Phytochemistry xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Phytochemistry

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

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

Jessie Moniodis [a,b,c,*], Christopher G. Jones [a], E. Liz Barbour [a], Julie A. Plummer [a], Emilio L. Ghisalberti [b], Joerg Bohlmann [c]

[a] School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley, WA 6009, Australia
[b] School of Chemistry and Biochemistry (M310), University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia
[c] Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T1Z4, Canada

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Sandalwood
Santalum spicatum
Santalaceae
Fragrance oil
Sesquiterpene
α-Santalol
β-Santalol
E,E-farnesol

ABSTRACT

The fragrant heartwood oil of West Australian sandalwood (*Santalum spicatum*) contains a mixture of sesquiterpene olefins and alcohols, including variable levels of the valuable sesquiterpene alcohols, α- and β-santalol, and often high levels of *E,E*-farnesol. Transcriptome analysis revealed sequences for a nearly complete set of genes of the sesquiterpenoid biosynthetic pathway in this commercially valuable sandalwood species. Transcriptome sequences were produced from heartwood xylem tissue of a farnesol-rich individual tree. From the assembly of 12,537 contigs, seven different terpene synthases (TPSs), several cytochromes P450, and allylic phosphatases were identified, as well as transcripts of the mevalonic acid and methylerythritol phosphate pathways. Five of the *S. spicatum* TPS sequences were previously unknown. The full-length cDNA of *Sspi*TPS4 was cloned and the enzyme functionally characterized as a multi-product sesquisabinene B synthase, which complements previous characterization of santalene and bisabolol synthases in *S. spicatum*. While *Sspi*TPS4 and previously cloned sandalwood TPSs do not explain the prevalence of *E,E*-farnesol in *S. spicatum*, the genes identified in this and previous work can form a basis for future studies on natural variation of sandalwood terpenoid oil profiles.

1. Introduction

The heartwood of several members of the sandalwood genus (*Santalum*, Santalaceae) is highly prized for the sesquiterpene-rich oil of their mature trees (Adams, 1995; Brennan and Merlin, 1993). *Santalum spicatum* is an important oil-bearing species, which has contributed substantially to the economic development of Western Australia. Major essential oil components include α- and β-santalol (1 and 2, see Fig. S1 for structures), α-bisabolol (7) and *E,E*-farnesol (8) along with olefin components such as santalenes (5 and 6), dendrolasin (9), sesquisabinene (10) and sesquiphellandrene (11). Variation in oil composition is present across the natural range of distribution of *S. spicatum* (Moretta, 2001). This variation and the higher levels of *E,E*-farnesol (8) contribute to the lower value of *S. spicatum* oil compared to *Santalum album* oil in the fragrance industry, where the preferred quality of sandalwood oils is largely defined by high levels of α- and β-santalol (1 and 2) (Howes et al., 2004; Verghese et al., 1990). Elucidation of sesquiterpenoid biosynthesis in sandalwood provides a basis to understand composition of oil quality and its variability. This in turn may support efforts towards tree improvement for oil yields and quality in sustainable sandalwood plantations and can afford opportunities for biotechnological sandalwood oil production. In nature, terpenoids are thought to protect sandalwood trees, as heartwood extractives are active against fungi and bacteria (Hammer et al., 1998; Jirovetz et al., 2006).

In plants, terpenoids are produced by terpene synthases (TPSs) from linear prenyl diphosphates, geranyl diphosphate (GPP) for monoterpenes, farnesyl diphosphate (FPP) for sesquiterpenes and geranylgeranyl diphosphate (GGPP) for diterpenes (Chen et al., 2011). GPP, FPP and GGPP are derived from the condensation of isopentyl diphosphate (IPP) (21) and one, two or three molecules, respectively, of dimethylallyl diphosphate (22) (DMAPP). DMAPP (21) and IPP (22) are produced from two pathways, the mevalonate (MEV) pathway and the methylerythritol phosphate (MEP) pathway (Fig. S2). Chemical diversity of terpenoids in plants results from a manifold different rearrangements of carbocation intermediates in the reactions catalyzed by TPSs (Davis and Croteau, 2000; Degenhardt et al., 2009). Many TPSs are multi-product enzymes

* Corresponding author at: School of Plant Biology (M084), University of Western Australia (UWA), 35 Stirling Hwy, Crawley, WA 6009, Australia. Tel.: +61 8 6488 1782; fax: +61 8 6488 1108.
E-mail address: jessie.moniodis@research.uwa.edu.au (J. Moniodis).

http://dx.doi.org/10.1016/j.phytochem.2014.12.009
0031-9422/

and variations of as little as a single amino acid substitution in conserved regions of TPSs can alter product profiles. Plant genomes typically contain families of many similar yet functionally diverse TPSs (Chen et al., 2011). Several cytochrome P450 (P450) family enzymes produce additional alcohol functionalities of sandalwood terpenoids, such as the santalols and bisabolols, and contribute another level of terpenoid diversity (Diaz-Chavez et al., 2013). Variations in the genomic and biochemical makeup as well as expression variations of TPSs and P450s may explain some of the variations of sesquiterpenoids present in natural *S. spicatum* populations.

Substantial progress has been made on the elucidation of TPSs and P450s of sesquiterpene biosynthesis in three different *Santalum* species, *Santalum spicatum*, *S. album* and *Santalum austrocaledonicum*. The primary focus has been on the santalols and bisabolols, while biosynthesis of several other less abundant bicyclic sesquiterpenes has also been explained (Diaz-Chavez et al., 2013; Jones et al., 2008, 2011). TPS cDNAs encoding santalene synthases are highly conserved with 94–96% sequence identity across the three species, which was reflected in the conservation of enzyme functions indicative of an important ecological role in the *Santalum* genus. In contrast, a bisabolol synthase-like TPS showed substantial sequence and functional variations, producing α-bisabolol (7) in *S. spicatum* and β-bisabolene (12) in *S. album* and *S. austrocaledonicum*. Of these three species, *S. album* appears to be the least chemically diverse (Jones et al., 2008) while substantial differences in composition have been noted for *S. spicatum* and *S. austrocaledonicum* (Butaud et al., 2003; Moretta, 2001; Page et al., 2010).

To establish a larger set of genes that contribute to the biosynthesis and possibly the variation of *S. spicatum* sesquiterpenoids, transcriptome sequencing and mining were used, this being a proven approach for identification of genes of terpenoid biosynthesis in non-model species (Zerbe et al., 2013). Reported herein is gene discovery in the transcriptome of *S. spicatum* heartwood xylem, the main sandalwood oil accumulating tissue, and the identification and characterization of *Spi*TPS4 sesquisabinene B synthase.

2. Results and discussion

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

To identify the core biosynthetic steps of sesquiterpene formation in *S. spicatum*, a transcriptome established by 454-sequencing of RNA from xylem tissue containing ray parenchyma cells, where sandalwood oil is thought to be synthesised, was explored (Jones et al., 2008). The transcriptome library of 489,364 reads was produced from a single farnesol (8)-rich tree and assembled into 12,537 apparently unique contig sequences. Contigs were classified into functional ontology groups (Fig. S3). More than half (55%) of the contigs had matches of known functions in other species. Of these, the majority (28%) were annotated with "cellular" or "metabolic processes". Genes involved in secondary metabolism were found in "metabolic processes" and "response to stimuli" groups, which comprised 14% and 7% of the transcriptome, respectively. The transcriptome included candidate TPS, P450 and allylic phosphatase sequences, as well as sequences for genes of the MEV and MEP pathways (Table 1). All steps of the MEV pathway, except for phosphomevalonate kinase, were found in the heartwood xylem transcriptome. In contrast, only two enzymes of the MEP pathway were represented. These results are consistent with the MEV pathway being the primary route by which sesquiterpene precursors are produced in plants and the heartwood xylem tissue being particularly rich in sesquiterpenoids. The most abundant transcript of the terpenoid pathway in the xylem transcriptome was HMG-CoA reductase 1 (HMGR1) classified based on sequence relatedness with *Arabidopsis thaliana* HMGR1 and HMGR2 (GenBank accession No. AEE35849 and AEC06618, respectively), a critical step for isoprenoid biosynthesis in plants (Chye et al., 1992; Goldstein and Brown, 1990). The MEP pathway, which provides the isoprenoid building blocks for monoterpenes and diterpenes was underrepresented in the transcriptome of *S. spicatum* heartwood matching the low abundance or lack of these compounds.

2.2. Terpene synthases

Seven different TPS-like partial sequences were identified in the xylem transcriptome, including previously characterized *S. spicatum* santalene synthase (*Sspi*SSy) and α-bisabolol synthase (*Sspi*BS) (Jones et al., 2011). Five TPS sequences were new for *S. spicatum* and like other *Santalum* TPSs were phylogenetically associated with the TPS-a and TPS-b subfamilies (Fig. 1) of angiosperm mono- and sesquiterpene synthases (Bohlmann et al., 1998, 1997). The full-length cDNA of *Sspi*TPS4, which was represented with five sequence reads in the transcriptome library, was obtained by 5′-RACE, expressed in *Escherichia coli* and functionally characterized. For the other four new TPS sequences, *Sspi*TPS5, −6, −7 and −8, two of which were present only with one read, full-length cDNAs could not be recovered from the available RNA.

2.3. SspiTPS4 is a multi-product sesquisabinene B synthase

The predicted protein encoded by *Sspi*TPS4 showed high amino acid homology to a sesquisabinene synthase from *S. album* (93% similarity, 87% identity; GenBank accession number ADP37190) (Jones et al., 2011). *Sspi*TPS4 encodes a 566 amino acid protein, which clustered with the TPS-b subfamily. Like *Sspi*SSy and *Sspi*BS (Jones et al., 2011), *Sspi*TPS4 did not contain an N-terminal transit peptide for plastidial targeting. *Sspi*TPS4 contains the R(R/P)X_8W motif which is conserved in mono- and sesquiterpene synthases near the N-terminus and the aspartate rich (DDxxD) metal ion-binding region of the active site domain (Fig. S4). A genomic sequence of *Sspi*TPS4 of 3,405 bps with a gene structure of 6 introns and 7 exons was obtained similar to other TPS-b genes (Fig. S5). Intron/exon boundaries were typical of TPS genes with a 3′ ∼ NGT, 5′ AGN∼ pattern.

The recombinant *Sspi*TPS4 protein expressed in *E. coli* had a molecular mass of 66 kDa similar to other sesquiterpene synthases. *Spi*TPS4 converted GPP and FPP substrates to mono- and sesquiterpene products, respectively. However, only the sesquiterpene products were constituents of *S. spicatum* oil. *Sspi*TPS4 was most active with *E,E*-FPP and Mg^{2+} producing a profile of six main sesquiterpenes (10–14) (Fig. 2). The most abundant product was sesquisabinene B (10) (58%) in addition to β-bisabolene (12) (18%), γ-curcumene (13) (12%), β-sesquiphellandrene (11) (9%), α-acoradiene (14) (2%) and another unidentified component (1%) (Table 2). No activity was detected when *Sspi*TPS4 was assayed with *E,Z*-FPP. Likewise, assays with Mn^{2+} instead of Mg^{2+} showed no activity, which is in agreement with the known preference of sesquiterpene synthases for Mn^{2+} instead of Mg^{2+} and in contrast to monoterpene synthases which typically prefer Mn^{2+} (Bohlmann et al., 1998). When *Sspi*TPS4 was incubated with GPP and Mg^{2+}, six different products were formed; β-pinene (15) (24%), myrcene (16) (21%), sabinene (17) (18%), α-pinene (18) (16%), α-terpineol (19) (11%) and linalool (20) (10%). The same compounds were produced when Mn^{2+} was replaced with Mg^{2+}, however levels of product formation were markedly lower (Fig. 3). Monoterpenes have been reported from sandalwood only in very low concentrations (Valder et al., 2003). Lack of the plastid targeting sequence suggests that *Sspi*TPS4 is active in the cytosol, where it can access FPP (23) as a substrate, while GPP (24) is

Please cite this article in press as: Moniodis, J., et al. The transcriptome of sesquiterpenoid biosynthesis in heartwood xylem of Western Australian sandalwood (*Santalum spicatum*). Phytochemistry (2015), http://dx.doi.org/10.1016/j.phytochem.2014.12.009