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*.

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