Biosynthesis of tovarol and other sesquiterpenoids in Thapsia laciniata Rouy

Andersen, Trine Bundgaard; Rasmussen, Silas Anselm; Christensen, Søren Brøgger; Simonsen, Henrik Toft

Published in:
Phytochemistry

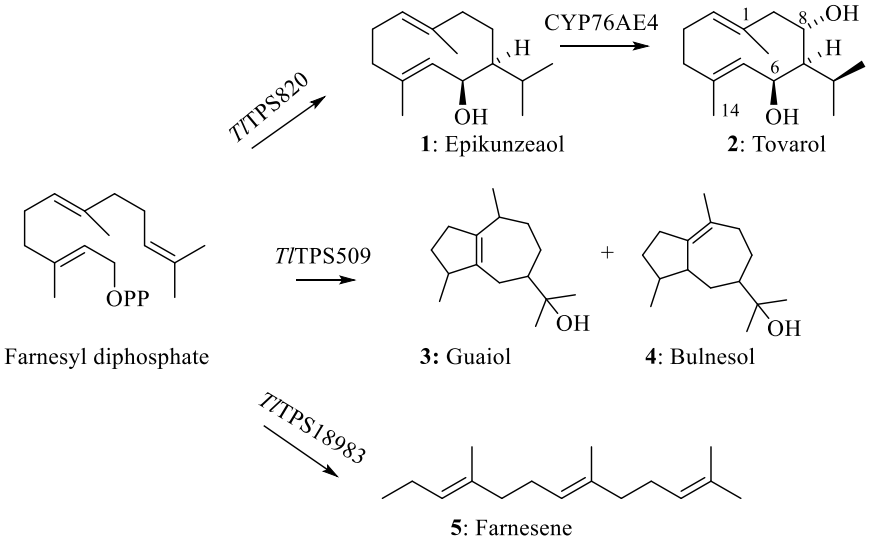
Link to article, DOI:
10.1016/j.phytochem.2018.10.027

Publication date:
2019

Document Version
Peer reviewed version

Graphical abstract

The biosynthesis of sesquiterpenoids in *Thapsia laciniata* Rouy was investigated, and three terpene synthases was characterized.



Biosynthesis of tovarol and other sesquiterpenoids in *Thapsia laciniata* Rouy

Trine Bundgaard Andersen[1], Silas Anselm Rasmussen[2], Søren Brøgger Christensen[3], Henrik Toft Simonsen[2,*]

1: Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

2: Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kgs. Lyngby, Denmark

3: Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark

*corresponding author: hets@dtu.dk

Abstract

The genus *Thapsia* produces a wide variety of sesquiterpenoids. The Mediterranean plant *Thapsia laciniata* Rouy is known to have a product profile that differs from several other species in the genus. Thus, the biosynthesis of sesquiterpenoids in *Thapsia laciniata* Rouy was investigated. Here we describe three terpene synthases, *Tl*TPS820, *Tl*TPS509 and *Tl*TPS18983. *Tl*TPS18983 is a multi-product enzyme with farnesene as the major product, while *Tl*TPS509 produces guaiol and bulnesol along with other major and several minor unknown products. *Tl*TPS820 is orthologous to *Tg*TPS2 from *Thapsia garganica* L. and is an epikunzeaol synthase. *Tg*CYP76AE2 from *Thapsia garganica* performs a triple hydroxylation of epikunzeaol at C-13 to make dihydrocostunolide. It was therefore investigated if the cytochrome P450, *Tl*CYP76AE4 was able to use epikunzeaol as a substrate. It was found that *Tl*CYP76AE4 hydroxylates epikunzeaol at C-8 to yield tovarol instead of dihydrocostunolide.

Keywords

Thapsia laciniata, Apiaceae, sesquiterpenoid, CYP76AE4, epikunzeaol, tovarol, guaiol, bulnesol

1. Introduction

The Apiaceae family is known to include species of great nutritional and pharmacological value, such as *Daucus carota* L. (carrot) and *Thapsia garganica* L. (deadly carrot). Species from the genus *Thapsia* have been used in traditional medicine for millennia for a variety of ailments including rheumatism, insect bites, acne, colds and for purging (Andersen et al., 2015b). The presence of biologically active compounds, several of which are sesquiterpenoids, may justify the traditional use (Simonsen et al., 2013). While many species in Apiaceae have been studied for their chemical composition, only a few genes involved in biosynthesis of sesquiterpenoids have been characterized. The initial enzymes involved in the specialized biosynthesis of sesquiterpenoids are the sesquiterpene synthases (sqTPS). A few sqTPSs have been characterized in Apiaceae, these include a sqTPS from *D. carota* producing (*E*)-*β*-caryophyllene and *α*-humulene (Yahyaa et al., 2015), a δ-cadinene synthase (*Tg*TPS1) and a epikunzeaol synthase (*Tg*TPS2) from *T. garganica* (Pickel et al., 2012) and the germacrene D synthase from *Thapsia laciniata* (Andersen et al., 2015a). Of these, the epikunzeaol synthase *Tg*TPS2 from *T. garganica* has been of particular interest. The structure of epikunzeaol makes this a likely precursor for thapsigargin. Thapsigargin is a potent inhibitor of the mammalian sarco-endoplasmic reticulum Ca^{2+}-ATPase (Thastrup et al., 1990) with a complex structure that requires several unknown biosynthetic steps. The recently discovered cytochrome P450, *Tg*CYP76AE2, from *T. garganica* was found to make a triple hydroxylation of epikunzeaol, thereby oxidizing C-13 to a carboxylic acid, which enabled formation of the lactone ring present in thapsigargin (Andersen et al., 2017). While several species in the *Thapsia* genus produce sesquiterpene lactones, these specific type of compounds are not found in *T. laciniata* (Drew et al., 2012).

The *Thapsia* genus is still in need of revision and several species are unresolved. *T. laciniata* Rouy has also been known as *Thapsia villosa* type II L. and is still listed as unresolved on http://www.theplantlist.org/. Studies of marker sequences in the *Thapsia* genus, however, support the assignment of *T. laciniata* as an independent species that is distinct from the classic *T. villosa* (Weitzel et al., 2014). *T. laciniata* is commonly found in the southern parts of France and in Catalonia, Spain. A variety of monoterpenoids and sesquiterpenoids have been reported from this species, with guaiols and germacrenols as the major sesquiterpenoids in the roots (Drew et al., 2012). In *T. laciniata* the only enzyme described so far is the germacrene D synthase, *Tl*TPS7414 (Andersen et al., 2015a). Here, we present the finding of an orthologue of *Tg*TPS2 namely *Tl*TPS820 that also produces epikunzeaol. Two additional sqTPSs, *Tl*TPS18983 and *Tl*TPS509 give further insight into the diverse sesquiterpenoid profile in *T. laciniata*. In *T. laciniata* there is no equivalent sequence to *Tg*CYP76AE2, which explains the lack of sesquiterpene lactones. The orthologue *Tl*CYP76AE4 instead catalyzes the hydroxylation of epikunzeaol at C-8 to yield tovarol.

2. Results

2.1. Identification of genes

A homology-based search using BLAST was initiated to investigate if *T. garganica* and *T. laciniata* harbored orthologue genes for sesquiterpenoid biosynthesis. *Tg*TPS1 and *Tg*TPS2 from *T. garganica* were used as baits in the *T. laciniata* Illumina transcriptome dataset (Drew et al., 2013). For *Tg*TPS2, a gene with 97.2% nucleotide identity was discovered. The homology search for *Tg*TPS1 did not yield any hits over 90% in sequence identity. Three additional sqTPSs were found; the germacrene D synthase, *Tl*TPS7414 (Andersen et al., 2015a) and the unknown *Tl*TPS509 and *Tl*TPS18983 (Drew et al., 2013). The data set was also mined for cytochrome P450s with high sequence similarity to *Tg*CYP76AE2, which is the gene responsible for sesquiterpene lactone formation in *T. garganica*. One gene was found with 92% amino acid sequence identity and 96% nucleotide sequence identity to *Tg*CYP76AE2. The gene was named *Tl*CYP76AE4 (Nelson, 2009). All genes were cloned into the pEAQ vector system for transformation into *A. tumefaciens* (Peyret and Lomonossoff, 2013). *Nicotiana benthamiana* was used as an expression host to examine *Tl*TPS820, *Tl*TPS18983, *Tl*TPS509 and *Tl*CYP76AE4. Samples were extracted with either hexane or by HS-SPME depending of the volatility of the enzymatic products.

2.2. Characterization of *Tl*TPS820 expressed in *N. benthamiana*

Leaves from *N. benthamiana* expressing *Tl*TPS820 were extracted with hexane and analyzed by GC-MS. The analysis yielded a sesquiterpenoid profile identical to *Tg*TPS2 (Figure 1). The main product of *Tg*TPS2 is epikunzeaol (**1**) (Figure 2), thus *Tl*TPS820 is also an epikunzeaol synthase and an orthologue of *Tg*TPS2. As described previously, the high injection port temperature at 250°C lead to cope-rearrangement of **1** (Andersen et al., 2017). Thus, **1** was partly degraded to two unknown alcohols.

2.3. Characterization of *Tl*TPS509 expressed in *N. benthamiana*

HS-SPME-GC-MS analysis of *N. benthamiana* leaves expressing *Tl*TPS509 resulted in the detection of two major sesquiterpene alcohols, guaiol (**3**) (RI: 1585, 34% of the area) and bulnesol (**4**) (RI: 1654, 20% of the area) (Figure 3). Compound **3** and **4** exhibited similar EI MS spectrum to those of bulnesol and guaiol in NIST and Wiley GC-MS spectrum libraries. The product profile showed several additional compounds, and we could identify two minor products; guaiene (RI 1418, Rt: 8.6)) and bulnesene (RI 1487, Rt: 9.18) (Figure 3). Several peaks with a *m/z* of 220 were also observed but not identified. A hexane extract of *N. benthamiana* leaves, expressing *Tl*TPS509 was analyzed by GC-MS using a Programmed Temperature Vaporising injector (PTV) port. By using a low initial temperature of 60°C followed by a gradual increase to 250°C it was expected that thermal rearrangements were avoided. Again, **3** and **4** were detected as major compounds, but not guaiene and bulnesene, which were probably below the detection limit. Furthermore, neither of the unknown compounds with *m/z* 220 were detected in the PTV analysis.

2.3.1. Isolation and structure elucidation of the *Tl*TPS509 products, bulnesol and guaiol

In order to verify the identity of **3** and **4** by NMR, the compounds were isolated in two purification steps. An initial purification of the compounds was performed on a silica column. **3** and **4** both eluted with 10% ethyl acetate in hexane and this fraction was further fractionated on a preparative GC-MS and **3** and **4** were separated and both precipitated as a colorless oil in the fraction collector. The ^{1}H and ^{13}C NMR resonance values of **3** and **4** were all in accordance with those reported by Raharivelomanana et al., (1995), which confirm the identity of **3** and **4**. The spectra are provided in Supplementary data figure 12 and 13.

2.4. Characterization of *Tl*TPS18983 expressed in *N. benthamiana*

Volatiles from *N. benthamiana* leaves expressing *Tl*TPS18983 were captured by HS-SPME and analyzed by GC-MS. The analysis of these plants showed that expression of *Tl*TPS18983 generate at least 15 products. The identification of one of the major products, farnesene (**5**), was based on comparison of the mass spectrum of the compound with NIST and Wiley GC-MS spectrum libraries (Figure 4, MS spectrum is provided in supplementary data Figure 3). It was not possible to identify the remaining products. Using an injection port temperature of 160 °C instead of 250 °C did not alter the composition of the chromatogram, which supports that these compounds are not a result of the GC-MS analysis method. Extraction of the products was also attempted using organic solvents such as hexane and pentane mimicking procedures that were successful for other sqTPS products. However, for *Tl*TPS18983 we did not detect any compounds by GC-MS using organic solvent extraction.

2.5. Characterization of *Tl*CYP76AE4 using epikunzeaol from *Tl*TPS820 as substrate

To investigate if *Tl*CYP76AE4 would be able to use **1** as a substrate, *Tl*TPS820 and *Tl*CYP76AE4 were co-expressed in *N. benthamiana*. This resulted in complete conversion of **1** as seen in Figure 1. One new product was detected in the GC-MS chromatogram with an injection port temperature of 250°C. However, the peak appeared to consist of two compounds that were co-eluting on the GC-MS. It was not possible to separate these products further. The GC-MS analysis indicated that the parent ions had a *m/z* of 238. To further support that *Tl*TPS820 yields **1**, *Tl*CYP76AE4 was also expressed with *Tg*TPS2 and was shown to yield the same new product(s) (supplementary data Figure 1). From the chromatograms and spectrums, it was clear that *Tl*CYP76AE4 yielded a new product. Co-expression of *Tg*TPS2, *Tg*CYP76AE2 and *Tl*CYP76AE4 did not result in additional new products.

2.5.1.Isolation and structure elucidation of tovarol

Compound **2** was isolated from the hexane extract of leaves of *N. benthamiana* expressing *Tl*TPS820 and *Tl*CYP76AE4. The isolation was performed by semi-preparative normal-phase HPLC and the purity of **2** was confirmed by LC-HRMS. High resolution mass spectrometry (HRMS) established the accurate mass of **2** to *m/z* 261.1826 ($[M+Na]^+$) to which the molecular formula was deduced to be $C_{15}H_{26}O_2$ (calcd. for $C_{15}H_{26}O_2$

133 $[M+Na]^+$ *m/z* 261.1825, Δ -0.2 ppm). **2** was structurally elucidated by interpretation of the ^1H, ^{13}C and ^1H-^1H
134 COSY, ^1H-^{13}C ed-HSQC, ^1H-^{13}C HMBC and ^1H-^1H ROESY NMR data. The data for **2** were in agreement
135 with the germacradiendiol tovarol (Figure 2), which has previously been isolated from *Thapsia* species (De
136 Pascual Teresa et al., 1985; Teresa et al., 1986). Due to the low quality of the published spectra and the absence
137 of ^{13}C NMR data a full structure elucidation was performed. The DEPT spectrum of the metabolite revealed
138 that **2** contained two quaternary carbon atoms, six methine carbon atoms, three methylene groups and four
139 methyl groups. Even though most of the peaks in the recorded spectrum were broad, presumably because of a
140 conformational equilibrium (Fischer et al., 1979; Triana et al., 2005), an in-depth investigation was possible
141 (Table 1). The use of 2D spectra also enabled an assignment of the signals in the ^{13}C NMR spectrum. The *E,E*-
142 configuration was established through a ROESY correlation between H-1 and H-5. Model building reveals
143 that only the *E,E*-configuration allows H-1 and H-5 to be near in space. ROESY correlation between H-6, H-
144 15 and H-7 revealed that H-6 and H-7 were cis-disposed. Comparison with a reference spectrum of tovarol
145 (Prof. G. Appendino personal communication) revealed that previous isolated tovarol possessed the same
146 stereochemistry at C-8 as **2**, and revealed that the hydroxyl group at C-8 is α-disposed. A ROEYS correlation
147 between H-7 (1.05 ppm) and H-9 resonating at 1.78 ppm revealed that this H-9 must be α-disposed and
148 consequently the signal at 2.49 ppm was H-9β. The absolute configuration was assumed based on the
149 observation that sesquiterpenes from Apiaceae plants have H-7 α-disposed (Fischer et al., 1979; Simonsen et
150 al., 2013). **2** is therefore concluded to be tovarol (De Pascual Teresa et al., 1985).

151 ## 3. Discussion

152 A variety of sesquiterpenoids are produced in *Thapsia* and the sesquiterpenoid profile often differs between
153 species (Christensen et al., 1997). Previous analysis of *T. garganica* and *T. laciniata* have shown that these
154 species produce sesquiterpenoids with different backbones and side-group modifications (Drew et al., 2012).
155 *T. laciniata* has at least four functional sesquiterpene synthases and two of them, *Tl*TPS18983 and *Tl*TPS509,
156 are multi-product synthases. Compared to the two sqTPS found in *T. garganica* this correlates well with the
157 higher number of metabolites described in *T. laciniata* (Christensen et al., 1997; Drew et al., 2012, 2013). In
158 *Daucus carota* 20 sqTPS (TPS-a family) are predicted, which show that even between closely related species
159 the number of sqTPS's can vary tremendously (Banasiak et al., 2016; Keilwagen et al., 2017).

160 Several farnesene synthases have been described in literature. *Tl*18983 is however the first farnesene synthase
161 from Apiaceae. The finding of a farnesene synthase is supported by the previous detection of β-farnesene in
162 *T. laciniata* (Drew et al., 2012). It is interesting that *Tl*18983 is able to produce a large number of products,
163 which cannot merely be characterized as minor side products, at least judging by expression in *N. benthamiana*.
164 Analyzing a multiproduct sqTPS is complex due to the difficulty of separating the products by GC-MS, and
165 due to very volatile nature of these compounds. Sesquiterpenoids in many species are a part of the volatile
166 mixture of compounds that is emitted by the plants. It can be hypothesized that the products produced by

*Tl*TPS18983 are a part of the volatile signaling compounds of *T. laciniata,* as also seen in other plants (Flamini et al., 2003; Máday et al., 1999; Wang et al., 2015).

Two major products of *Tl*TPS509 were identified as guaiol (**3**) and bulnesol (**4**). Previous investigation of *T. laciniata* sesquiterpenoids (Drew et al., 2012; Lemmich et al., 1984) also showed that **3** is produced by this species. Three guaiane esters, (4*S*, 5*S*, 7*S*, 8*S*)-8-p-Coumaroyloxy-1(10)-guaiene-11-ol, (4*S*, 5*S*, 7*S*, 8*S*)-8-Feruloyloxy-1(10)-guaiene-11-ol, and (4*S*, 5*S*, 7*S*, 8*S*)-8-Senecioyloxy-1(10)-guaien-11-ol have also been found in *T. laciniata* (Lemmich et al., 1991). All of these are derived from the structure of **4** and have different side chain modifications at C-8. This indicates that bulnesol is a possible precursor of these three structures. The described modifications at C-8 are expected to involve a hydroxylation at C-8 followed by the addition of side chains by acyl transferases. Currently, *T. laciniata* is the only species in the genus *Thapsia* that has been found to produce guaiols (Christensen et al., 1997). Furthermore, α-guaiene was previously found in *T. laciniata* (Drew et al., 2012). This is a minor product of *Tl*TPS509 and is the non-hydroxylated compound with the same type of 5- and 7-ring structure as guaiols. The same is also observed for bulnesene and bulnesol. Co-expression of CYP76AE4 and *Tl*TPS509 did not provide any new products. Thus, an orthologue of CYP76AE4 might be responsible for the hydroxylation at C-8, though this remains to be examined.

*Tl*TPS820 produces epikunzeaol (**1**) as a main product. The discovery of the two orthologues sqTPS genes (*Tg*TPS2 and *Tl*TPS820) in *T. laciniata* and *T. garganica* shows that the epikunzeaol synthase can serve as a marker for the *Thapsia* genus. It also shows that the species within *Thapsia* have evolved different downstream products by using **1** as a substrate. While *Tg*CYP76AE2 and *Tl*CYP76AE4 are both able to use **1** as a substrate, they perform hydroxylation on different carbon positions. This difference is also reflected in the metabolite profiles of the two species. In contrast to most *Thapsia* species, *T. laciniata* does not contain guaianolides, for which epidihydrocostunolide (the product of *Tg*CYP76AE2 in *T. garganica*) appears to be a precursor (Andersen et al., 2017). Tovarol has currently not been detected from *T. laciniata* and could be involved in the biosynthesis of a downstream product yet to be discovered. However, tovarol and several derivatives was isolated from the closely related *Thapsia minor* Hoffmanns. & Link (De Pascual Teresa et al., 1985). It is likely that similar biosynthetic routes are present in *T. laciniata* and *T. minor* since these two species are very closely related (Banasiak et al., 2016; Weitzel et al., 2014). It remains to be investigated if *T. minor* harbors orthologue genes to those described here and in *T. garganica*.

The products detected by GC-MS from *Tg*CYP76AE2 were shown to be Cope re-arrangements of dihydrocostunolide (Andersen et al., 2017). It is highly probable that **2** from *Tl*CYP76AE4 has also undergone Cope re-arrangement during exposure to the high injection port temperature in the GC-MS (Figure 1). Through the isolation of **2** it was clear from our LC data that *Tl*CYP76AE4 only produces **2** as the product.

The expression of genes, involved in sesquiterpene biosynthesis in *N. benthamiana* has previously been reported to result in the conjugation of produced sesquiterpenoids. This was shown for artemisinic acid in the artemisinin pathway, which was conjugated to a diglucoside (van Herpen et al., 2010). In costunolide biosynthesis, the analysis was affected by the conjugation of costunolide to glutathione or cysteine (Liu et al., 2011). LC-MS analysis of **2** from *Tl*CYP76AE4 did not reveal any conjugation to the two alcohol groups, which therefore allowed for detection by GC-MS as well. **1** on the other hand from *Tl*TPS820 and *Tg*TPS2 was found both in its free form in GC-MS analysis and as a disaccharide in LC-MS analysis.

4. Conclusion

We have described the activity of three sesquiterpene synthases from *Thapsia laciniata*. *Tl*TPS820 was found to be an epikunzeaol synthase whereas *Tl*TPS18983 and *Tl*TPS509 are multi-product sesquiterpene synthases, with farnesene as the major product of *Tl*TPS18983 and guaiol and bulnesol as the major products of *Tl*TPS509. *Tl*TPS820 is an orthologue to *Tg*TPS2 from *Thapsia garganica* and co-expression with the cytochrome P450 *Tl*CYP76AE4 led to the biosynthesis of tovarol. Thus, we have shown that *Tl*CYP76AE4 hydroxylate epikunzeaol at C-8 to yield tovarol.

5. Experimental section

5.1. Plant material

Thapsia laciniata Rouy, Apiaceae (Banasiak et al., 2016; Weitzel et al., 2014) roots were collected in early June 2008, just west of Cannes, France (GPS 43.540958, 6.816158). Voucher specimens of *T. laciniata* (HTS 2008-01) are stored at herbarium C (Natural History Museum of Denmark, Copenhagen).

5.2. Identification and cloning of genes

The sesquiterpene synthases *Tl*TPS18983 (MG680745), *Tl*TPS509 (MG680746) and *Tl*TPS820 (MG680744) were found in a transcriptome dataset from *T. laciniata* root (SRX252523) (Drew et al., 2013). The full-length sequences were obtained from a cDNA library from *T. laciniata* root material. The discovery of *Tg*TPS2 (JQ290345) and *Tg*CYP76AE2 (KX826943) has been presented previously (Andersen et al., 2017; Pickel et al., 2012). *Tg*CYP76AE2 was blasted into the transcriptome dataset to search for a homologues sequence. The resulting P450 was named, *Tl*CYP76AE4 (MG680747) (Nelson, 2009).

Forward and reverse primers for all genes were designed with USER-overhangs, to enable cloning into a pEAQ-USER compatible version of the pEAQ-*HT* vector (Supplementary data, table 1) (Luo et al., 2016).

pEAQ-*HT* harbors the viral suppressor p19 and was kindly provided by George Lomonosonoff (John Innes Research Centre, Norwich, UK) (Peyret and Lomonossoff, 2013). USER cloning was performed as previously depicted (Nour-Eldin et al., 2006).

5.3. Expression of *Tl*TPS18983, *Tl*TPS509 *Tg*TPS2, *Tl*TPS802, *Tg*CYP76AE2 and *Tl*CYP76AE4 in *Nicotiana benthamiana*

Nicotiana benthamiana plants were grown from seeds at 24 °C / 19 °C (day/night) for 5 weeks before transformation. The transformation of *Agrobacterium tumefaciens* and infiltration of *N. benthamiana* with *A. tumefaciens* was achieved as previously described (Bach et al., 2014). Plants were harvested 5 days after infiltration. In short, 10 ml LB containing kanamycin, rifampicin and carbenicillin was inoculated with several agrobacterium colonies containing the plasmid of interest. Cultures were grown overnight at 28 °C and 200 rpm. Cell pellets were washed twice with water before a final resuspension in water followed by a dilution to OD_{600} 0.5. Resuspended *A. tumefaciens* carrying plasmids containing *At*tHMGR, *Tl*TPS or *Tl*CYP76AE4 were mixed 1:1:1 and infiltrated into leaves of at least three *N. benthamiana* plants by use of a syringe. Plants were placed at 24 °C / 19 °C (day/night) and harvested five days after infiltration. As controls, plants were infiltrated with *A. tumefaciens* carrying plasmids with *At*tHMG, or *At*tHMGR plus *Tl*TPSs.

The ~100 plants needed for purification of tovarol were infiltrated by use of vacuum. Three *A. tumefaciens* cultures containing *At*tHMGR, *Tl*TPS820 or *Tl*CYP76AE4 were grown overnight at 28 °C and 200 rpm in 500 ml LB (containing kanamycin, rifampicin and carbenicillin) from 20 ml starter cultures. Cell pellets were washed twice with water before final resuspension in water followed by a dilution to OD_{600} 0.5. Resuspended *A. tumefaciens* carrying plasmids containing *At*tHMGR, *Tl*TPS820 or *Tl*CYP76AE4 were mixed 1:1:1. Plants were submerged in a 1 L suspension of *A. tumefaciens* and infiltrated by use of vacuum at 50-100 mbar for 1 min (Andersen-Ranberg et al., 2016).

5.4. GC-MS detection of sesquiterpenoids from liquid extractions

Two leaf discs (diameter 3 cm) from *N. benthamiana*, expressing genes from *T. laciniata* were extracted with 1.2 ml hexane for GC-MS analysis to provide one sample, a minimum of three biological replicates were examined. Samples were analyzed on a Shimadzu GCMS-QP2010. The column used was an Agilent HP-5MS UI, 20 m, 0.18 mm diameter × 0.18 μm film thickness. The pressure was kept at 66.7 kPa giving a column flow of 1 mL/min. The injection port temperature was set to 250 °C. The oven temperature was set to 60 °C for 3 min, and then increased to 160 °C with a rate of 7 °C/min. The temperature was further increased to 300 °C at a rate of 50 °C /min, held for 5 min, finally increased to 320 °C at 50 °C/min and maintained for 3 min. The carrier gas was H_2 and the ionization electron energy was 70 eV. The ion source temperature was 230 °C with an interface temperature 280 °C. The total run time was 28.49 min. All data were analyzed using the

Shimadzu software Lab Solutions, GCMS Solutions version 2.50 SU3, with the 2008 libraries provided by NIST and Wiley.

5.5. HS-SPME-GC-MS analysis of sesquiterpenoids

*Tl*TPS18983 and *Tl*TPS509 were also analyzed on GC-MS by use of HS-SPME fibers. HS-SPME fibers were purchased from Supelco Sigma-Aldrich, Denmark. A whole leaf was placed in a 20 ml glass vial and extracted at 60 °C for 20 min with a SPME fiber exposed to the air above the leaf (Andersen et al., 2015a). Immediately hereafter, the SPME fiber was run on the GC-MS. Samples were analyzed on a Shimadzu GCMS-QP2010. The column used was an Agilent HP-5MS UI, 20 m, 0.18 mm diameter × 0.18 µm film thickness. The pressure was kept at 66.7 kPa giving a column flow of 1 mL/min. The injection port temperature was set to 160 °C or 250 °C. The oven temperature was set to 60 °C for 3 min, and then increased to 160 °C with a rate of 7 °C/min. The temperature was further increased to 300 °C at a rate of 50 °C/min, held for 5 min, finally increased to 320 °C at 50 °C/min and maintained for 3 min. The carrier gas was H_2 and the ionization electron energy was 70 eV. The ion source temperature was 230 °C with an interface temperature 280 °C. The total run time was 28.49 min. All data were analyzed using the Shimadzu software Lab Solutions, GCMS Solutions version 2.50 SU3, with the 2008 libraries provided by NIST and Wiley.

5.6. Analytical LC-MS detection of sesquiterpenoids

Two leaf discs (diameter 3 cm) were ground in liquid nitrogen. Samples were extracted with 1200 µl 80 % methanol and sonicated for 30 min. Before LC-MS analysis samples were filtered through a 0.45 µm filter. Analytical LC-MS was carried out using an Agilent 1100 Series LC (Agilent Technologies, Santa Clara, CA, USA) coupled to a Bruker HCT-Ultra ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). Separation was achieved on a Gemini-NX C_{18} column (Phenomenex; 2×150 mm, 3 µm) maintained at 35 °C. The mobile phases were: A, water with 0.1 % (v/v) HCOOH; B, acetonitrile with 0.1 % (v/v) HCOOH. The gradient program was: 0 to 1 min, isocratic 12 % B; 1 to 33 min, linear gradient 12 to 80 % B; 33 to 35 min, linear gradient 80 to 99 % B; 35 to 38 min, isocratic 99 % B; 38-47 min, isocratic 12 % B. The flow rate was 0.2 mL/min. The mass spectrometer was operated in positive mode and the range *m/z* 100-1000 was acquired.

High-resolution mass spectrometry of **2** was acquired on a maXis HD mass spectrometer (Bruker Daltronics, Bremen, Germany) coupled to an Ultima 3000 series HPLC-DAD (Thermo, Waltham, Massachusetts, USA). Separation was achieved on a Kinetex C_{18} column (Phenomenex, Torrance, USA, 150×2 mm, 2.6 µm) using a linear gradient consisting of A, H_2O with 20 mM HCOOH and B, HPLC-MS grade MeCN with 20 mM HCOOH, from 10-100 % B in 10 min with a flow rate of 0.4 ml/min, the column was maintained at 40 °C. The mass spectrometer was operated in positive ESI acquiring the *m/z* range of 75-1250.

5.7. Isolation of guaiol and bulnesol from *Tl*TPS509 by preparative GC-MS

For isolation of guaiol and bulnesol, a large-scale hexane extraction was made from *N. benthamiana* leaves expressing truncated *At*HMGR and *Tl*TPS509. The leaves from approximately 100 five-week-old plants were extracted with hexane. The crude hexane extract was subjected to an initial preparative separation on a silica column and eluted stepwise in hexane-ethyl acetate, from hexane to hexane:ethyl acetate (80:20) with 2 % increments. The final purification of products from *Tl*TPS509 were performed on a preparative GC-MS using an Agilent 7890B GC installed with an Agilent 5977A inert MSD, GERSTEL Preparative Fraction Collector (PFC) AT 6890/7890 and a GERSTEL CIS 4C Bundle injection port. For separation by GC, a RESTEK Rtx-5 column (30m x 0.53mm ID x 1 µm df) with H_2 as the carrier gas was used. A split piece with a split of 1:100 (MS:PFC) was placed at the end of this column. Sufficient amounts of sesquiterpene product for NMR analysis (0.5-1 mg) were obtained with 5 µL injections of extract repeated 100 times. The injection port was put in solvent vent mode with a carrier gas flow of 100 mL/min until minute 0.17, combined with an injection speed of 1.5 mL/min. The purge flow was set to 3 mL/min from minute 0.17 to 2.17. The injection temperature was held at 40 °C for 0.1 min followed by ramping at 12 °C/sec until 320 °C, which was then held for 2 min. The column flow was set to 7.5 mL, which was held constant throughout the GC program. The GC program was set to hold at 60 °C for 1 min, ramp 20 °C/min to 320 °C and 320 °C held for 3 min. The temperature of the transfer line from GC to PFC and the PFC itself was set to 250 °C. The PFC was set to collect the selected peaks based on the MS spectrum. The MS was set in scan mode from *m/z* 35 to *m/z* 500, with a threshold of 150. Solvent cut-off was set to 4 min, and the temperature of the MS source and the MS quadrupole was set to 300 °C and 150 °C, respectively. Traps were kept at -20°C, in case products were prone to rearrangement, and rinsed with chloroform-d (Euriso-top, 99.8 atom % D).

5.8. Isolation of tovarol from *Tl*CYP76AE4 by HPLC

For isolation of tovarol, a large-scale hexane extraction was made from *N. benthamiana* leaves expressing truncated *At*HMGR, *Tl*TPS820 and *Tl*CYP76AE4. The leaves from approximately 100 five-week-old plants were extracted with hexane. The crude hexane extract was subjected to an initial preparative separation on a Isolera autoflasher using a 10 g diol functionalized silica column (Biotage, Uppsala, Sweden) eluted stepwise in hexane-ethyl acetate, from hexane to hexane:ethyl acetate (80:20) with 2 % increments. Final isolation of tovarol was achieved by semi-preparative HPLC utilizing a 250 × 10 mm, 7 µm Nucleosil PEI column (Macherey-Nagel) eluted in a linear hexane-isopropanol gradient from 0-5 % isopropanol in 20 min, on a Waters 600 HPLC equipped with a Waters 996 PDA detector. The purity was confirmed by LC-MS as described in 5.6.

5.9. ^{1}H and ^{13}C NMR spectroscopic analysis

NMR-spectra were acquired using a 600 MHz Bruker Avance III HD NMR spectrometer (^{1}H operating frequency 600.13 MHz) equipped with a Bruker SampleJet sample changer and a cryogenically cooled gradient inverse triple-resonance 1.7-mm TCI probe-head (Bruker Biospin, Rheinstetten, Germany) optimized for ^{13}C and ^{1}H. Samples were analyzed at 300 K. Proton spectra, at 600.03 MHz, were obtained using 30°-pulses, a spectral width of 12 kHz, collecting 16 scans with a length of 65536 data points with a relaxation delay of 1.0 sec. Carbon spectra were acquired at 150.88 MHz with 30°-pulses, a spectral width of 36 kHz, collecting 256 scans with a length of 65536 data points and with a relaxation delay of 2.0 sec. The ^{13}C nuclei were ^{1}H-decoupled using the Waltz-16 composite pulse-decoupling scheme. FID's were exponentially multiplied with a line-broadening factor of 1.0 Hz before Fourier transformation. The 2D experiments were recorded using Bruker standard parameter settings. The isolated tovarol was dissolved in MeCN-d_3 (99.8 atom % D) prior to NMR analysis.

6. Acknowledgements

Trine Bundgaard Andersen, Henrik Toft Simonsen and Søren Brøgger Christensen were supported by SPOTLight, a grant from the Danish Council for Strategic Research. George Lomonosonoff (John Innes Research Centre, Norwich, UK) provided the pEAQ-*HT* plasmid. NMR equipment used in this work was purchased via grant #10-085264 from The Danish Research Council for Independent Research | Nature and Universe. We thank Professori Ordinari Giovanni Battista Appendino, Università degli Studi del Piemonte Orientale, Italy for reference spectra of tovarol.

Figure legends

Figure 1: GC-MS chromatogram of hexane extracts from *N. benthamiana* expressing *At*tHMGR+ *Tl*TPS820, *At*tHMGR+*Tg*TPS2, and *At*HMGR+ *Tl*TPS820 + *Tl*CYP76AE4. The chromatograms show that *Tg*TPS2 and *Tl*TPS820 both produce epikunzeaol (**1**) as previously published for *Tg*TPS2 (Andersen et al., 2017), and with co-expression of *Tl*CYP76AE4 this product disappear and a cope rearranged product of tovarol (**2**) is found. This confirms the functionality of both enzymes.

Figure 2: The biosynthetic routes catalyzed by the four described *Thapsia laciniata* enzymes *Tl*TPS820, *Tl*TPS509, *Tl*TPS18983, and *Tl*CYP76AE4 leading to tovarol (**2**).

Figure 3: HS-SPME GC-MS analysis of *Tl*TPS509 expressed together *At*HMGR, the control shows the expression of *At*HMGR alone. Guaiol (**3**) is seen as the major product (Rt 9.89 min), with bulnesol (**4**), Rt: 10.2, a-guaiene (Rt: 8.6), and a-bulnesene (Rt: 9.18) as minor products. * show unidentified sesquiterpene like products, which is either ketone formed during the biosynthesis or breakdown products of the alcohols formed during the GC-MS analysis. Based on this *Tl*TPS509 is a guaiol synthase. MS spectra are presented in supplementary data

Figure 4: a) HS-SPME GC-MS analysis of *Tl*TPS18983 expressed together *At*HMGR, the control sample show the expression of *At*HMGR alone. farnesene (**5**) is seen as the major product (Rt 8.75 min), * show unidentified sesquiterpene like products. Based on this *Tl*TPS18983 is a multiproduct terpene synthase with farnesene as the major product. MS spectra are presented in supplementary data.

Tables

Table 1: NMR data for tovarol (**2**). The spectrum was recorded in CD_3CN at 600 MHz. Most of the signals are broad, which is probably because of a conformational equilibrium. Numbering is shown in Figure 2.

Position	δ_H mult (*J* Hz)	δ_C type
1	4.96, br.qd (*J* 12.0, 5.1)	131.3 CH
2β	2.39, dd, (*J* 12.0, 5.1)	25.6 CH_2
2α	2.06, ddt, (*J* 12.0, 5.1, 2.0)	
3	2.14, br.dd, (*J* 12.0 ,5.1)	39.7 CH*2*
	2.06, br.ddt, (*J* 12.0, 5.1, 2.0)	
4		132.3 C
5	5.13, br.d (*J* 6.6)	133.6 CH
6	4.57, br.d (*J* 6.6)	69.2 CH
7	1.05, overlaid	54.7 CH
8	4.16, br.ddd, (*J* 11.7, 5.9, 0.7)	73.2 CH
9β	2.49, dd (*J* 11.9, 6.0)	45.6 CH_2
9α	1.78, t (*J* 11.9)	
10		135.5 C
11	2.00 dq, (*J* 10.3, 6.6)	26.4 CH
12	1.02, d, *(*J 6.6)	21.9 CH_3
13	1.05, d, (*J* 6.2) overlaid	21.4 CH_3
14	1.64, s	24.2 CH_3
15	1.42, s	16.8 CH_3
OH	2.14, br.s overlaid	

7. References

Andersen-Ranberg, J., Kongstad, K.T., Nielsen, M.T., Jensen, N.B., Pateraki, I., Bach, S.S., Hamberger, B., Zerbe, P., Staerk, D., Bohlmann, J., Møller, B.L., Hamberger, B., 2016. Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. Angew. Chemie - Int. Ed. 55, 2142–2146.

Andersen, T.B., Cozzi, F., Simonsen, H.T., 2015a. Optimization of biochemical screening methods for volatile and unstable sesquiterpenoids using HS-SPME-GC-MS. Chromatography 2, 277–292.

Andersen, T.B., Lopez, C.Q., Manczak, T., Martinez, K.A., Simonsen, H.T., 2015b. Thapsigargin – from *Thapsia* L. to Mipsagargin. Molecules 20, 6113–6127.

Andersen, T.B., Martinez, K.A., Rasmussen, S.A., Boughton, B.A., Jørgensen, K., Andersen-Ranberg, J., Nyberg, N., Christensen, S.B., Simonsen, H.T., 2017. Localization and *in-vivo* characterization of *Thapsia garganica* CYP76AE2 indicates a role in thapsigargin biosynthesis. Plant Physiol. 174, 56–72.

Bach, S.S., Bassard, J.-É., Andersen-Ranberg, J., Møldrup, M.E., Simonsen, H.T., Hamberger, B., 2014. High-throughput testing of terpenoid biosynthesis candidate genes using transient expression in *Nicotiana benthamiana*, in: Rodríguez-Concepción, M. (Ed.), Plant Isoprenoids, Methods in Molecular Biology. Springer New York, New York, USA, pp. 245–255.

Banasiak, Ł., Wojewódzka, A., Baczyński, J., Reduron, J.-P., Piwczyński, M., Kurzyna-Młynik, R., Gutaker, R., Czarnocka-Cieciura, A., Kosmala-Grzechnik, S., Spalik, K., 2016. Phylogeny of Apiaceae subtribe Daucinae and the taxonomic delineation of its genera. Taxon 65, 563–585.

Christensen, S.B., Andersen, A., Smitt, U.W., 1997. Sesquiterpenoids from *Thapsia* species and medicinal chemistry of the thapsigargins. Fortschritte der Chemie Org. Naturstoffe. Prog. Chem. Org. Nat. Prod. 71, 129–167.

De Pascual Teresa, J., Morán, J.R., Hernández, J.M., Grande, M., 1985. Tovarol and other germacrane derivatives from *Thapsia villosa*. Phytochemistry 24, 1779–1783.

Drew, D.P., Dueholm, B., Weitzel, C., Zhang, Y., Sensen, C.W., Simonsen, H.T., 2013. Transcriptome analysis of *Thapsia laciniata* Rouy provides insights into terpenoid biosynthesis and diversity in Apiaceae. Int. J. Mol. Sci. 14, 9080–9098.

Drew, D.P., Rasmussen, S.K., Avato, P., Simonsen, H.T., 2012. A comparison of headspace solid-phase microextraction and classic hydrodistillation for the identification of volatile constituents from *Thapsia* spp. provides insights into guaianolide biosynthesis in Apiaceae. Phytochem. Anal. 23, 44–51.

399 Fischer, N.H., Olivier, E.J., Fischer, H.D., 1979. The biogenesis and chemsitry of sesquiterpene lactones.
400 Fortschritte der Chemie Org. Naturstoffe. Prog. Chem. Org. Nat. Prod. 38, 47–390.

401 Flamini, G., Cioni, P.L., Morelli, I., 2003. Differences in the fragrances of pollen, leaves, and floral parts of
402 garland (*Chrysanthemum coronarium*) and composition of the essential oils from flowerheads and leaves.
403 J. Agric. Food Chem. 51, 2267–2271.

404 Keilwagen, J., Lehnert, H., Berner, T., Budahn, H., Nothnagel, T., Ulrich, D., Dunemann, F., 2017. The terpene
405 synthase gene family of carrot (*Daucus carota* L.): Identification of QTLs and candidate genes associated
406 with terpenoid volatile compounds. Front. Plant Sci. 8, 1930.

407 Lemmich, E., Jensen, B., Rasmussen, U., 1984. (8R,14S)-8-angeloyloxythapsan-14-ol, a sesquiterpene with a
408 novel carbon skeleton, from *Thapsia villosa*. Phytochemistry 23, 809–811.

409 Lemmich, E., Wagner Smitt, U., Sandholm Jensen, J., Brøgger Christensen, S., 1991. Guaiane esters from
410 *Thapsia villosa*. Phytochemistry 30, 2987–2990.

411 Liu, Q., Majdi, M., Cankar, K., Goedbloed, M., Charnikhova, T., Verstappen, F.W.A., de Vos, R.C.H.,
412 Beekwilder, J., van der Krol, S., Bouwmeester, H.J., 2011. Reconstitution of the costunolide biosynthetic
413 pathway in yeast and *Nicotiana benthamiana*. PLoS One 6, e23255.

414 Luo, D., Callari, R., Hamberger, B., Wubshet, S.G., Nielsen, M.T., Andersen-Ranberg, J., Hallstrom, B.M.,
415 Cozzi, F., Heider, H., Lindberg Moller, B., Staerk, D., Hamberger, B., 2016. Oxidation and cyclization
416 of casbene in the biosynthesis of *Euphorbia* factors from mature seeds of *Euphorbia lathyris* L. Proc.
417 Natl. Acad. Sci. U. S. A. 113, E5082-9.

418 Máday, E., Szöke, É., Muskáth, Z., Lemberkovics, É., 1999. A study of the production of essential oils in
419 chamomile hairy root cultures. Eur. J. Drug Metab. Pharmacokinet. 24, 303–308.

420 Nelson, D., 2009. The Cytochrome P450 Homepage. Hum. Genomics 4, 59–65.

421 Nour-Eldin, H.H., Hansen, B.G., Nørholm, M.H.H., Jensen, J.K., Halkier, B.A., 2006. Advancing uracil-
422 excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 34,
423 e122.

424 Peyret, H., Lomonossoff, G.P., 2013. The pEAQ vector series: the easy and quick way to produce recombinant
425 proteins in plants. Plant Mol. Biol. 83, 51–58.

426 Pickel, B., Drew, D.P., Manczak, T., Weitzel, C., Simonsen, H.T., Ro, D., 2012. Identification and
427 characterization of a kunzeaol synthase from *Thapsia garganica*: Implications for the biosynthesis of the

pharmaceutical thapsigargin. Biochem. J. 448, 261–271.

Raharivelomanana, P., Bianchini, J.-P., Cambon, A., Azzaro, M., Faure, R., 1995. Two-dimensional NMR of sesquiterpenes. 8—complete assignment of 1H and 13C NMR spectra of seven sequiterpene alcohols from *Neocallitropsis pancheri*. Magn. Reson. Chem. 33, 233–235.

Simonsen, H.T., Weitzel, C., Christensen, S.B., 2013. Guaianolide sesquiterpenoids - Their pharmacology and biosynthesis, in: Ramawat, K.G., Merillon, J.M. (Eds.), Handbook of Natural Products - Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes, Handbook of Natural Products. Springer-Verlag, Berlin, Germany, pp. 3069–3098.

Thastrup, O., Cullen, P.J., Drobak, B.K., Hanley, M.R., Dawson, A.P., 1990. Thapsigargin, a tumor promoter, discharges intracellular Ca2+ stores by specific inhibition of the endoplasmic reticulum Ca2(+)-ATPase. Proc. Natl. Acad. Sci. U. S. A. 87, 2466–2470.

Triana, J., López, M., Pérez, F.J., González-Platas, J., Quintana, J., Estévez, F., León, F., Bermejo, J., 2005. Sesquiterpenoids from *Pulicaria canariensis* and their cytotoxic activities. J. Nat. Prod. 68, 523–531.

van Herpen, T.W., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J., Beekwilder, J., 2010. *Nicotiana benthamiana* as a production platform for artemisinin precursors. PLoS One 5, e14222.

Wang, G.P., Yu, X.D., Fan, J., Wang, C.S., Xia, L.Q., 2015. Expressing an (E)-beta-farnesene synthase in the chloroplast of tobacco affects the preference of green peach aphid and its parasitoid. J. Integr. Plant Biol.

Weitzel, C., Rønsted, N., Spalik, K., Simonsen, H.T., 2014. Resurrecting deadly carrots: Towards a revision of *Thapsia* (Apiaceae) based on phylogenetic analysis of nrITS sequences and chemical profiles. Bot. J. Linn. Soc. 174.

Yahyaa, M., Tholl, D., Cormier, G., Jensen, R., Simon, P.W., Ibdah, M., 2015. Identification and characterization of terpene synthases potentially involved in the formation of volatile terpenes in carrot (*Daucus carota* L.) roots. J. Agric. Food Chem. 63, 4870–4878.