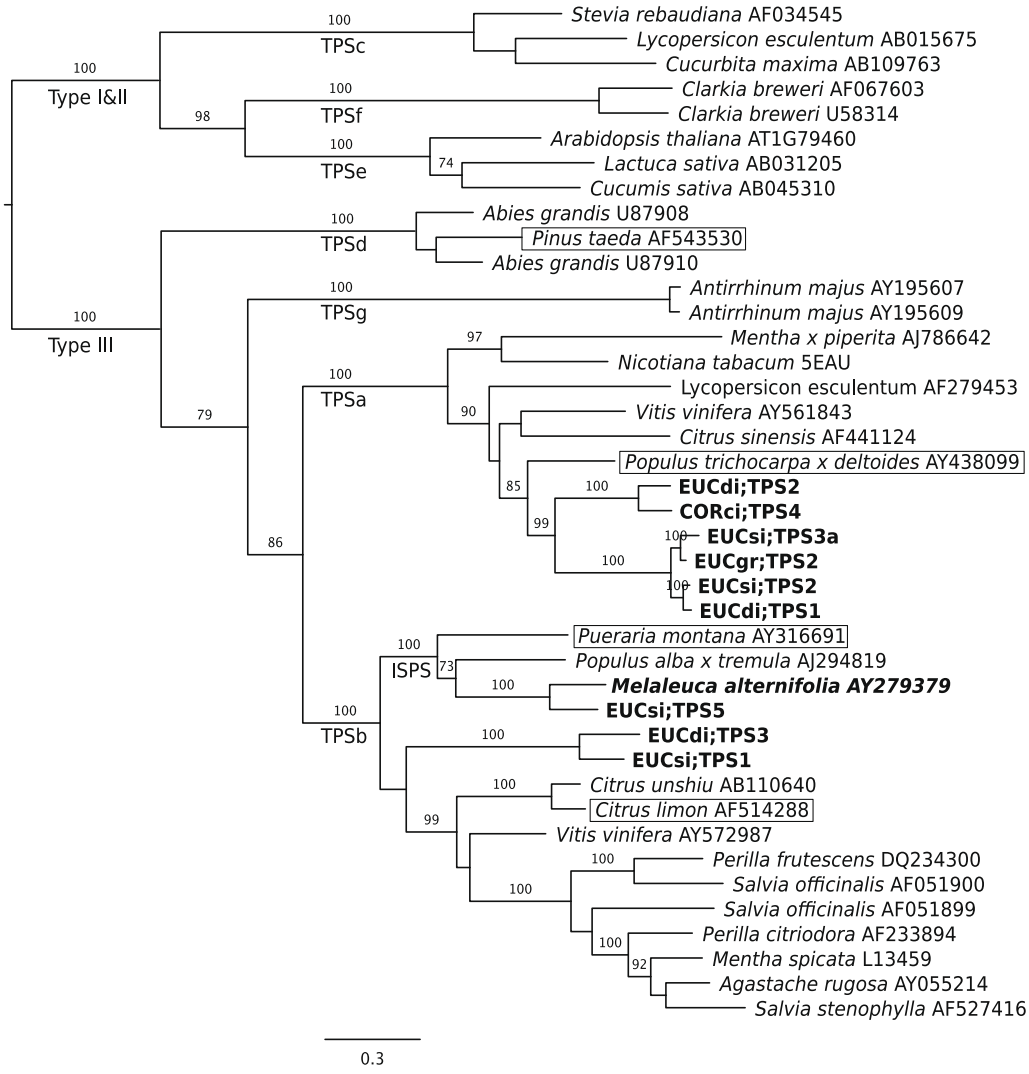

Fig. 3. Maximum likelihood tree of full-length terpene synthase protein sequences from Australian Myrtaceae. Branch labels indicate confidence values based on 1000 bootstrap replicates. Terminal node labels in bold represent sequences from Myrtaceae. Sequences used as reference points for the fragment phylogeny are boxed.

functional phylogeny can nevertheless give valuable insight into the origins of the different chemistries which dominate the ecosystems of the Australian continent.

C. citriodora and *L. petersonii* had been chosen for this study because both of their leaf oils contain over 93% group C monoterpenes (the aliphatic aldehydes citronellal and citral). While citronellal is a significant essential oil product, its biosynthesis has not been described from plants. We have found that the degenerate primer used to obtain close to 80 individual terpene synthases from Myrtaceae did not isolate any TPSb sequences from either of these species. This may simply indicate that the implicated TPS genes differ significantly at the primer binding site. However, as the primer had been designed around the DDXXD motif, conserved in not only angiosperm but also gymnosperm terpene synthases (Bohlmann et al., 1998), we propose that alternative origins for these terpenes, such as prenyl diphosphate synthases with changed catalytic activity (Blanchard and Karst, 1993), should also be taken into consideration.

3.2. Variation in intron splicing

In some of the species from which we isolated multiple TPS clones, we observed several sequence length polymorphisms, which usually showed a shift in the protein coding frame. From *E. rubida*, *E. sideroxylon*, *E. nicholii* and *C. citrinus*, the observed mRNA length discrepancies are summarised in Fig. 5. All the length differences in the fragments coincide with the splicing site of intron 4, as confirmed by comparison to genomic sequences *from Arabidopsis thaliana*. Both TPS1x from *E. sideroxylon* and TPS2x from *E. rubida* have deletions of 14 bp at the start of exon 5, producing a frame shift and resulting in premature stop codons three amino acids downstream of the deletion in both sequences. The terpene synthase fragments TPS4x from *C. citrinus*, TPS4x from *E. nicholii* and TPS2x from *E. bancroftii* have insertions of 87, 87 and 88 bp in the same position. Furthermore, EUCba;TPS2x is missing the remainder of the exons up until the poly-A tail, and CALci;TPS4x is missing exon sequence from 96 bp into exon 5 all the way to the polyadenylation site. In *E. nicholii* TPS4x, through the failure to excise all or part of intron 4, downstream premature stop codons are introduced which are expected to result in truncated translation products missing the majority of the catalytic domain.

In plants, phenotypic variation due to inefficient RNA splicing has only been shown in *E. nitens* CCR genes in relation to microfibril angle (Thumma et al., 2005). The phenomenon is believed to be caused by changes in exon splicing enhancers (ESE) (Hastings and Krainer, 2001), and is well documented in humans as the cause of several medical conditions (Denson et al., 2006; Maciolek et al., 2006; Mine et al., 2003).

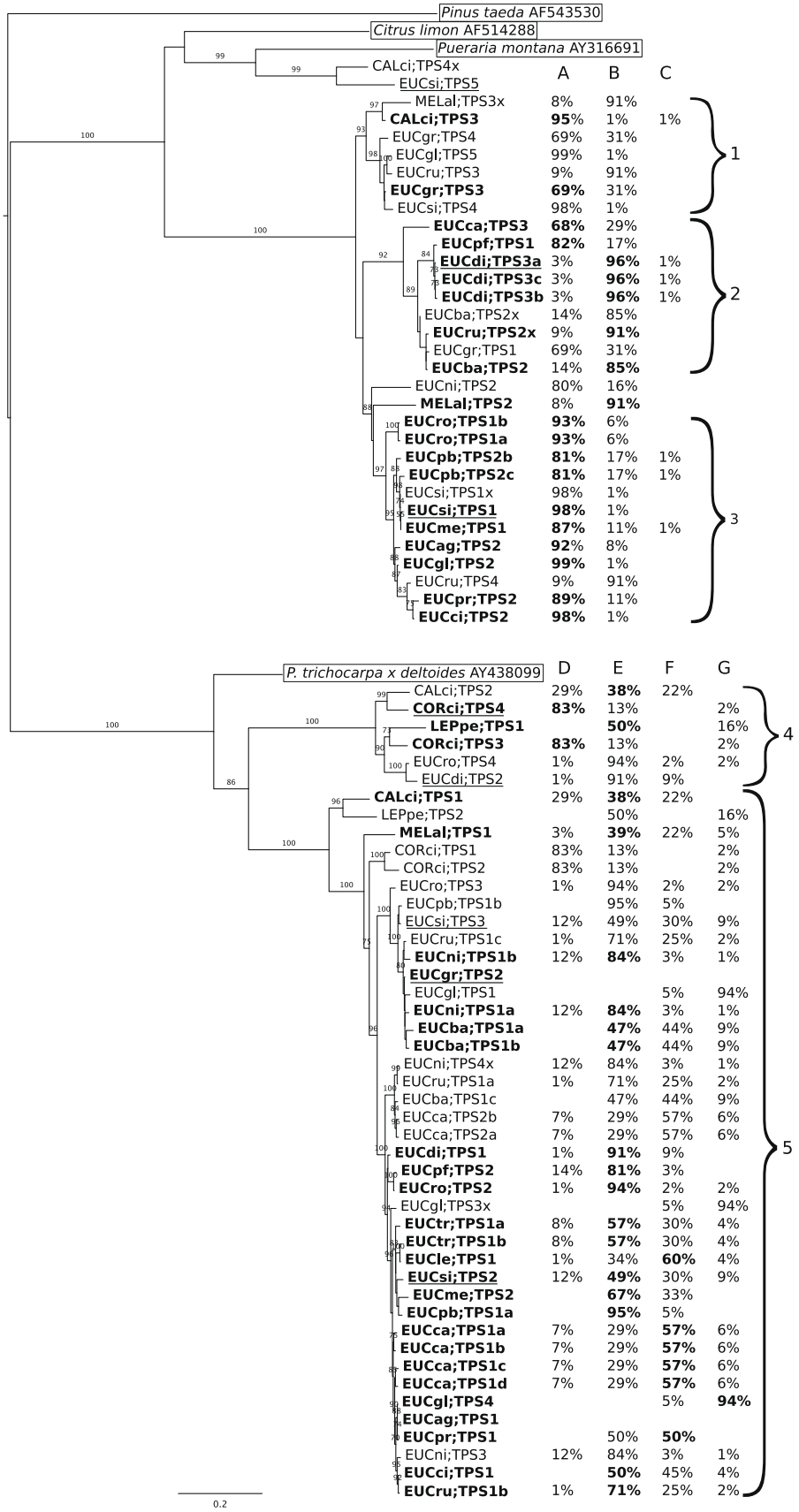

Fig. 4. Maximum likelihood tree of terpene synthase c-terminal fragments from Australian Myrtaceae. Branch labels indicate confidence values based on 1000 bootstrap replicates. Terminal node labels in bold indicate the most abundant transcript from the species. Full-length sequences of the underlined transcripts are also represented in Fig. 3. Columns A–G indicate the proportion of biochemically similar compounds in the relevant leaf oil fraction, as indicated in Fig. 1.

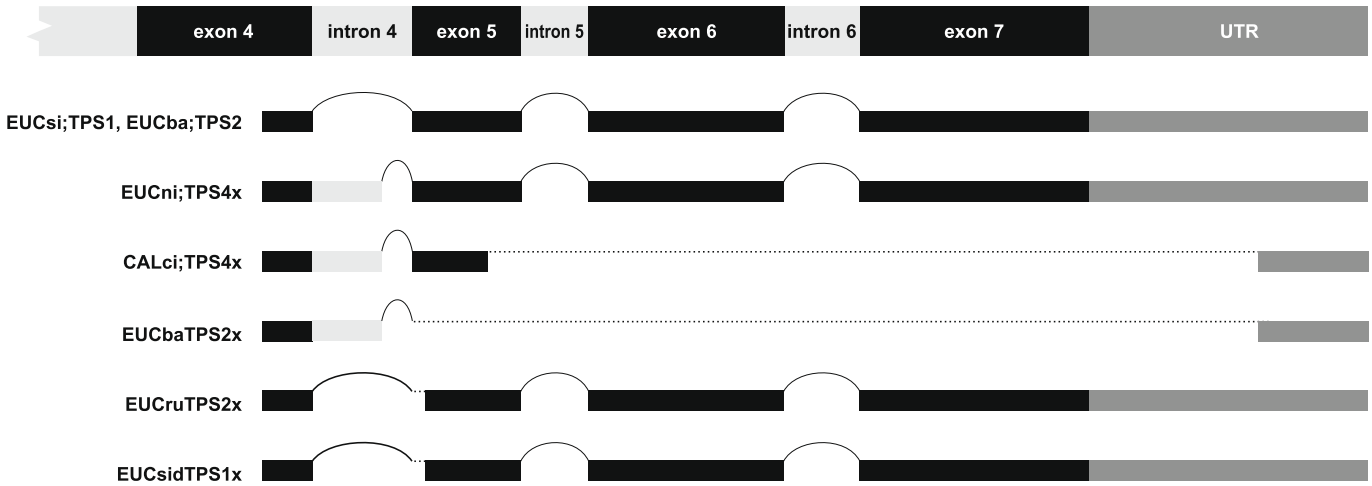

Fig. 5. Splice variants in Myrtaceae TPS transcripts. The top reference is based on TPSa and TPSb gene structure in A. thaliana. Introns are grey, spliced introns are represented by arcs, and missing exon sequence is indicated by dotted lines.

Of the three splicing sites covered by the TPS fragments in this study, we observed a strong tendency for mis-splicing in just one. As the surrounding exon sequences may be quite different (EUCsi;TPS1x belongs in TPSb whereas EUCru;TPS2x belongs in TPSa), this may indicate a highly conserved motif which influences correct splicing in Myrtaceae. Both correctly and incorrectly spliced variants of EUCsi;TPS1 and EUCba;TPS2 could be amplified from cDNA, and this may indicate either allelic variation of alternate variants, or the co-occurrence of different splicing forms of the same mRNA transcript. Based on the frequency of the phenomenon among the transcripts identified, it appears that the loss of function in specific terpene synthases due to altered splicing may be one of the causes of intra-specific variability observed in Myrtaceae.

3.3. Concluding remarks

We have successfully isolated terpene synthases from some of the most important oil bearing species of Myrtaceae. The sequences obtained are transcripts from young leaf actively synthesising terpenes, and are therefore characteristic of the terpene synthases contributing to the final leaf oil. Using molecular data from both taxonomically and chemically similar species we have been able to place these sequences in a chemotaxonomic context. We have shown strong patterns indicating the persistence of function in this otherwise highly variable gene family, and have functionally characterised the first terpene synthase genes, a cineole and a caryophyllene synthase, from any species of Myrtaceae.

Understanding the correlations between sequence and phenotype is increasingly important as whole genome sequencing is underway in Myrtaceae. To make the best use of the genomic information that is becoming available, we need to understand how individual genes contribute to the final phenotype – in this case leaf oil composition, and what the origins of chemical variability are in an evolutionary as well as functional sense. The results of this paper make this possible by providing a starting point for future functional characterisation and expression studies.

4. Experimental

4.1. Plant material and chemical analysis

Young expanding leaf was collected in liquid nitrogen from the Australian National Botanic Gardens and stored at −80 °C prior to extraction of DNA, RNA, or leaf oil.

For chemical analysis, 100 mg of leaf, ground to powder in liquid nitrogen, was extracted in 500 μl pentane in the dark for 24 h at room temperature. An Agilent model 6890N gas chromatograph was used with helium as the carrier gas at a flow rate of 1 ml min^{-1}, 25:1 split injection (injector temperature 250 °C, injection volume 1 μl), using an Alltech AT-35 column ((35%-phenyl)-methylpolysiloxane, 60 m, 0.25 mm i.d., 0.25 mm thickness, Alltech, USA) and a temperature program from 100 °C (5 min hold) at 20 °C min^{-1} to 200 °C, followed by 5 °C min^{-1} to 250 °C (4 min hold). The coupled mass spectrometer was an Agilent model 5973 with a quadrupole mass selective detector, transfer line temperature 230 °C, source temperature 230 °C, quadrupole temperature 150 °C, ionisation potential 70 eV, and a scan range of 40–350 atomic mass units (amu). Compounds were identified by comparison of retention times and mass spectra to reference spectra in the Wiley and National Institute of Standards and Technology libraries.

4.2. Isolation of nucleic acids

RNA was extracted using RNeasy micro kits (Qiagen, Valencia, CA) from 100 mg of leaf ground in liquid nitrogen. The extraction buffer was modified by adding 2% PVP (Sigma–Aldrich P5288) and/or 120 mg ml^{-1} sodium-isoascorbate (to saturation) (Sigma–Aldrich 496332) to inhibit the oxidative conjugation of phenolics to the RNA (Suzuki et al., 2003). Extraction was performed without adjuvants, with the addition of only PVP, only sodium isoascorbate, or both, until final RNA concentrations of 100 ng μl^{-1} or more were obtained. RNA concentrations were calculated from UV absorbance measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

4.3. cDNA synthesis, amplification, cloning and sequencing

Between 0.5 and 5 μg of RNA was used for cDNA first strand synthesis using RevertAid™ M-MuLV reverse transcriptase (MBI Fermentas, Burlington, Ontario, Canada) primed by single-nucleotide anchored 35-mer polythymidosine primer ($T_{35}V$). 3′-RACE was carried out using 5 μl of first strand reaction as template for a 50 μl reaction, with 35 repeats of the following thermal program: 30 s at 94 °C, 30 s at 52 °C, and 120 s at 72 °C. DNA polymerisation was primed by $T_{35}V$ and a degenerate forward primer (T[ATGC]-GATGAT[AG]TTTA[CT]GATGT[GC]TATGG) based on the DDXXD divalent ion binding motif characteristic of most terpene synthases.

To obtain full-length clones, we performed 5′-RACE using the SMART® 5′ RACE protocol (BD Biosciences Clontech, Palo Alto,

CA). cDNA incorporating the SMART 5′ adapter was synthesised from purified mRNA (Oligotex-resin, QIAGEN, Valencia, CA), and used as a template for subsequent PCR reactions.

Based on knowledge of both the 3′ and 5′ ends, nested primer pairs were designed to include as much as possible of both the 3′ and 5′ untranslated regions, and amplification of full-length clones was carried out using *Pfu* proofreading polymerase (Invitrogen, Carlsbad, CA). PCR products were resolved in and excised from 1% TAE-agarose gels. Prior to ligation the bands were cleaned using the QIAquick membrane purification system (QIAGEN, Valencia, CA). The full-length amplicons were ligated into the pGEM-T vector (Promega, Madison, WI) which we subsequently used to transform JM-109 chemically competent *Escherichia coli* cells (Promega, Madison, WI). The inserts were amplified from colony lysates using AmpliTaq (Invitrogen, Carlsbad, CA) DNA polymerase and M13 primers, and all PCR products in the expected size range were cleaned from agarose gel using the QIAquick membrane purification system (QIAGEN Valencia, CA). Purified DNA was sequenced using BigDye v3.1 dye terminator chemistry on an ABI3100 capillary sequencer (Applied Biosystems, Foster City, CA).

4.4. Heterologous expression and functional characterisation

EUCsi;TPS1 was amplified using the following primers: eusidTPSfwd (ATGGTAACCTGCATTAGCGCGCCCAATTCGTGACATGCGCTT) and eusidTPSrev (ATGGTAACCTGCATTATATCAATCGAGGGGCACGGATTCAAATA), and EUCdi;TPS2 using eudivTPSfwd (ATGGTAGGTCTCAGCGCTCTCTTCCGATTTCAGCAACTCC) and eudivTPSrev (ATGGTAGGTCTCATATCACTGCACTGGGTCTATGAGCACC). PCR was performed using Advantage 2 polymerase mix (BD Biosciences, Palo Alto, CA). The resulting PCR product was directly inserted as a *Bsp*MI fragment into the expression vector pASK-IBA7 (IBA GmbH, Göttingen, Germany). Expression and partial purification of the recombinant protein followed the procedure described in Köllner et al. (2004). To determine the catalytic activity of the recombinant protein, enzyme assays containing 50 μl of the bacterial extract and 50 μl assay buffer (10 mM MOPSO [pH 7.0], 1 mM dithiothreitol, 10% [v/v] glycerol) with 10 μM substrate ((*E,E*)-GPP (Echelon Biosciences, Salt Lake City, UT, USA) and (*E,E*)-FPP, respectively), a divalent metal cofactor (10 mM $MgCl_2$), 0.2 mM Na_2WO_4 and 0.1 mM NaF in a Teflon-sealed, screw-capped 1 ml GC glass vial were performed. A solid phase microextraction (SPME) fibre consisting of 100 μm polydimethylsiloxane (SUPELCO, Belafonte, PA, USA) was placed into the headspace of the vial for 30 min incubation at 30 °C. For analysis of the adsorbed reaction products, the SPME fibre was directly inserted into the injector of the gas chromatograph.

A Shimadzu model 2010 gas chromatograph was employed with the carrier gas He at 1 ml min^{-1}, splitless injection (injector temperature: 220 °C, injection volume: 1 μl), a Chrompack CP-SIL-5 CB-MS column ((5%-phenyl)-methylpolysiloxane, 25 m × 0.25 mm i.d. × 0.25 μm film thickness, Varian, USA) and a temperature program from 50 °C (3-min hold) at 6 °C min^{-1} to 180 °C (1 min hold). The coupled mass spectrometer was a Shimadzu model QP2010Plus with a quadrupole mass selective detector, transfer line temperature: 230 °C, source temperature: 230 °C, quadrupole temperature: 150 °C, ionisation potential: 70 eV and a scan range of 50–300 amu. Compounds produced by EUCsi;TPS1 and EUCdi;TPS2 were identified by comparison of mass spectra and retention times to those of authentic standards or using the Wiley mass spectra library.

4.5. Sequence analysis and phylogenetic modelling

The sequences' base-calls were reviewed in FinchTV sequence analysis software (Geospiza Inc., Seattle, WA). The assembly of contigs and the removal of cloning vector sequence was performed using BioEdit 5.09 (Ibis Therapeutics, Carlsbad, CA). The assembled sequences were subsequently compared to sequences published on GenBank using BLASTx translated protein search (Altschul et al., 1990). Preliminary alignments of the sequences were generated using ClustalW (Thompson et al., 1994), and manual adjustments were made to regions of ambiguous alignment. Maximum likelihood phylogenies were calculated using RAxML (Stamatakis et al., 2005) with 1000 bootstrap replicates. Full-length sequence alignments and phylogenies were calculated using the Dayhoff similarity matrix on the amino acid sequence, to correspond to the family and subfamily demarcations in the terpene synthase phylogeny (Bohlmann et al., 1998).

Acknowledgements

This work was supported partly by grant #DP0877063 from the Australian Research Council, HSF 05-6 from the Hermon Slade Foundation, and the inclusion of *Melaleuca alternifolia* by grant #ANU74 from the Rural Industries Research and Development Corporation.

We would like to thank Bronwyn Matheson for her expert technical support in expanding the study to multiple species, and the Australian National Botanic Gardens, especially Frank Zich for access to plant material.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Andrew, R.L., Peakall, R., Wallis, I.R., Wood, J.T., Knight, E.J., Foley, W.J., 2005. Marker-based quantitative genetics in the wild?: the heritability and genetic correlation of chemical defenses in *Eucalyptus*. Genetics 171, 1989–1998.

Blanchard, L., Karst, F., 1993. Characterization of a lysine-to-glutamic acid mutation in a conservative sequence of farnesyl diphosphate synthase from *Saccharomyces cerevisiae*. Gene 125, 185–189.

Bodén, M., Hawkins, J., 2005. Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics 21, 2279–2286.

Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc. Natl. Acad. Sci. USA 95, 4126–4133.

Boland, D.J., Brophy, J.J., House, A.P.N., 1991. Eucalyptus Leaf Oils. Inkata Press, Sydney.

Brophy, J.J., Goldsack, R.J., Punruckvong, A., Bean, A.R., Forster, P.I., Lepschi, B.J., Doran, J.C., Rozefelds, A.C., 2000. Leaf essential oils of the genus *Leptospermum* (Myrtaceae) in eastern Australia. Part 7. *Leptospermum petersonii*, *L. liversidgei* and allies. Flavor Frag. J. 15, 342–351.

Butcher, P.A., Matheson, A.C., Slee, M.U., 1996. Potential for genetic improvement of oil production in *Melaleuca alternifolia* and *M. linariifolia*. New Forests 11, 31–51.

Chen, F., Ro, D., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E., Tholl, D., 2004. Characterization of a root-specific arabidopsis terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. Plant Physiol. 135, 1956–1966.

Coppen, J.J.W., 2002. Eucalyptus: The Genus Eucalyptus. Taylor and Francis, London.

Denson, J., Xi, Z.Y., Wu, Y.C., Yang, W.J., Neale, G., Zhang, J.O., 2006. Screening for inter-individual splicing differences in human GSTM4 and the discovery of a single nucleotide substitution related to the tandem skipping of two exons. Gene 379, 148–155.

Dudareva, N., Cseke, L., Blanc, V.M., Pichersky, E., 1996. Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. Plant Cell 8, 1137–1148.

Edwards, P.B., Wanjura, W.J., Brown, W.V., 1993. Selective herbivory by Christmas Beetles in response to intraspecific variation in *Eucalyptus* terpenoids. Oecologia 95, 551–557.

Emanuelsson, O., Nielsen, H., Von Heijne, G., 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8, 978–984.

Gang, D.R., 2005. Evolution of flavors and scents. Annu. Rev. Plant Biol. 56, 301–325.

Hastings, M.L., Krainer, A.R., 2001. Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. 13, 302–309.

Jones, T.H., Potts, B.M., Vaillancourt, R.E., Davies, N.W., 2002. Genetic resistance of *Eucalyptus globulus* to autumn gum moth defoliation and the role of cuticular waxes. Can. J. Forest Res. 32, 1961–1969.

Kampranis, S.C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N.A., Anssour, S., Dunwell, J.M., Degenhardt, J., Makris, A.M., Goodenough, P.W., Johnson, C.B., 2007. Plant Cell 19, 1994–2005.

Keszei, A., Brubaker, C.L., Foley, W.J., 2008. A molecular perspective on terpene formation in Australian Myrtaceae. Aust. J. Bot. 56, 197–213.

Köllner, T.G., Schnee, C., Gershenzon, J., Degenhardt, J., 2004. The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell 16, 1113–1115.

Lawler, I.R., Foley, W.J., Eschler, B.M., Pass, D.M., Handasyde, K., 1998. Intraspecific variation in Eucalyptus secondary metabolites determines food intake by folivorous marsupials. Oecologia 116, 160–169.

Lücker, J., El Tamer, M.K., Schwab, W., Verstappen, F.W.A., van der Plas, L.H.W., Bouwmeester, H.J., Verhoeven, H.A., 2002. Monoterpene biosynthesis in lemon (Citrus limon). CDNA isolation and functional analysis of four monoterpene synthases. FEBS J. 269, 3160–3171.

Maciolek, N.L., Alward, W.L.M., Murray, J.C., Semina, E.V., McNally, M.T., 2006. Analysis of RNA splicing defects in PITX2 mutants supports a gene dosage model of Axenfeld-Rieger syndrome. BMC Med. Genet. 7, 59.

Martin, D.M., Fäldt, J., Bohlmann, J., 2004. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. Plant Physiol. 135, 1908–1927.

Mine, M., Brivet, M., Touati, G., Grabowski, P., Abitbol, M., Marsac, C., 2003. Splicing error in E1 alpha pyruvate dehydrogenase mRNA caused by novel intronic mutation responsible for lactic acidosis and mental retardation. J. Biol. Chem. 278, 11768–11772.

Moore, B.D., Wallis, I.R., Pala-Paul, J., Brophy, J.J., Willis, R.H., Foley, W.J., 2004. Antiherbivore chemistry of *Eucalyptus* – Cues and deterrents for marsupial folivores. J. Chem. Ecol. 30, 1743–1769.

Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G., Dudareva, N., 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J. 55, 224–239.

Schwab, W., 2003. Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62, 837–849.

Sharkey, T.D., Yeh, S., Wiberley, A.E., Falbel, T.G., Gong, D., Fernandez, D.E., 2005. Evolution of the isoprene biosynthetic pathway in kudzu. Plant Physiol. 137, 700–712.

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

Shelton, D., Zabaras, D., Chohan, S., Wyllie, S.G., Baverstock, P., Leach, D., Henry, R., 2004. Isolation and partial characterisation of a putative monoterpene synthase from *Melaleuca alternifolia*. Plant Physiol. Biochem. 42, 875–882.

Shepherd, M., Chaparro, J.X., Teasdale, R., 1999. Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. Theor. Appl. Genet. 99, 1207–1215.

Spraul, M.H., Nitz, S., Drawert, F., 1991. Photochemical-reactions of p-mentha-1, 3, 8-triene and structural related p-menthadienes. Tetrahedron 47, 3037–3044.

Stamatakis, A., Ludwig, T., Meier, H., 2005. RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic trees. Concurrency-Pract. Ex. 17, 1705–1723.

Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E., Potts, B.M., 2002. Higher-level relationships among the Eucalypts are resolved by ITS-sequence data. Aust. Syst. Bot. 15, 49–62.

Suzuki, Y., Hibino, T., Kawazu, T., Wada, T., Kihara, T., Koyama, H., 2003. Extraction of total RNA from leaves of *Eucalyptus* and other woody and herbaceous plants using sodium isoascorbate. Biotechniques 34, 988.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680.

Thumma, B.R., Nolan, M.R., Evans, R., Moran, G.F., 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp.. Genetics 171, 1257–1265.

Toyota, M., Koyama, H., Mizutani, M., Asakawa, Y., 1996. (−)-ent-Spathulenol isolated from liverworts is an artefact. Phytochemistry 41, 1347–1350.

Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R.B., Christianson, D.W., 2002. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc. Natl. Acad. Sci. USA 99, 15375–15380.

Wildung, M.R., Croteau, R.B., 2005. Genetic engineering of peppermint for improved essential oil composition and yield. Transgenic Res. 14, 365–372.

Wilson, P.G., O'Brien, M.M., Heslewood, M.M., Quinn, C.J., 2005. Relationships within Myrtaceae sensu lato based on a matK phylogeny. Plant Syst. Evol. 251, 3–19.