Biosynthesis of the phenolic monoterpenes, thymol and carvacrol, by terpene synthases and cytochrome P450s in oregano and thyme

Dissertation

Zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von Diplom-Biologe

Christoph Crocoll

geboren am 11. Februar 1977 in Kassel

Gutachter:

1. Prof. Dr. Jonathan Gershenzon, Max-Planck-Institut für chemische Ökologie, Jena
2. Prof. Dr. Christian Hertweck, Hans-Knöll-Institut, Jena
3. Prof. Dr. Harro Bouwmeester, Wageningen University, Wageningen

Tag der öffentlichen Verteidigung: 11.02.2011

Biosynthesis of the phenolic monoterpenes, thymol and carvacrol, by terpene synthases and cytochrome P450s in oregano and thyme

Christoph Crocoll - Max-Planck-Institut für chemische Ökologie - 2010

Contents

1 General introduction

Plants are sophisticated light-driven "green" factories able to synthesize an immense number of bio-active natural products (Jensen and Møller, 2010). These natural products are also referred to as secondary products or secondary metabolites since they are not directly essential for the basic processes of growth and development (Theis and Lerdau, 2003). The investigation of plant natural products has a long history that started about 200 years ago with the isolation of morphine by Friedrich Wilhelm Sertürmer. Since then the number of described secondary metabolites has risen to over 200,000 (Hartmann, 2007). They can be divided into two major classes, the first class formed by nitrogen-containing substances, such as alkaloids, amines, cyanogenic glycosides, non-protein amino acids and glucosinolates, and the second class consisting of nitrogen-free substances which are represented by polyketides, polyacetylenes, saponins, phenolics and terpenes. Many of the secondary metabolites were found to serve plants as defenses against herbivores, pathogens and abiotic stresses (Huang et al., 2010).

In human society, plants play an irreplaceable role as food sources, not only for their nutritional value but also as spices and herbs which help preserve food or improve its taste. The plant compounds that add flavor to our food are mainly secondary products, such as capsaicin, an alkaloid, which is responsible for the hot taste of chili; or thymol, a terpene, which is one of the main flavoring components in herbs like oregano (*Origanum* sp.) or thyme (*Thymus* sp.).

Oregano and thyme belong to the Lamiaceae plant family which harbors many other aromatic plants of great scientific and economic interest such as rosemary, sage, mint, and marjoram. The aroma associated with these plants arises from the essential oil found in peltate glandular trichomes on the aerial parts of the plant. These glandular trichomes consist of highly specialized secretory cells in which the components of the essential oil are synthesized and subsequently accumulate in a subcuticular storage cavity (Gershenzon et al., 1989; Turner et al., 1999). The composition of the essential oils of oregano, thyme and marjoram is dominated by mono- and sesquiterpenes (Skoula and Harborne, 2002; Stahl-Biskup, 2002).

These substances are responsible for the aroma and flavor of these herbs, and the extracted essential oils are used for the manufacturing of perfumes and cosmetics as well as for medicinal and pharmaceutical purposes as antimicrobial or antiseptic agents (Kintzios, 2002; Stahl-Biskup, 2002). Mono- and sesquiterpenes are also thought to help defend the plant against herbivores and pathogens (Gershenzon and Dudareva, 2007).

Two monoterpenes of the Lamiaceae that have attracted much attention, thymol and carvacrol, are found in thyme and oregano but not in marjoram. These phenolic monoterpenes are especially known for their antiherbivore, antimicrobial, pharmaceutical and antioxidant

activities (Isman, 2000; Hummelbrunner and Isman, 2001; Ultee et al., 2002; Sedy and Koschier, 2003; Braga et al., 2008). They are even used to treat bee hives against the varroa mite without harming the bees (Floris et al., 2004). According to a prediction by (Poulose and Croteau, 1978a), the pathway for thymol formation proceeds from *γ*-terpinene via the aromatic compound, *p*-cymene, as an intermediate. However, despite extensive efforts to breed oregano or thyme varieties with a larger proportion of thymol and carvacrol for pharmaceutical use and the interest in these terpenes as plant defenses, no genes or enzymes responsible for thymol or carvacrol formation from *γ*-terpinene or *p*-cymene have been described.

Questions addressed in this thesis

As stated above plant secondary compounds are of great importance not only for plant defense but also for pharmaceutical and medicinal purposes. Understanding the mechanisms underlying the formation and regulation of plant secondary compounds is essential to further investigate their roles in plant defense and develop new strategies to make these compounds more readily available for pharmaceutical and nutritional usage. In this thesis the biosynthesis of the two highly bioactive compounds, thymol and carvacrol, is studied on different levels. Chapter I covers the molecular and biochemical mechanisms of mono- and sesquiterpene biosynthesis in different oregano cultivars emphasizing the role of terpene synthases in producing the basic terpene carbon skeletons. In Chapter II, the next biosynthetic step in the formation of thymol and carvacrol is investigated in the closely related plants oregano, thyme and marjoram. Here, the focus lies on the oxidation of the basic terpene skeleton by cytochrome P450s. Five novel P450 genes from oregano and thyme chemotypes are described, and the biochemical characterization of three of them by heterologous expression in a yeast system is presented. In Chapter III some of the limitations of the yeast expression system are circumvented by over-expression in the model plant *Arabidopsis thaliana*. As a result, a new pathway for thymol and carvacrol formation is proposed that is different from the prediction in the literature.

Terpenes

Terpenes (also known as terpenoids or isoprenoids) form the largest group of natural products with more than 30,000 different structures (Buckingham, 1998) spread over the widest assortment of structural types with hundreds of different monoterpene, sesquiterpene, diterpene and triterpene carbon skeletons (Degenhardt et al., 2010). The majority of terpenes have been isolated from plants where their enormous structural variability leads to a great functional diversity. Terpenes play important roles in almost all basic plant processes, including growth, development, reproduction and defense (Gershenzon, 1999). For example, phytol, the side chain of the photosynthetic pigment chlorophyll, is the most abundant plant terpenoid (Davis and

Croteau, 2000). Still, comparatively little is known about the actual role of most terpenes in nature despite this immens number and the importance of natural products in medicine, agriculture and industry (Gershenzon and Dudareva, 2007).

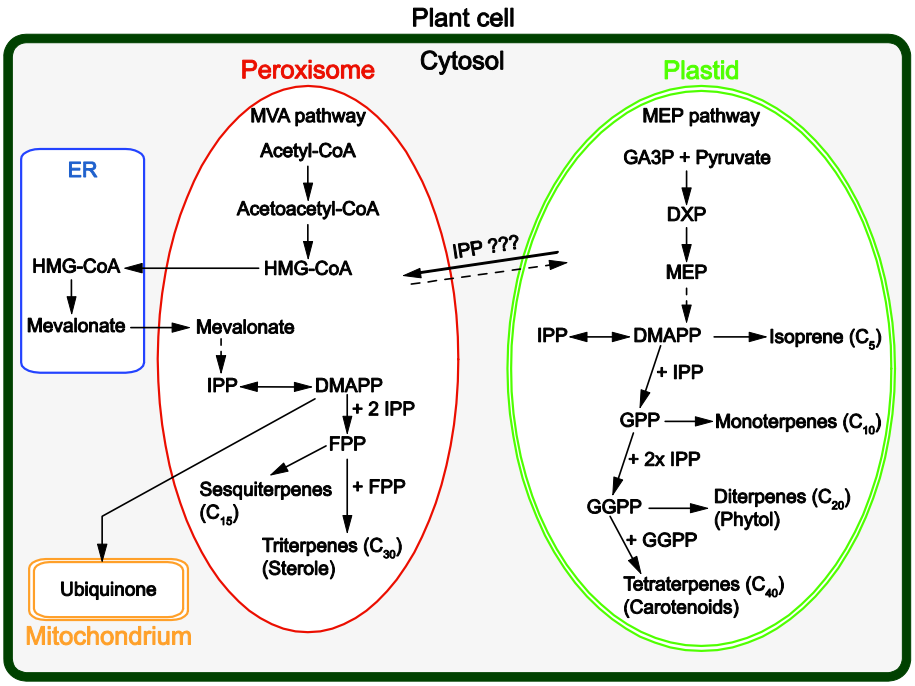


Figure 1 Compartmentation of plant terpene biosynthesis. The Mevalonic acid (MVA) pathway is located in the cytosol in peroxisomes and in the endoplasmatic reticulum (ER). It starts with three units of Acetyl-CoA and the final product farnesyl pyrophosphate (FPP) is the precursor molecule for all sesquiterpenes. The Methyl-erythritol-phosphate (MEP) pathway is located in the plastids and the initial substrates are glyceraldehyde-3-phosphate (GA3P) and pyruvate. Geranyl diphosphate (GPP) is the precursor for all monoterpenes and geranyl geranyldiphsophate (GGPP) the precursor for diterpenes. Carotenoids are derived from two units of GGPP. DMAPP (dimethylallyl diphosphate) is the backbone to which different numbers of the isomer IPP (isopentenyl diphosphate) are added to form GPP, FPP or GGPP. Ubiquinone is formed in mitochondria. An exchange of IPP between different compartments is still under investigation. (Redrawn after (Sapir-Mir et al., 2008; Sallaud et al., 2009)).

Much more is known about the biosynthesis and localization of terpenes within the plant cell. Terpenes are formed from the universal five-carbon building blocks, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are both synthesized by the plastidic methylerythritol pathway and the cytosolic mevalonate pathway (Gershenzon, 1999; Sapir-Mir et al., 2008; Sallaud et al., 2009) (Fig. 1). DMAPP and IPP are fused by prenyltransferases to form geranyl diphosphate (GPP, C_{10}), the usual precursor of the monoterpenes, and DMAPP and two units of IPP are fused to form farnesyl diphosphate (FPP, C_{15}) the precursor of most sesquiterpenes. Next, the linear carbon skeletons of GPP and FPP are converted to the basic terpene skeletons by terpene synthases, a widespread class of enzymes

responsible for the huge structural diversity of mono- and sesquiterpenes since these enzymes often form multiple products (Tholl, 2006; Degenhardt et al., 2010).

Terpene biosynthesis in Lamiaceae

Plant terpene biosynthesis is often restricted to special morphological structures like idioblasts (e.g. oil cells in *Laurus* sp.) or ducts (e.g. resin ducts in *Pinus* sp.) or trichomes (e.g. glandular trichomes in Lamiaceae and Asteraceae) that can store these lipophilic compounds in high concentrations (Fahn, 1988).

Studies on Lamiaceae species such as mint (*Mentha* sp.) and sweet basil (*Ocimum basilicum*) have provided many insights into essential oil biosynthesis. In both plants, the essential oil is produced in glandular trichomes situated on the aerial parts of the plants. These glandular trichomes consist of a cluster of secretory cells covered by a subcuticular storage cavity where the essential oil accumulates (Gershenzon et al., 1989; Turner et al., 1999) (Figure 2).

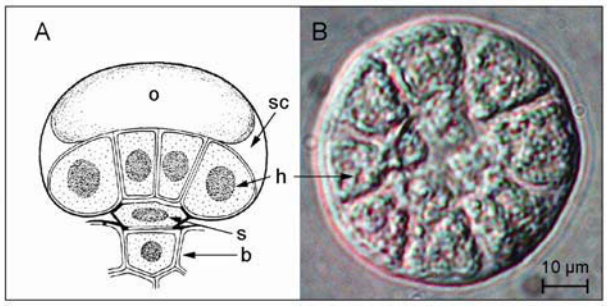


Figure 2 Peltate glandular trichomes of Lamiaceae. (**A**) Schematic drawing of a peltate glandular trichome with a basal cell (b) and the head cells (h) based on one stalk cell (s). Secreted terpenes are stored in the subcuticular cavity (sc) as an oil drop (o). (Modified after Fahn, 1988). (**B**) Micrograph of the isolated head cells of a glandular trichome from oregano. Eight larger cells are arranged around four smaller inner cells.

Terpene synthase genes have been isolated and characterized from several Lamiaceae species, including the genus *Mentha* where, among others, a limonene synthase, an (*E*)-β-farnesene synthase, and a *cis*-muuroladiene synthase have been identified (Colby et al., 1993; Haudenschild et al., 2000). From other Lamiaceae, terpene synthase genes are known from *Salvia officinalis* (Wise et al., 1998), *S. pomifera* and *S. fruticosa* (Kampranis et al., 2007), *Rosmarinus officinalis* (Tan, 2007), *Lavandula angustifolia* (Landmann et al., 2007) and *Ocimum basilicum* (Iijima et al., 2004a; Iijima et al., 2004b).

Terpene synthases

Considering the entire protein, plant monoterpene synthase with about 550-650 amino acids are longer than sesquiterpene synthases (550–580 amino acids) due to an N-terminal signal peptide that targets the initial translation product towards the plastids (Turner et al., 1999; Degenhardt et al., 2010). Terpene synthases share several highly conserved motifs in their amino acid sequence, such as the RR motif and the DDxxD motif. The RR motif is located in the N-terminus and appears to be essential during the isomerization of geranyl pyrophosphate to linalyl pyrophosphate, a common mechanism of cyclic monoterpene formation (Williams *et al.*, 1998) (Fig. 3). The aspartate-rich DDxxD motif has an important function in binding the Mn^{2+}- or Mg^{2+}-pyrophosphate complex of the diphosphate substrates GPP or FPP (Cane *et al.*, 1996, Ashby and Edwards, 1990).

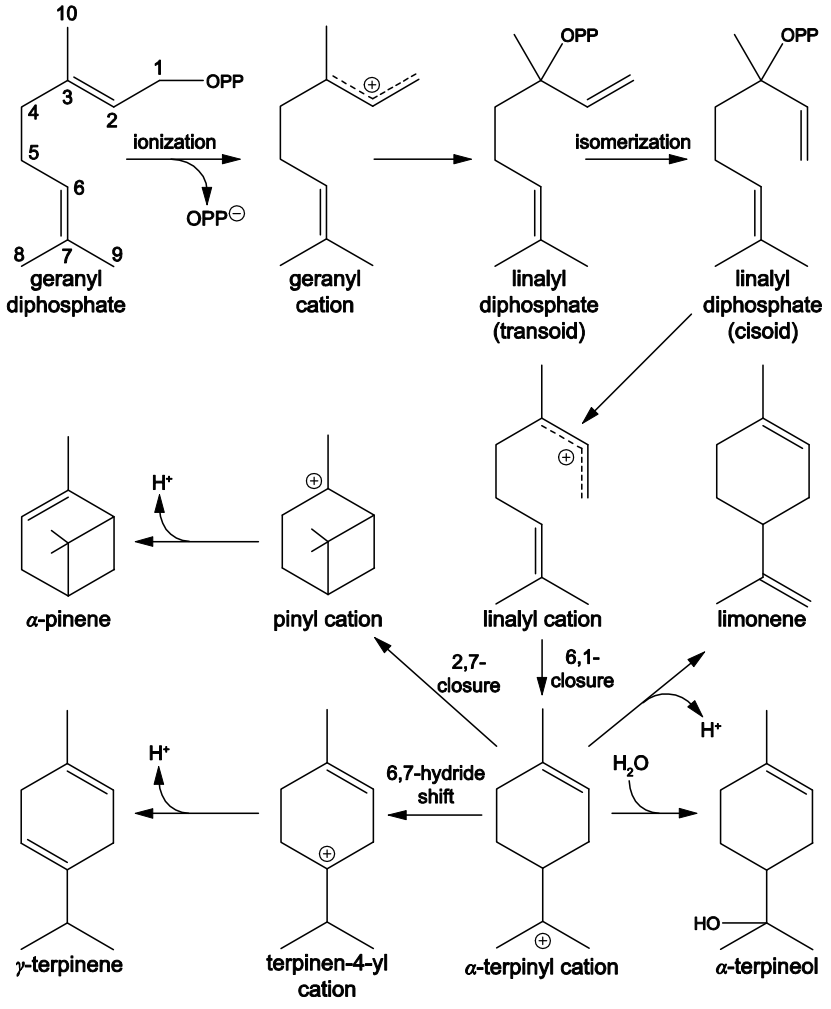


Figure 3 Reaction mechanism of cyclic monoterpenes formation from geranyl diphosphate (GPP). The reaction starts with the ionization of GPP and the resulting geranyl cation is isomerized to a linalyl intermediate capable of cyclizations. The initial cyclic *α*-terpinyl cation is the central intermediate and can be subject to further cyclizations, hydride shifts and rearrangements before the reaction is terminated by deprotonation or water capture. The numbering of carbon atoms in intermediates and products corresponds to that of GPP.

Recently, 213 plant mono- and sesquiterpene synthases and the reaction mechanisms leading to the structural diversity have been reviewed in detail (Degenhardt et al., 2010). A short overview of the initial steps in monoterpene formation from the precursor GPP is shown in Figure 3. Ionization of GPP results in a geranyl cation which is most likely the precursor for acyclic monoterpenes such as geraniol, linalool, myrcene and *(E)*-*β*-ocimene. Isomerization of the initial geranyl cation leads to a linalyl cation which is involved in the formation of cyclic monoterpenes via the *α*-terpinyl cation. The *α*-terpinyl cation is the central precursor for all cyclic monoterpenes such as limonene or *γ*-terpinene. Oxidation, reduction, rearrangement (via hydride shifts and additional ring closures), conjugation, double bond isomerization and hydration result in a high diversity of monoterpene structures. The initial terpene synthase products (olefins and monoalcohols) are often further oxidized or conjugated. One prominent class of enzymes, the cytochrome P450s, plays an important role in these downstream modifications.

Cytochrome P450s – a widespread enzyme family

Cytochrome P450s monooxygenases (P450s) form an ubiquitous class of enzymes known from all kinds of organisms including animals, bacteria, fungi, archaea, protists and even viruses (http://drnelson.uthsc.edu/P450.statistics.Aug2009.pdf). The total P450 count reached over 11,000 in 2009 which is 39 % (or > 3000) more than in February 2008 (http://drnelson.uthsc.edu/p450stats.Feb2008.htm). The largest group of P450s (4,266 enzymes) is found in plants which is one reason for the immense amount of structurally diverse natural products in this kingdom.

Table 1 Total numbers of known cytochrome P450 monooxygenases from different organisms as of August 2009 (http://drnelson.uthsc.edu/P450.statistics.Aug2009.pdf). CYP = cytochrome P450

Organism	Total number	CYP families	CYP subfamilies
Plants	4,266	126	464
Animals (w/o insects)	1,607	69	169
Insects	1,675	59	338
Fungi	2,570	459	1,011
Protists	247	62	119
Bacteria	905	196	409
Archaea	22	12	14
Viruses	2	2	2
Total	11,294	977	2519

Cytochrome P450 monooxygenases are heme-dependent mixed function oxidases that utilize NADPH and / or NADH to reductively cleave atmospheric dioxygen producing a functionalized organic product and a molecule of water (Schuler, 1996). All plant cytochrome P450s described to date are bound to membranes of the endoplasmic reticulum through a short hydrophobic segment of their N-terminus (Williams et al., 2000; Werck-Reichhart et al., 2002). P450s need to be coupled to an electron donating protein, a cytochrome P450 reductase or a cytochrome b5, which is also anchored to the endoplasmic reticulum by its N- or C-terminus (Schuler, 1996; Werck-Reichhart et al., 2002).

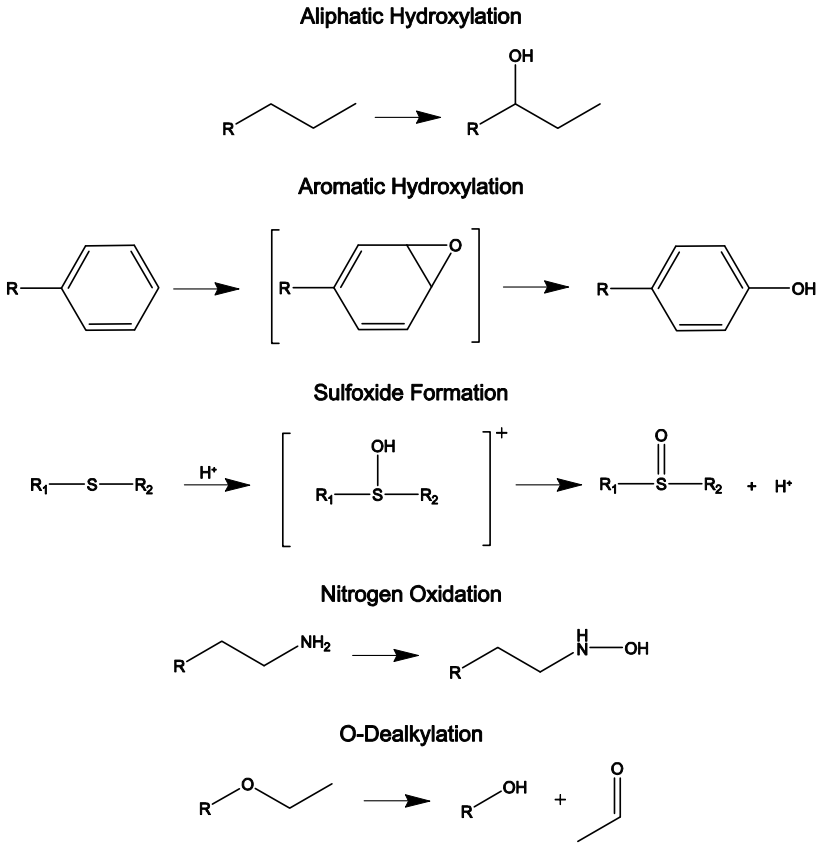


Figure 4: Examples of reactions catalyzed by different cytochrome P450 enzymes.

Although the sequence identity between distantly related P450s can be as low as that between two random sets of 500 amino acids (a common sequence length of P450s), a few absolutely conserved sequence motifs are known, the WxxxR motif, the GxE/DTT/S motif, the ExLR motif, the PxxFxPE/DRF motif and the PFxxGxRxCxG/A motif. The latter, most conserved P450 motif (PFxxGxRxCxG/A) represents the heme binding loop and is often considered as a 'signature' sequence for P450 proteins (Feyereisen, 2005). This motif is responsible for the characteristic 450 nm absorption of the Fe^{II}-CO complex of cytochrome P450 (Mansuy and Renaud, 1995). As shown by crystal structure analysis of mostly bacterial P450s, the three-dimensional structure reveals quite high conservation which would not be expected from the

high overall sequence diversity (Poulos et al., 1985; Feyereisen, 2005). Cytochrome P450s catalyze a large variety of reactions like aliphatic and aromatic hydroxylations, N-, O-, and S-dealkylation, oxidative deamination, sulfoxide formation or N-oxidation. An overview of common reactions is shown in Figure 4.

The nomenclature for P450s comprises the CYP prefix, followed by an Arabic numeral which designates the family (all members nominally > 40 % identical), a capital letter designates the subfamily (all members nominally > 55 % identical) and an Arabic numeral designates the individual gene (all numbers and letters in italics) or protein (no italics) (Nebert et al., 1991, 1991; Nelson et al., 1993; Nelson et al., 1996). This would mean that the mint limonene-3-hydroxylase CYP71D13 is P450 number 13 of the subfamily D of P450 family 71. The family names also indicate the organism the P450 is from; the plant P450s are found in families 51, 71-99 and 701-804 (http://drnelson.uthsc.edu/CytochromeP450.html).

The catalytic cycle of cytochrome P450s

Figure 5 shows the catalytic cycle of cytochrome P450s which has been mostly validated by experimental evidence (Groves, 2005; Makris et al., 2005; Ortiz de Montellano, 2010). The resting enzyme is in the ferric state and has a thiolate proximal ligand while the distal ligand is usually a water molecule (**A**). Upon substrate (RH) binding water is displaced if present. This causes a shift in the redox potential of the heme iron atom which enables electron transfer from the redox partner, NAD(P)H (**B**). The higher redox potential can cause a reduction of ferric (Fe^{III}) into ferrous (Fe^{II}) (**C**) which is followed by binding of molecular dioxygen to result in the instable ferrous dioxy complex (**D**). An additional reduction generates the ferric peroxy anion which is protonated to yield the ferric hydroperoxo complex (**E**). This second electron transfer can be the rate-limiting step of the catalytic cycle. A ferryl intermediate (**F**) is formed from the unstable ferric hydroperoxo intermediate by protonation and heterolytic cleavage of the O-O bond, also known as *compound I*. Different formulations are possible for this state (Ortiz de Montellano, 2010). Finally, the ferryl intermediate reacts with the substrate to produce the hydroxylated metabolite (**G**). After release of the product and re-equilibration with water the resting ferric state is reached again (**A**).

The exact mechanism of how the oxygen is transferred to the substrate is still under investtigation. Most likely, a radical rebound mechanism applies for hydrocarbon hydroxylations. The cleavage of the C-H bond results in the formation of a radical in the substrate. In a second step, the Fe^{IV+}-OH intermediate and the substrate are connected via the oxygen as shown in Figure 5 (**G**) and finally the hydroxylated product dissociates (Meunier et al., 2004; Shaik et al., 2005). The radical rebound mechanism is largely supported by studies involving the use of

hydrocarbons as radical clocks (He and Ortiz de Montellano, 2004; Groves, 2005; Jiang and Ortiz de Montellano, 2009; Ortiz de Montellano, 2010).

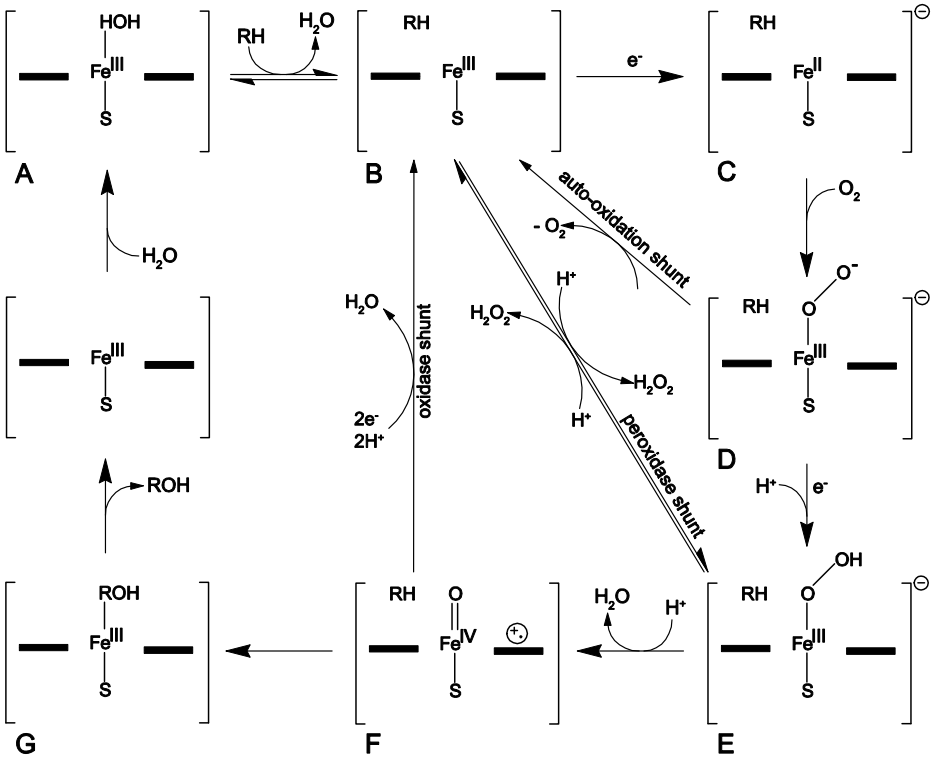


Figure 5: Summary of the cytochrome P450 catalytic cycle. The heme group is represented by two solid bars with the iron (Fe) between them. The S represents the thiolate of the protein. RH is the hydrocarbon substrate and ROH is the hydroxylated product. The **+.** over one heme bar in (**F**) indicates the position of the radical cation on the porphyrin ring. The different steps of the catalytic cycle are described in detail within the previous paragraph. (Redrawn after (Karuzina and Archakov, 1994; Ortiz de Montellano, 2010)).

In addition to the main pathway three shunt pathways are known which result in the dissociation of oxygen without hydroxylation of the substrate (Karuzina and Archakov, 1994). The *autoxidation shunt* leads to an elimination of oxygen from the instable ferrous dioxy complex (**D**). The *peroxide shunt* describes the elimination of hydrogen peroxide from the ferric hydroperoxo complex (**E**). Via the *oxidase shunt* the ferryl intermediate or *compound I* (**F**) is transformed back into intermediate (**B**) by reduction of oxygen into water which consumes two equivalents of NAD(P)H. All three shunt pathways result in an unwanted uncoupling of co-factor comsumption without hydroxylation of the substrate.

As shown in the catalytic cycle, electron transfer to the heme is required for oxygen activation. Different types of electron transfer chains are known for cytochrome P450s. In plants and other eukaryotes, class II P450s bound to the endoplasmic reticulum are the most common (Hannemann et al., 2007). Adjacent to the P450 is another membrane-bound protein, a cytochrome P450 oxidoreductase (CPR) containing the prosthetic groups FAD and FMN, which

are both required to transfer redox equivalents from NADPH to the P450. The estimated ratio between CPR and P450 in the microsomal membrane is 1:15 which requires some control to ensure efficient electron transfer from CPR to P450s (Shephard et al., 1983). Recently 54 CPRs from vascular plants have been reviewed in detail (Jensen and Møller, 2010). In different plant species, the number of CPR homologs was found to vary between one and three. In general, plant P450s can be catalytically supported by CPRs from other plant species which is reflected by the high conservation of amino acid residues involved in CPR-P450 interactions (Bak et al., 2000; Tattersall et al., 2001). Nevertheless, the efficiency of the electron transfer may still depend on the CPR homolog present, and thus different CPRs may differentially influence cytochrome P450 performance (Hasemann et al., 1995; Jensen and Møller, 2010).

Secondary modifications of terpenes by cytochrome P450s

In general, cytochrome P450 monooxygenases play important roles in plant survival both by detoxifying xenobiotics and synthesizing defense compounds with high bioactivity against enemies (Schuler, 1996; Ohkawa et al., 1999; Werck-Reichhart et al., 2002). Compounds like the cyanogenic glucosides in *Sorghum bicolor* (Halkier and Møller, 1991; Halkier et al., 1995; Bak et al., 1998) or glucosinolates in *Arabidopsis* (Du et al., 1995; Wittstock and Halkier, 2000; Chen et al., 2003) are produced in pathways containing multiple P450s. The impact of cytochrome P450s in plants becomes even more obvious when realizing the large number of P450s that have been described. In the model plant *Arabidopsis thaliana*, 246 putative P450 genes and 26 pseudogenes were annotated, which represent approximately 1 % of its gene complement (Paquette et al., 2000; Werck-Reichhart et al., 2002; Schuler and Werck-Reichhart, 2003; Nelson et al., 2004) and in rice the number of 457 annotated P450s is even more impressive (Schuler and Werck-Reichhart, 2003).

Biosynthetic P450s catalyze steps in pathways leading to a range of plant compounds like lignin intermediates, sterols, furanocoumarins and terpenes (Bolwell et al., 1994; Schuler, 1996).

Within the terpenes, mono-, sesqui-, di-, and triterpenes are all substrates for cytochrome P450s (Bolwell et al., 1994) and these enzymes play therefore an important role in generating some of the enormous structural diversity of plant terpenoid secondary metabolites (Ro et al., 2005). In loblolly pine diterpene resin acid biosynthesis, multisubstrate, multifunctional P450s catalyze an array of consecutive oxidation steps with several different alcohol and aldehyde intermediates (Ro et al., 2005). P450s also catalyze the hydroxylation of sesquiterpenes in chicory (*Cichorium intybus* L.) (de Kraker et al., 2003), and P450-mediated modifications of monoterpenes are known to occur for acyclic monoterpenes, such as geraniol-10-hydroxylation in *Catharanthus roseus* (Meijer et al., 1993) or the bicyclic monoterpenes, sabinene and pinene (Karp et al., 1987).

The largest diversity of monoterpenes hydroxylations by cytochrome P450s is reported for limonene. Several limonene-using P450s have been identified from mint, caraway and *Perilla* (Karp et al., 1990; Bouwmeester et al., 1998; Mau et al., 2010). In *Mentha*, the biosynthesis of menthol and carvone involves hydroxylation steps catalyzed by cytochrome P450 monooxygenases (Colby et al., 1993; Lupien et al., 1999; Haudenschild et al., 2000). Hydroxylation reactions catalyzed by cytochrome P450s are often highly regiospecific (Schuler, 1996) which is important since the position of hydroxylation can determine the downstream fate of the products in biosynthetic pathways. In mint, the oxygenation pattern of monoterpenes is determined by the regiospecificity of P450 mediated (-)-*S*-limonene-hydroxylation either at carbon C3 or C6 (Lupien et al., 1995). The hydroxylation at carbon C6 results in the formation of carveol which leads to the accumulation of carvone in spearmint, whereas hydroxylation at carbon C3 results in the formation of isopiperitenol which is subsequently transformed into the commercially important menthol in peppermint (Karp et al., 1990; Lupien et al., 1995; Croteau et al., 2005). The structural similarity of these highly regiospecific limonene hydroxylases to those enzymes forming thymol and carvacrol is discussed in Chapters II and III.

2 Chapter I

Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis

2.1 Abstract

The aroma, flavor and pharmaceutical value of cultivated oregano (*Origanum vulgare* L.) is a consequence of its essential oil which consists mostly of monoterpenes and sesquiterpenes. To investigate the biosynthetic pathway to oregano terpenes and its regulation, we identified and characterized seven terpene synthases, key enzymes of terpene biosynthesis, from two cultivars of *O. vulgare*. Heterologous expression of these enzymes showed that each forms multiple mono- or sesquiterpene products and together they are responsible for the direct production of almost all terpenes found in *O. vulgare* essential oil. The correlation of essential oil composition with relative and absolute terpene synthase transcript concentrations in different lines of *O. vulgare* demonstrated that monoterpene synthase activity is predominantly regulated on the level of transcription and that the phenolic monoterpene alcohol thymol is derived from γ-terpinene, a product of a single monoterpene synthase. The combination of heterologously-expressed terpene synthases for *in vitro* assays resulted in blends of mono- and sesquiterpene products that strongly resemble those found *in vivo*, indicating that terpene synthase expression levels directly control the composition of the essential oil. These results provide the possibility / tools to develop strategies for metabolic engineering and direct breeding of *O. vulgare* cultivars with higher quantity of essential oil and an improved composition.

2.2 Introduction

Throughout the plant kingdom, many species produce aromatic essential oils that contain mixtures of terpene or phenylpropanoid metabolites. The well-known culinary herb, oregano, *Origanum vulgare* L., contains an essential oil whose composition is dominated by monoterpenes and sesquiterpenes. These substances are responsible for the aroma and flavor of oregano, its pharmaceutical uses as an antimicrobial and antiseptic agent, and its anti-oxidant activity, often touted as a health benefit (Kintzios, 2002). They are also thought to help defend the plant against herbivores and pathogens (Gershenzon and Dudareva, 2007). Two components of the essential oil, the phenolic monoterpenes, thymol and carvacrol, are especially known for their antiherbivore, antimicrobial, pharmaceutical and antioxidant activities (Isman, 2000; Hummelbrunner and Isman, 2001; Sedy and Koschier, 2003; Braga et al., 2008). However, the composition and quantity of essential oil varies strongly between populations and accessions of *O. vulgare* (Kintzios, 2002) and within the genus *Origanum* (Kokkini and Vokou, 1989; Vokou et al., 1993). Despite extensive efforts to breed *O. vulgare* varieties with a larger proportion of thymol and carvacrol for pharmaceutical use and the interest in terpenes as plant defenses, little is known about the molecular factors determining the biosynthesis and composition of the essential oil.

Studies on other Lamiaceae species such as mint (*Mentha* sp.) and sweet basil (*Ocimum basilicum*) have provided many insights into essential oil biosynthesis. In both plants, the essential oil is produced in glandular trichomes situated on the aerial parts of the plants. The glandular trichomes consist of a cluster of secretory cells covered by a subcuticular storage cavity where the essential oil accumulates (Gershenzon et al., 1989; Turner et al., 1999). Within the secretory cells, the five-carbon building blocks for terpenes, isopentenyl diphosphate and its isomer dimethylallyl diphosphate, are synthesized by the plastidic methylerythritol pathway and the cytosolic mevalonate pathway (Gershenzon, 1999). Two of these five-carbon units are fused to form geranyl diphosphate (GPP), the usual precursor of the monoterpenes, and three of these units are fused to form farnesyl diphosphate (FPP) the precursor of most sesquiterpenes. Next, the linear carbon skeletons of GPP and FPP are converted to the basic terpene skeletons by terpene synthases, a widespread class of enzymes responsible for the huge structural diversity of monoterpenes and sesquiterpenes (Tholl, 2006). Terpene synthase genes have been isolated and characterized from several Lamiaceae species, including the genus *Mentha* where a limonene synthase, an (*E*)-*β*-farnesene synthase, and a *cis*-muuroladiene synthase have been identified (Colby et al., 1993; Haudenschild et al., 2000). From other Lamiaceae, terpene synthase genes are known from *Salvia officinalis* (Wise et al., 1998), *S. pomifera* and *S. fruticosa* (Kampranis et al., 2007), *Rosmarinus officinalis* (Tan, 2007) and *Ocimum basilicum* (Iijima et al., 2004b). The initial terpene synthase products (olefins and monoalcohols) are often further oxidized or

conjugated. In *Mentha*, the biosynthesis of menthol and carvone involves hydroxylation steps catalyzed by cytochrome P450 monooxygenases (Colby et al., 1993; Lupien et al., 1999; Haudenschild et al., 2000).

These previous results have many implications for monoterpene biosynthesis in oregano. For example, the phenolic monoterpene thymol, a major antioxidant and antimicrobial component, is likely to be derived from GPP by a terpene synthase followed by oxidative steps. In *Thymus vulgaris*, thymol formation was suggested to proceed via the olefins *γ*-terpinene and *p*-cymene (Poulose and Croteau, 1978a). A second group of monoterpenes common to *O. vulgare* essential oil is characterized by the bicyclic monoterpene alcohols, *cis*- and *trans*-sabinene hydrate which impart the typical marjoram flavor (found in *Origanum majorana*) (Kintzios, 2002). These are likely to be direct products of terpene synthases based on previous isolation of sabinene hydrate synthases (Hallahan and Croteau, 1988, 1989). In most *Origanum* species, the monoterpenes are accompanied by varying amounts of sesquiterpenes (Skoula et al., 1999). These are also likely to be direct terpene synthase products based on accumulated knowledge of sesquiterpene synthases (Tholl, 2006).

In order to learn more about the pathway and molecular mechanisms regulating terpene production in *O. vulgare* L., we isolated and characterized seven terpene synthases responsible for the production of the major essential oil components. Correlations of terpene synthase transcript levels from qRT-PCR and essential oil composition in clonal *O. vulgare* lines were then employed to examine the role of these enzymes in the biosynthesis of various terpenes and its control.

2.3 Materials and Methods

Plant Material

Oregano (*Origanum vulgare* L.) plants were propagated from stem cuttings and grown in the greenhouse with additional illumination by sodium lamps (16 h day at 20-22 °C and 8 h night at 18-20 °C), and 30-55 % humidity. Plants were potted in commercial soil (Tonsubstrat, Klasmann, Geeste/Groß-Hesepe, Germany) and watered every 2 to 3 days with tap water. Two cultivars of *Origanum vulgare* L. (cv. 'd06-01' and cv. 'f02-04') were selected from the collection of *Origanum* species of the botanical garden of Università di Bari, Italy, which were chosen for the presence of both sabinyl- and *p*-cymyl compounds in the essential oil. An additional criterion was a low density of hairy trichomes which facilitates the extraction of glandular trichomes. Herbarium specimens of the two genotypes are kept at the Herbarium of the Institute of Applied Botany, University of Veterinary Medicine, Vienna, Austria. For the cultivar d06-01, three clonal lines were chosen (designated d2, d5 and d8) and from the cultivar f02-04 four lines were chosen (designated f2 through f5). Additional lines were derived by selfing f02-04 plants (seven lines chosen designated ff1, ff2, ff4, ff5, ff6, ff7 and ff8) and crossing line d06-01 with f02-04 (four lines chosen designated df5 through df8).

Terpene extraction from leaves

For terpene extractions correlated to the RNA-hybridization analysis, leaf material of the *O. vulgare* lines d2, d5, d8, f2-f5, df5-df8, ff2, ff4, ff6, and ff7 was harvested in June 2006. Young expanding leaves from five plants of each line were pooled and frozen in liquid N_2 immediately after harvest and ground to a fine powder with mortar and pestle. The powder (50-100 mg) was soaked in 1 ml ethyl acetate:pentane (2:1) containing an internal standard (menthol, 50 ng/µl) for 24 h at room temperature with constant rotation. The solution was cleared with activated charcoal for 5 min and dried over a column of 500 mg water-free Na_2SO_4. All extractions were performed in triplicate.

For terpene extractions correlated to the qRT-PCR analysis, leaf material of the lines d2, d5, f4, f5, df6, df8, ff1, ff2, ff4, ff5, ff6, ff7, ff8 was harvested in November 2009. Young expanding leaves from three plants of each line were individually harvested and extracted as described above.

Determination of glandular trichome abundance

The number of glandular trichomes in their secretory stage was determined by imaging leaves with an optical scanner at 1200 dpi (HP Scanjet 8000, Hewlett Packard, Palo Alto, CA, USA). All visible glands were counted and calculated per cm^2 on both sides of the leaf. Three

developmental stages of leaves were chosen: not fully expanded (<1 week), mature expanded (1-2 weeks) and old expanded (> 3 weeks).

GC-MS Analysis of plant volatiles

Products of terpene synthase assays were identified by gas chromatography (Agilent Hewlett-Packard 6890, Agilent Technologies, Santa Clara, CA, USA) coupled to a mass spectrometer (Agilent Hewlett-Packard 5973, Agilent Technologies) or a flame ionization detector (FID). For analyses, 2 µl of pentane or ethyl acetate:pentane (2:1) extracts were injected with an injector temperature of 230 °C. Alternatively, a solid phase micro extraction fiber exposed to leaf volatiles (30 min, 30 °C) was introduced into the injector at a temperature of 180 °C. The terpenes were separated on a DB5-MS column (30 m length, 0.25 mm inner diameter and 0.25 µm film (J&W Scientific, Santa Clara, CA, USA); GC-program 40 °C for 2 min, first ramp 5 °C min^{-1} to 175 °C, second ramp 90 °C min^{-1} to 250 °C, final 3 min hold). GC-MS carrier gas: helium at 1 ml min^{-1}; GC-FID carrier gas: hydrogen at 2 ml min^{-1}. All terpene products were identified by using Agilent Technologies software with the Wiley275.L and NIST98.L MS libraries, as well as by comparison of mass spectra and retention times with those of authentic standards (Sigma-Aldrich Chemicals, Steinheim, Germany). The amounts of the individual terpenes were determined by GC-FID with monoterpene standards. Spearman's rank correlation coefficient was calculated between the terpene amounts and transcript levels.

Isolation of peltate glandular trichome clusters and cDNA library construction

Peltate glandular trichomes were isolated from young leaves using a method modified from the literature (Gershenzon et al., 1992). In brief, approx. 7 g of young, not fully expanded leaves were harvested, soaked in ice cold, distilled water containing 0.05 % Tween 20 for 2 h. The water was then decanted and the leaves were washed twice with ice-cold, distilled water, and abraded using a cell disrupter (Bead Beater, Biospec Products, Bartlesville, USA). The chamber was filled with the plant material, 65 ml of glass beads (0.5-1.0 mm diameter), XAD-4 resin (1 g/g plant material), and ice-cold extraction buffer (25 mM HEPES pH 7.3, 12 mM KCl, 5 mM $MgCl_2$, 0.5 mM K_2HPO_4, 0.1 mM $Na_4P_2O_7$, 5 mM DTT, sucrose (2.4 g l^{-1}), D-sorbitol (26.4 g $^{-1}$), methyl cellulose (6 g l^{-1}), and polyvinylpyrrolidone (10 g l^{-1}, PVP; M_r 40,000)) to full volume. Glands were abraded from the leaves with three pulses of 1 min at medium speed, with a 1 min pause between pulses. Following abrasion, the glands were separated from leaf material, glass beads, and XAD-4 resin by passing the supernatant of the chamber through a 500 µm mesh cloth (SEFAR NITEX, Sefar AG, Heiden, Switzerland). The residual plant material and beads were rinsed twice with 10 ml ice-cold isolation buffer (extraction buffer without methylcellulose) and passed through the 500 µm mesh. The combined 500 µm filtrates were

then consecutively filtered through membranes with decreasing mesh size (350 µm, 200 µm, and 100 µm). Finally, clusters of secretory cells (approx. 60 µm in diameter) were collected by passing the 100 µm filtrate through a 20 µm mesh. An aliquot of the isolated cell clusters was checked for integrity and purity with a light-microscope before being transferred to a 1.5 ml reaction tube and frozen in liquid nitrogen prior to RNA extraction.

RNA from isolated glandular trichomes was prepared and further purified using Trizol reagent (Invitrogen, Carlsbad, USA) and the RNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. A directional cDNA library was constructed from 1µg total RNA using the Creator Smart cDNA library construction kit (BD Bioscience Clontech, Mountain View, CA, USA) following the manufacturer's protocol, except using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) to perform first strand synthesis of the cDNA.

cDNA library Sequence analysis

Sequencing of 2,364 clones from a cDNA library of *O. vulgare* cultivar f02-04 was performed at the Purdue University Genomics Core Facility (Purdue University, West Lafayette, IN, USA). Sequences were assembled with the SeqMan program (Lasergene DNAStar V5.05, Madison, USA) and the resulting contigs were compared by BLAST search via the NCBI sequence database (http://www.ncbi.nlm.nih.gov/BLAST/). Sequences with similarities to terpene synthase genes from other plants were used to design primers for the isolation of full length cDNA clones by RACE-PCR. For this procedure, the primers were used with the cDNA libraries of the cultivars d06-01 and f02-04 and the BD SMART RACE cDNA Amplification Kit (BD Bioscience Clontech, Mountain View, CA, USA). The components of the PCR reaction were: 0.8 µl Adv. Taq DNA Polymerase Mix (5 U/µl), 5 µl 10 x Adv. Taq PCR-buffer, 1 µl dNTPs (10 mM each), 5 µl universal primer mix and 1 µl gene-specific primer (10 pmol/µl), 0.5-1 µl cDNA and PCR grade water added to a final volume of 50 µl. The PCR was conducted with an initial denaturation at 94 °C for 2 min, 30-35 cycles of denaturation at 94 °C for 30 s, annealing ranging from 52 °C to 57 °C for 30 s, extension at 68 °C for 60 to 150 s, and a final step at 70 °C for 5 min. PCR fragments were analyzed by cloning into pCR4-TOPO vector (TOPO TA cloning kit for sequencing, Invitrogen, Carlsbad, CA, USA) and subsequently sequenced. RACE-PCR was repeated several times to verify the correct 5' and 3' ends of the cDNAs. cDNA fragments from at least two independent RACE-PCR reactions were fully sequenced to prevent sequence errors.

The SeqMan program (Lasergene DNAStar V5.05, Madison, WI, USA) was used for contig assembly. The resulting contigs were compared to nr/nt nucleotide collection databases using the BLASTN search algorithm. All amino acid alignments were conducted using ClustalX

(protein weight matrix: Gonnet series; gap opening: 10.00; gap extension: 0.20; delay divergent sequences: 30 %) (Thompson et al., 1997). The resulting tree was constructed by the neighbor joining algorithm with a bootstrap sample of 1000 and visualized by Treecon 1.3b (Van de Peer and De Wachter, 1994).

RNA extraction from leaf material and RNA hybridization analysis

Whole leaf RNA was isolated from the same frozen, homogenized leaf material of *O. vulgare* that had been used for terpene extractions in June 2006. RNA was isolated with the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and quantified by spectrophotometry and gel electrophoresis. Total RNA (6 µg) was run on a denaturing RNA gel (1 % (w/v) agarose, 10 % (v/v) NorthernMax 10 x denaturing gel buffer, Ambion, Austin, TX, USA), for approx. 70 min at 100 V. The RNA was blotted onto a Hybond-XL membrane (Amersham, Piscataway, NJ, USA) by capillary transfer using NorthernMax 10 x running buffer (Ambion, Austin, Texas, USA) overnight. RNA was UV-crosslinked to the membrane two times with a fluence of 120 mJ cm^{-2}. For both prehybridization and hybridization, UltraHyb buffer (Ambion, Austin, TX, USA) was used. Probes were labeled with ([α^{32}P]dATP) using the Strip-EZ PCR Kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. Probes were amplified as ssDNA from short fragments of terpene synthase open reading frames. Fragments with lengths from 424 to 577 bp and large sequence differences from corresponding regions of other terpene synthase genes were chosen (*Ovtps*1: bp 1223-1699, *Ovtps*2: bp 1085-1635, *Ovtps*3: bp 1156-1595, *Ovtps*4: bp 1103-1526, *Ovtps*5: bp 1210-1647, *Ovtps*6: bp 975-1551, for primer details see Supplementary Material, Table S1). All probes were designed using gene sequences from cultivar d06-01. Probes were hybridized overnight at 42 °C. Membranes were washed twice with 3x SSC at 42 °C, once with 1x SSC at 42 °C, once with 0.1 x SSC at 42 °C, once with 0.1 x SSC at 50 °C and depending on the signal strength once with 0.1 x SSC at 55 to 68 °C. All SSC buffers contained 0.1 % SDS. Washed blots were sealed in PVC bags and blots were exposed to BioMax MS films (Kodak, Carestream Health, Rochester, NY, USA) with an intensifying screen for 14-15 h at -80 °C. Films were developed in a film developer (Konica medical film developer SRX-101A, Konica, Tokyo, Japan) according to the manufacturer's instructions. Additionally, storage phosphor screens (Amersham Bioscience, Uppsala, Sweden) were exposed to the blots for 4-5 h and analyzed with a Storm 840 scanner (Molecular Dynamics, Sunnyvale, CA, USA).

RNA isolation from leaf material for qRT-PCR

Total RNA was extracted from homogenized *O. vulgare* leaf material harvested in November 2009. The RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). To remove residual genomic DNA, RNA was treated with RNAse-free DNAse I (Qiagen, Hilden Germany). RNA samples were analyzed on an Agilent Bioanalyzer 2100 and RNA 6000 Nano Labchip using the Expert software (Agilent version B.02.02.SI258) to determine qualtity, integrity and rRNA ratios. RNA was quantified by spectrophotometry.

For cDNA synthesis, Superscript III reverse polymerase (Invitrogen, Carlsbad, California, USA) was used according to the manufacturer's instructions but with reverse transcription of 3-5 µg total RNA in a 2 x scaled-up reaction volume.

Transcript quantification by relative and absolute qRT-PCR

All experiments were performed on a Stratagene Mx3000P (La Jolla, California, USA) using SYBR green I with ROX as an internal loading standard. Each 25-µl reaction contained cDNA corresponding to 2.5 ng total RNA. Controls included non-RT controls (using 2.5 ng total RNA without reverse transcription to monitor for genomic DNA contamination) and non-template controls (water template).

PCR thermocycles were run as follows: 10 min at 95 °C, 40 cycles of 30 s at 95 °C, 1 min at 56 °C, and 1 min at 72 °C. Fluorescence was read following each annealing and extension phase. All runs were followed by a melting curve analysis from 55-95 °C. The products of each primer pair were cloned and sequenced at least six times to verify primer specificity. The linear range of template concentration to threshold cycle value (C_t value) was determined by performing a series of sixfold dilutions (1- to 1,296-fold) using cDNA from three independent RNA extractions analyzed in three technical replicates. All primers were designed using BeaconDesigner (version 5.0; PremierBiosoft, Palo Alto, California, USA) and HPLC-purified (Invitrogen, Carlsbad, California, USA). Primers were designed for regions identical in all known alleles for the respective gene (for primer details see Supplementary Material, Table S1). Primer efficiencies for all primers pairs were calculated using the standard curve method (Pfaffl, 2001). The stability of reference gene expression in the different plant lines was tested by comparing C_t values between all lines with cDNA corresponding to 2.5 ng total RNA. All amplification plots were analyzed with the MX3000P^{tm} software to obtain C_t values. For relative qRT-PCR elongation factor 1 alpha (*OvEF1alpha*) was employed as housekeeping gene. Relative transcript values were calculated using plant line f5 as calibrator.

Absolute quantification of *Ovtps*2 and *Ovtps*5 copy number in each cDNA sample was conducted using a standard curve and the results were normalized against 1 mg fresh weight of

plant material. The standard curve was generated with purified plasmid containing the respective genes. A 10-fold dilution series from 10^1 to 10^6 copies was determined.

Functional expression of terpene synthases

The complete open reading frames of the putative terpene synthases *Ov*TPS1 through *Ov*TPS7 and corresponding 5' truncations were cloned into the bacterial expression vector pH9GW. This vector is a derivative of the pET-T7 (28a) vector (Novagen, Madison, WI, USA) containing a 9 x His-tag followed by a Gateway attR cassette (Yu and Liu, 2006). Gene amplification was performed with primers (see Supplementary Material, Table S1) using the Advantage Taq Polymerase (BD Bioscience Clontech, Mountain View, CA, USA) as directed by the manufacturer. The amplification products were cloned into pCR4-TOPO vector (Invitrogen, Carlsbad, CA, USA). Expression constructs were created with Gateway technology (Invitrogen, Carlsbad, CA, USA). Validated sequences were subcloned from the pCR4-TOPO vector into the pDONR207 vector with BP Clonase II (Invitrogen, Carlsbad, CA, USA). Subsequently, the constructs were cloned into the pH9GW expression vector with LR Clonase II according to the manufacturer's instructions. The expression constructs were verified by sequencing and transformed into the BL21 (DE3) strain (Invitrogen, Carlsbad, CA, USA). For gene expression, a starter culture of 5 ml Luria-Bertani (LB) medium with 50 µg ml^{-1} kanamycin was grown overnight at 37 °C. Starter cultures (50 µl) were used to inoculate 125 ml auto-induction terrific broth (TB) media (Overnight Express Instant TB Medium, Novagen, Madison, WI, USA) with 50 µg ml^{-1} of kanamycin. Cells were harvested by centrifugation for 5 min at 5500 x *g*. Pellets were resuspended in 3 ml extraction buffer (50 Mm MOPSO buffer at pH 7.0, 10 % (v/v) glycerol, 5 mM DTT, 5 mM Na-ascorbate, 0.5 mM PMSF) and disrupted by sonification (Bandelin Sonopuls HD 2070, Bandelin Electronics, Berlin, Germany) for 3 min, cycle 2, power 65 %. Cell fragments were removed by centrifugation at 16,100 x *g* for 30 min at 4 °C. The supernatant contained the expressed proteins. Buffer exchange into assay buffer (20 mM MOPSO buffer at pH 7.0, 10 % (v/v) glycerol, 1 mM DTT) was performed with 10DG columns (BioRad, Hercules, CA, USA). Enzyme activities were assayed in 1 ml volumes containing 200 µl of the extracted enzyme. For multi-enzyme assays, 50 to 175 µl of the individual enzyme extracts were added. Assays contained 10 µM GPP or FPP (both Echelon, Salt Lake City, USA). If not mentioned otherwise, metal ions were added at the concentrations 1 mM $MgCl_2$ and 0.5 mM $MnCl_2$ and phosphatase inhibitors at 0.1 mM $NaWo_4$ and 0.05 mM NaF. Terpene products were collected by solid phase micro extraction with a polydimethylsiloxane coated fibre. The fibre was exposed for 30 to 60 min to the head space above the assay mixture at 30 °C in a water bath.

Biochemical characterization of terpene synthases

For biochemical characterization, *Ov*TPS2-d0601 was cloned into the pHIS8-3 vector (Jez et al., 2000) using the Nco1 and Not1 restriction sites. Two liters of bacterial culture were harvested by centrifugation for 5 min at 5500 x *g* and resuspended in 100 ml lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, 1 % (v/v) Tween 20, 10 % (v/v) glycerol, 10 mM β-mercaptoethanol, 0.5 mg ml^{-1} lysozyme, DNAse I) and stirred for 2h at 4 °C. The cell suspension was centrifuged 1h at 22,000 x *g* and 4 °C. HIS-tagged protein was bound to 2 ml Ni-NTA agarose (Qiagen, Hilden, Germany), washed with 30 ml lysis buffer and 30 ml washing buffer identical to lysis buffer but without Tween 20. The protein was eluted with 10 ml elution buffer identical to washing buffer but with 250 mM imidazole and fractionated as aliquots of 1 ml. The fractions with the highest protein concentrations were pooled. Salt concentration was reduced by dialysis (dialysis tubing cutoff 8 kDa) overnight in dialysis buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM β-mercaptoethanol). Protein concentrations were measured according to (Bradford, 1976) by using the BioRad reagent (BioRad, Hercules, CA, USA) with bovine serum albumin (Biorad, Hercules, CA, USA) as standard according to the manufacturer's instructions. The terpene synthase activity assays were performed as described above with 1.75 µg purified *γ*-terpinene synthase protein (*Ov*TPS2). Assays were overlaid with 200 µl pentane containing 0.9 ng μl^{-1} *δ*-cadinene as an internal standard. After 60 min at 30 °C the assays were shaken at 1400 rpm for 2 min to partition the terpene volatiles into the solvent phase. The vial was frozen at -80 °C for 30 min to freeze the aqueous phase. The solvent phase was taken off for terpene product analysis and identification by gas chromatography.

2.4 Results

Origanum vulgare plant lines show high variability in their essential oil composition

To assess the variability of the terpene content of *Origanum vulgare*, eighteen clonal lines derived from two cultivars were selected for this study. Fifteen of these lines were chosen for terpene extraction and RNA hybridization analysis. Three of the lines, d2, d5, and d8, were clones of the cultivar d06-01 while the lines f2, f3, f4, and f5 were clones of the cultivar f02-04.

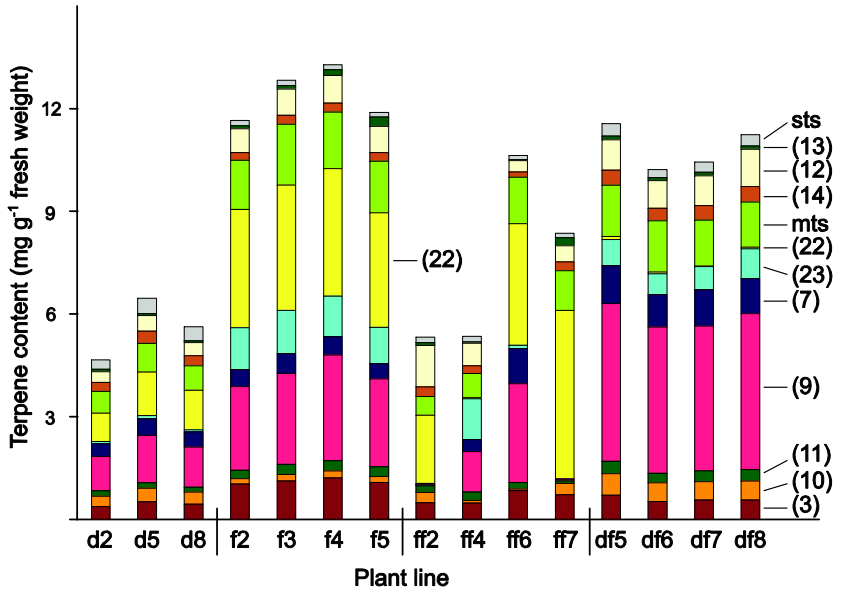


Figure 1 Terpene composition of the *O. vulgare* lines studied. The terpene profiles of three clonal lines of the cultivar d06-01 (d2, d5, d8), four clonal lines of cultivar f02-04 (f2, f3, f4, f5), four lines from a selfing of the cultivar f02-04 (ff2, ff4, ff6, ff7), and four lines from a cross of both cultivars (df5, df6, df7, df8) were measured. Monoterpenes and sesquiterpenes were extracted, identified by GC-MS and quantified by GC-FID. The trace of the flame ionization detector is shown. The compounds were identified as: 3, sabinene; 7, *p*-cymene; 9, γ-terpinene; 10, *cis*-ocimene; 11, *trans*-β-ocimene; 12, (-)-germacrene D; 13, bicyclo-germacrene; 14, (*E*)-β-caryophyllene; 22, *trans*-sabinene-hydrate; 23, thymol; mts, other monoterpenes: α-thujene, α-pinene, camphene, β-pinene, myrcene, α-terpinene, (+)*R*-limonene, β-phellandrene, *cis*-sabinene-hydrate, α-terpineol, carvacrol; sts, other sesquiterpenes: α-humulene, alloaromadendrene, 1,6-germacradiene-5-ol.

In addition to these parental clones, four other lines were generated by selfing of cultivar f02-04 (ff2, ff4, ff6, and ff7) and four further lines from a cross between both parent cultivars (df5, df6, df7, and df8). The total amounts of monoterpenes and sesquiterpenes in these lines ranged from 4.7 mg g^{-1} fresh weight for the cultivar d06-01 to a maximum of 13.4 mg g^{-1} in the cultivar f02-04 (Figure 1). This amount correlated with the density of peltate glandular trichomes on the leaf surfaces. A density of 951 ± 31 glandular trichomes cm^{-2} was measured on cultivar f02-04 on the surface of young leaves less than one week of age, while leaves of the cultivar d06-01

displayed only 560 ± 32 glandular trichomes cm^{-2} on comparable tissue. This two-fold difference in gland density was also observed in older leaves between one and three weeks of age. In all *O. vulgare* lines, the essential oil was dominated by monoterpenes, which provided between 67.4 % (ff2) and 94.1 % (ff6) of the total oil content (Figure 1). The major monoterpene components were γ-terpinene, *trans*-sabinene-hydrate, *p*-cymene, sabinene, *cis*- and *trans*-β-ocimene and thymol. The remainder of the essential oil consisted mostly of the sesquiterpenes (*E*)-β-caryophyllene, α-humulene and (-)-germacrene D. Despite the differences in total amounts, the composition of the lines of the parental cultivars (d2, d5, d8, f2-f5) was rather similar. However, a much higher variation of the terpene composition was observed in the lines that resulted from the selfing of cultivar f02-04 (ff lines). For example, line ff6 showed reduced levels of both ocimenes and thymol compared to the f parental lines, while lines ff2 and ff7 lacked γ-terpinene, *p*-cymene and thymol almost completely. In line ff4, the ratio of γ-terpinene to thymol was altered in favor of thymol, compared to the f parental lines, and *trans*-sabinene-hydrate could not be detected. The lines resulting from a hybridization between the two cultivars (df5-8) showed less variability in terpene composition but had consistently high amounts of γ-terpinene and low amounts of *trans*-sabinene-hydrate (Figure 1) (Supplementary Material, Table S2).

Origanum vulgare contains multiple terpene synthase genes with similarity to other Lamiaceae terpene synthases

In order to identify terpene synthases responsible for terpene formation in *O. vulgare*, we generated cDNA libraries from the cultivars d06-01 and f02-04. To enrich for terpene synthase cDNAs, glandular trichome cell clusters, the site of essential oil biosynthesis, were sheared off the leaf surface, purified and subjected to RNA extraction and cDNA synthesis. Sequencing of 2364 random clones from a cDNA library of cultivar f02-04 resulted in 69 ESTs that displayed sequence similarity to those of terpene synthases from other plants in BLAST searches (Supplementary Material, Table S3). These EST fragments were extended by several rounds of RACE-PCR to yield six individual full length cDNA clones which were designated putative terpene synthases *Ovtps*1 through 6. With all cDNAs, RACE-PCR was repeated at least twice to confirm the 5' end of the open reading frame. *Ovtps*7 was not represented in the cDNA library but was isolated by PCR with primers for the full-length clone of *Ovtps*1 (Supplementary Material, Table S1). These putative terpene synthases were also isolated from cultivar d06-01 and showed amino acid identities ranging from 94.0 % to 99.3 % compared to the respective genes from f02-04. The *Ovtps*4 of cultivar d06-01 contained a frame shift mutation that leads to early termination of the open reading frame. Within each of the cultivars, apparently only one allele could be isolated for nearly all of the genes suggesting that only one allele is actively

transcribed. Alternatively, both genes could have alleles with identical sequence. Two alleles were found only for *Ovtps*5 from cultivar d06-01 but one of the alleles was missing part of the open reading frame due to a frame shift mutation. However, the assignment of apparent alleles can only be tentative and based on high sequence similarity as *O. vulgare* is not a species where there is much genomic sequence information.

All genes showed the sequence motifs characteristic of terpene synthases, notably the double arginine motif (RRx_8W) which is commonly found in the N-terminal domain of monoterpene synthases (Bohlmann et al., 1998; Dudareva et al., 2003; Pechous and Whitaker, 2004). A highly conserved, aspartate-rich DDxxD motif involved in binding the bivalent metal ion cofactor (Starks et al., 1997) was found present in the C-terminal domain of six of the genes. In *Ovtps*4, this motif is altered to DDxxE which was previously observed only in 5-*epi*-aristolochene synthase from *Capsicum annuum* (CAA06614) (Zavala-Paramo et al., 2000)

Transit peptides were predicted for *Ov*TPS1 (35 aa), *Ov*TPS2 (24 aa) and *Ov*TPS7 (32 aa) using the ChloroP prediction site (http://www.cbs.dtu.dk/services/ChloroP/) (Emanuelsson et al., 2000) which target these proteins most likely to the plastid, the location of monoterpene biosynthesis in other Lamiaceae, and so these were designated as putative monoterpene synthases (Turner et al., 1999) (Table 1).

Table 1 Properties of terpene synthase genes isolated from *Origanum vulgare* cultivars d06-01 and f02-04

	ORF length [bp]			
	d06-01	f02-04	Transit peptide	Major product type
*Ovtps*1	1803	1803	35 AA	monoterpenes
*Ovtps*2	1782	1782	24 AA	monoterpenes
*Ovtps*3	1689	1689	No	sesquiterpenes
*Ovtps*4	1666[1]	1665	No	sesquiterpenes
*Ovtps*5	1659[2]	1659[3]	No	no *in vitro* activity
*Ovtps*6	1662	1662	No	sesquiterpenes
*Ovtps*7	1785[3]	1785	32 AA	monoterpenes

[1]*Ovtps*4-d06-01 contains a frameshift mutation. [2]Both alleles of *Ovtps*5-d06-01 are inactive *in vitro*. One allele contains a frame shift mutation. [3]*Ovtps*5-f0204 and *Ovtps*7-d06-01 show no *in vitro* activity. AA = amino acids.

*Ov*TPS3 through *Ov*TPS6 did not display transit peptides. These are likely to be targeted to the cytosol, and so were tentatively designated as sesquiterpene synthases. The sequence identity among putative monoterpene synthases at the amino acid level was within the range of 58-90 % and among sesquiterpene synthases 48-57 %, but the groups of mono- and sesquiterpene synthases share only 27-32 % identity with each other. However, *Ov*TPS*5*, considered a sesquiterpene synthase for lack of a transit peptide, displays an amino acid sequence identity of

74.7 and 76.7 % to the putative monoterpene synthase *Ov*TPS2 from cultivars d06-01 and f02-04, respectively. The lack of a transit peptide in this gene was verified by CELI-PCR (Ren et al., 2005).

Comparisons of the putative *O. vulgare* monoterpene synthase genes *Ovtps*1, *Ovtps*2, and *Ovtps*7 to terpene synthase genes known from other plants showed approx. 60 % amino acid sequence identity to monoterpene synthases of other Lamiaceae, including myrcene synthase (AY693649) and fenchol synthase (AY693648) of *Ocimum basilicum* (Iijima et al., 2004), sabinene synthase of *Salvia officinalis* (AF051901) (Wise et al., 1998), linalool synthase of *Mentha citrata* (AY083653), pinene synthase of *Rosmarinus officinalis* (EF495245) and linalool synthase of *Perilla frutescens var. crispa* (AF444798). The sequences of *Ovtps*3, *Ovtps*4 and *Ovtps*6 also showed approx. 60 % amino acid identity to sesquiterpene synthases of other Lamiaceae including (-)-germacrene D synthase (AY693644) and selinene synthase (AY693643) from *Ocimum basilicum* (Iijima et al., 2004b).

O. vulgare terpene synthases produce a variety of mo noterpenes and sesquiterpenes

The putative terpene synthases genes were expressed in a bacterial system to characterize their enzymatic activity. Expression of the *Ov*TPS1 full length clone from both cultivars did not result in active protein (Figure 2a), and the deletion of the first 35 amino acids predicted to encode the transit peptide did not result in the formation of active enzyme. A previous study on a terpene synthase from another Lamiaceae, limonene synthase from *Mentha spicata*, demonstrated formation of active enzyme after bacterial expression (in *E. coli)* following removal of the sequence coding for the N-terminal region up-stream of the double arginine motif (Williams et al., 1998). A similar truncation of *Ov*TPS1 that removed 59 amino acids of the N-terminal region resulted in active enzymes from both the d06-01 and f02-04 cultivars. The truncated *Ov*TPS1 converted geranyl diphosphate (GPP) to the monoterpene olefins sabinene and β-phellandrene with minor amounts of myrcene, γ-terpinene, α-thujene and α-pinene (Figure 2b). *Ov*TPS2 was active both as a full length protein and after deletion of the predicted plastidial targeting sequence of 24 amino acids. *Ov*TPS2-d06-01 converted GPP to γ-terpinene (80% of total products) and the minor products, α-thujene (7.2%), α-terpinene (6%), myrcene (2.6%), sabinene (1.3%), (+)-*R*-limonene (1.2%), α-pinene (1%) and α-phellandrene (0.8 %), as well as trace amounts of *p*-cymene and β-pinene (Figure 2c).

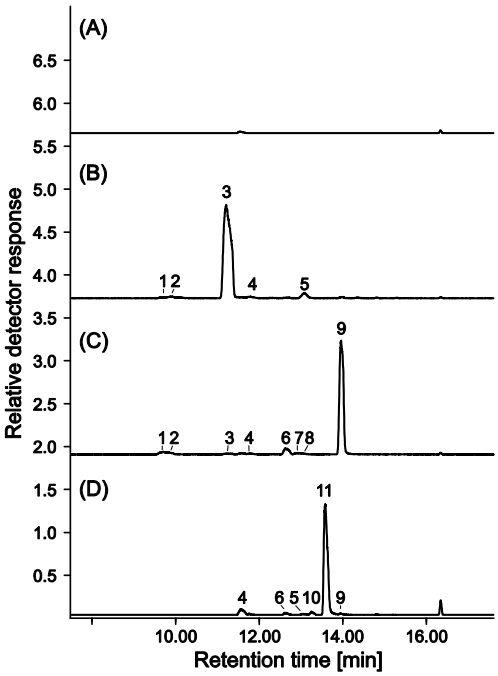


Figure 2 Monoterpene products of *Ov*TPS1, *Ov*TPS2 *Ov*TPS5 and *Ov*TPS7 measured *in vitro*. The enzymes were expressed in *E. coli*, extracted, and incubated with the substrate GPP and 0.5 mM Mn^{2+} ions. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) *Ov*TPS1 full length protein did not give any products, (**B**) Products of *Ov*TPS1 with a truncation of the N-terminal 59 amino acids: 1, α-thujene; 2, α-pinene; 3, sabinene; 4, myrcene; 5, β-phellandrene. (**C**) Products of the *Ov*TPS2 full length protein: 1, α-thujene: 2, α-pinene; 3, sabinene; 4, myrcene; 6, α-terpinene; 7, *p*-cymene; 8, (+)*R*-limonene; 9, γ-terpinene, (**D**) Products of *Ov*TPS7 with a truncation of the N-terminal 53 amino acids: 4, myrcene; 6, α-terpinene; 5, β-phellandrene; 10, *cis*-ocimene; 11, *trans*-β-ocimene and 9, γ-terpinene.

Since γ-terpinene is a major component of *Origanum vulgare* essential oil and a possible precursor of thymol (Poulose and Croteau, 1978a), we characterized the basic biochemical properties of this enzyme. *Ov*TPS2-06-01 showed the expected Michaelis-Menten saturation kinetics with K_m for GPP determined as 8.71 µM, a value typical for monoterpene synthases (Wise and Croteau, 1999). Mn^{2+} was preferred over Mg^{2+} as the bivalent metal ion cofactor with activity maxima at 1 µM and 25 µM, respectively. V_{max} for Mn^{2+} was 2.3-fold higher than for Mg^{2+} (Table 2). The K_m values for Mn^{2+} and Mg^{2+} are 0.695 mM and 3.4 mM, respectively. The pH optimum was determined as 6.8 with half maximal activities between 7.2 and 7.5, and between 6.0 and 6.5. The highest enzymatic activity was measured at 28 °C with half maxima at 20 and 35 °C. The lack of a transit peptide suggested that *Ov*TPS5 was a sesquiterpene synthase, yet this enzyme possessed high sequence similarity to other *O. vulgare* monoterpene synthases. However, in our bacterial expression system, the enzyme showed no activity in the presence of GPP or FPP, and could not be activated by a 5' truncation of amino acid residues just upstream of the RR motif (data not shown). The enzyme *Ov*TPS7 converted GPP to *trans*-β-ocimene and minor amounts of *cis*-β-ocimene, myrcene, α- and γ-terpinene. The same product profile was observed for the full-length protein and after an N-terminal deletion of 53 amino acids that truncated the protein two amino acids before the RR motif (Fig. 2d). None of the active monoterpene synthases, *Ov*TPS1, *Ov*TPS2 or *Ov*TPS7 converted farnesyl diphosphate (FPP) into sesquiterpene products (data not shown).

Table 2 Biochemical characterization of *Ov*TPS2 of cultivar d06-01, a *γ*-terpinene synthase**.** The kinetic constants for GPP were determined in the presence of 0.5 mM Mn^{2+} ions. Those for Mn^{2+} and Mg^{2+} were determined in the presence of 10 µM GPP. V_{max} for GPP is 6.18 µmol min^{-1} g $protein^{-1}$ and provides the reference point for measurements of V_{rel}. All values are ± SE

	K_m	V_{rel}
GPP	8.71 [µM]± 0.63	100
Mn^{2+}	0.70 [mM]± 0.02	104
Mg^{2+}	3.41 [mM]± 0.54	44

The remaining *O. vulgare* terpene synthase gene products were considered to be sesquiterpene synthases as they did not appear to have a transit peptide (Table 1, Supplementary Material, Fig. S1) and share high sequence identities with other sesquiterpene synthases. After heterologous expression, the terpene synthase *Ov*TPS3 converted FPP to (-)-germacrene D and trace amounts of bicyclo-germacrene (Figure 3a).

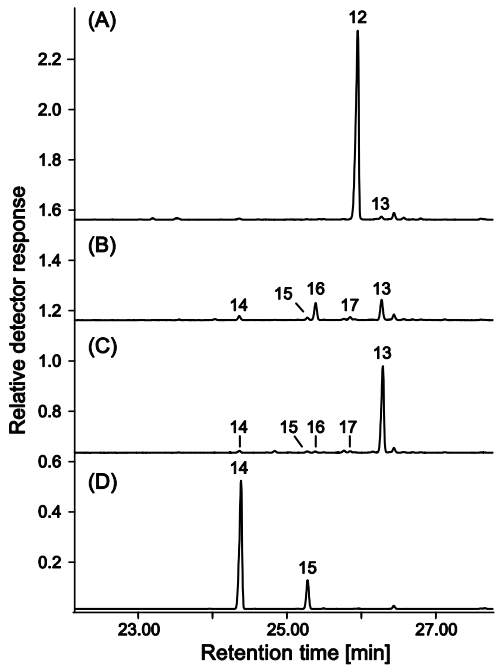


Figure 3 Sesquiterpene products of *Ov*TPS3, *Ov*TPS4 and *Ov*TPS6 measured *in vitro*. The enzymes were expressed in *E. coli*, extracted, and incubated with the substrate FPP and 1 mM Mg^{2+} or 0.5 mM Mn^{2+} ions. The resulting terpene products were identified by GC-MS. The total ion chromatogram is shown. (**A**) Products of *Ov*TPS3 in the presence of Mg^{2+} ions. (**B**) Products of *Ov*TPS4 in the presence of Mg^{2+} ions. (**C**) Products of *Ov*TPS4 in the presence of Mn^{2+} ions. (**D**) Products of *Ov*TPS6 in the presence of Mg^{2+} ions. Key to the terpene products were: 12, (-)-germacrene D; 13, bicyclo-germacrene, 14, (*E*)-*β*-caryophyllene; 15, *α*-humulene; 16, alloaromadendrene; 17, unknown sesquiterpene.

In the presence of GPP, the enzyme was also active producing moderate amounts of myrcene, limonene, terpinolene and linalool (Figure 4b). Expression of *Ov*TPS4 in *E. coli* resulted in an activity converting FPP to alloaromadendrene and bicyclo-germacrene (Figure 3b). However, product formation was dependent on the metal ions present in the *in vitro* assays. While alloaromadendrene and bicyclo-germacrene were produced in equivalent amounts in the presence of 1 mM magnesium ions, the formation of bicyclo-germacrene was greatly favored in the presence of 0.5 mM manganese ions (Figure 3c). The substrate GPP was converted by *Ov*TPS4 to terpinolene, limonene, geraniol and traces of myrcene, *α*-terpineol, *cis*-*β*-ocimene,

trans-β-ocimene and γ-terpinene (Figure 4c). Interestingly, the conversion of GPP was not affected by the presence of magnesium or manganese ions in the assay (Figure 4d). Expression of *Ov*TPS6 resulted in an enzyme producing (*E*)-β-caryophyllene and α-humulene from FPP (Figure 3d). The enzyme was virtually inactive in the presence of GPP (Figure 4e).

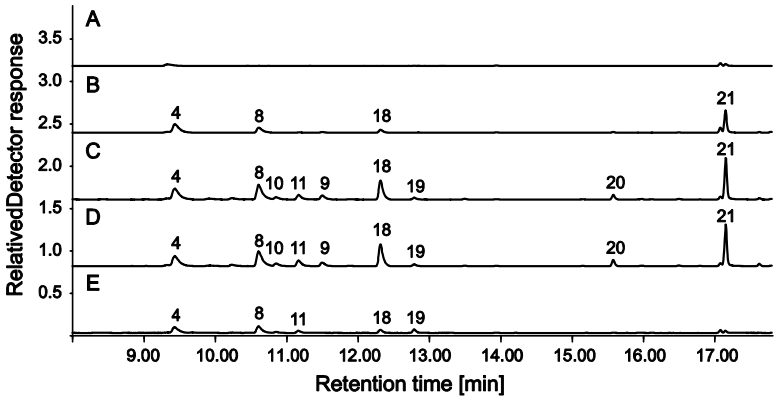


Figure 4 Monoterpene products of *Ov*TPS3, *Ov*TPS4 and *Ov*TPS6 measured *in vitro*. The enzymes were expressed in *E. coli*, extracted, and incubated with the substrate GPP. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry. The total ion chromatogram is shown. (**A**) Control vector without terpene synthase insert, (**B**) *Ov*TPS3 incubated in the presence of 1 mM Mg^{2+} ions, (**C**) *Ov*TPS4 incubated in the presence of 1 mM Mg^{2+} ions, (**D**) *Ov*TPS4 incubated in the presence of 1 mM Mn^{2+} ions, (**E**) *Ov*TPS6 incubated in the presence of 1 mM Mg^{2+} ions. Key to the products: 4, myrcene; 8, limonene, 9, γ-terpinene, 10, *cis*-ocimene; 11, *trans*-β-ocimene; 18, terpinolene; 19, linalool; 20, α-terpineol; 21, geraniol.

Terpene synthase transcript levels regulate monoterpene production in *O. vulgare*

To study the expression of each of the terpene synthases in *O. vulgare*, we first measured their transcript levels by hybridization analysis with total RNA from expanding leaves. Under the stringent hybridization conditions used, a cross reaction between the terpene synthases identified in this study is unlikely due to their sequence differences. Only *Ovtps*1 and *Ovtps*7 have a sequence identity of more than 90 % which does not allow for a clear separation of transcripts by RNA hybridization analysis even under the high stringency conditions used (Sambrook et al., 1989). The results showed clear terpene synthase transcript differences between the two cultivars. Lines of f02-04 had higher transcript levels for *Ovtps*1, *Ovtps*3, *Ovtps*5 and *Ovtps*6 than cultivar d06-01, but similar levels for *Ovtps*2 and *Ovtps*4 (Figure 5). The selfed lines of f02-04 (ff2, ff4, ff6, ff7) displayed a high variation in terpene synthase expression levels. For example, compared to the other selfed lines and lines of parent cultivar f02-04, lines ff2 and ff7 showed drastically reduced transcript levels for *Ovtps*2, while line ff6 had increased transcript levels of *Ovtps*5. The hybrid lines df5 through df8 exhibited transcript levels that were similar to those of cultivar f02-04 with lower levels of *Ovtps*3 and *Ovtps*4.

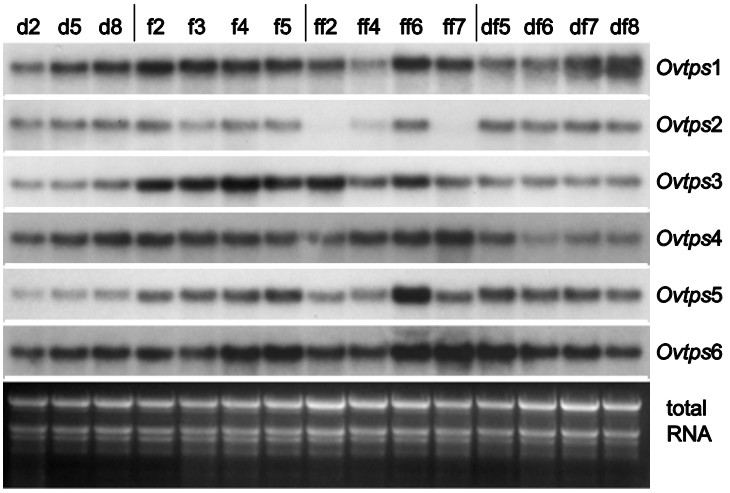


Figure 5 Expression analysis of *Ovtps1* through *Ovtps6 in planta.* Transcript levels of the terpene synthases in leaves of *O. vulgare* were measured by RNA hybridization analyses. The lines measured comprise the clones of cultivars d06-01 (d) and f02-04 (f), their crosses (df), and a selfing of f02-04 (ff). The lines are identical with those in Fig. 1. The bottom panel shows an ethidium-bromide-stained agarose gel with total RNA as control for equal RNA loading.

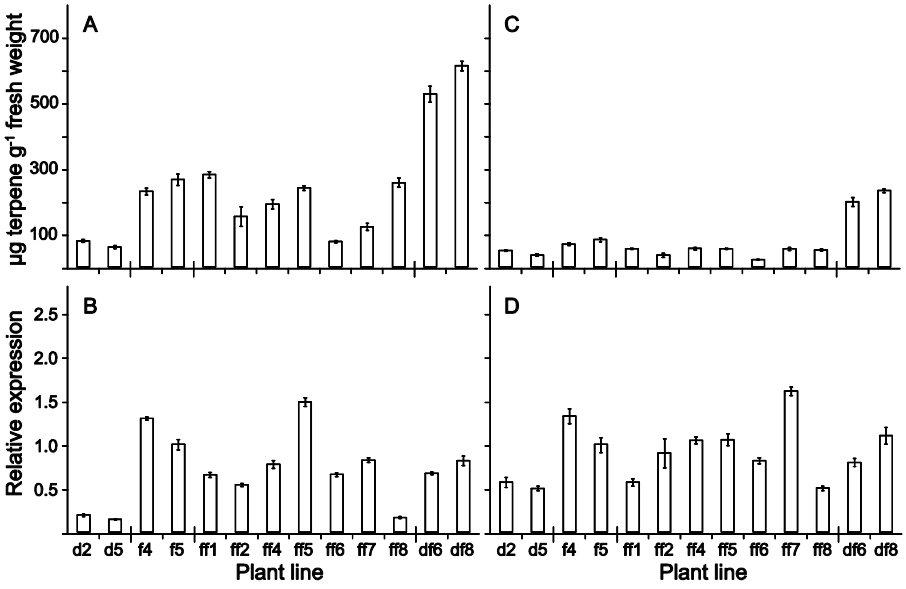


Figure 6 Correlation between relative transcript levels of *Ovtps*3 and *Ovtps*6 and sesquiterpene content in *O. vulgare* lines. (**A**) Amounts of germacrene D produced by selected *O. vulgare* lines. (**B**) Relative measurement of *Ovtps*3 transcript levels by qRT-PCR in same set of lines. (**C**) Amounts of *(E)-β*-caryophyllene produced by selected *O. vulgare* lines. (**D**) Relative measurement of *Ovtps*6 transcript levels by qRT-PCR in same set of lines. All experiments were done with three biological and three technical replicates per plant line. Plant line f5 was used as calibrator for experiments with rel. qRT-PCR.

In order to confirm and extend the results from RNA hybridization, relative and absolute qRT-PCR was employed for the monoterpene synthases *Ovtps*2 and *Ovtps*5 as well as the sesquiterpene synthases *Ovtps*3 and *Ovtps*6. For this experiments, two lines each of both cultivars and the hybrids (d2, d5, f4, f5, df6, df8) were chosen along with seven lines of the

selfed cultivar f02-04 (ff1, ff2, ff4, ff5, ff6, ff7, ff8) which displayed a high variation of terpene synthase expression. Relative quantification of transcripts of the four terpene synthases displayed a pattern similar to that of the RNA hybridization (Figures 6, 7).

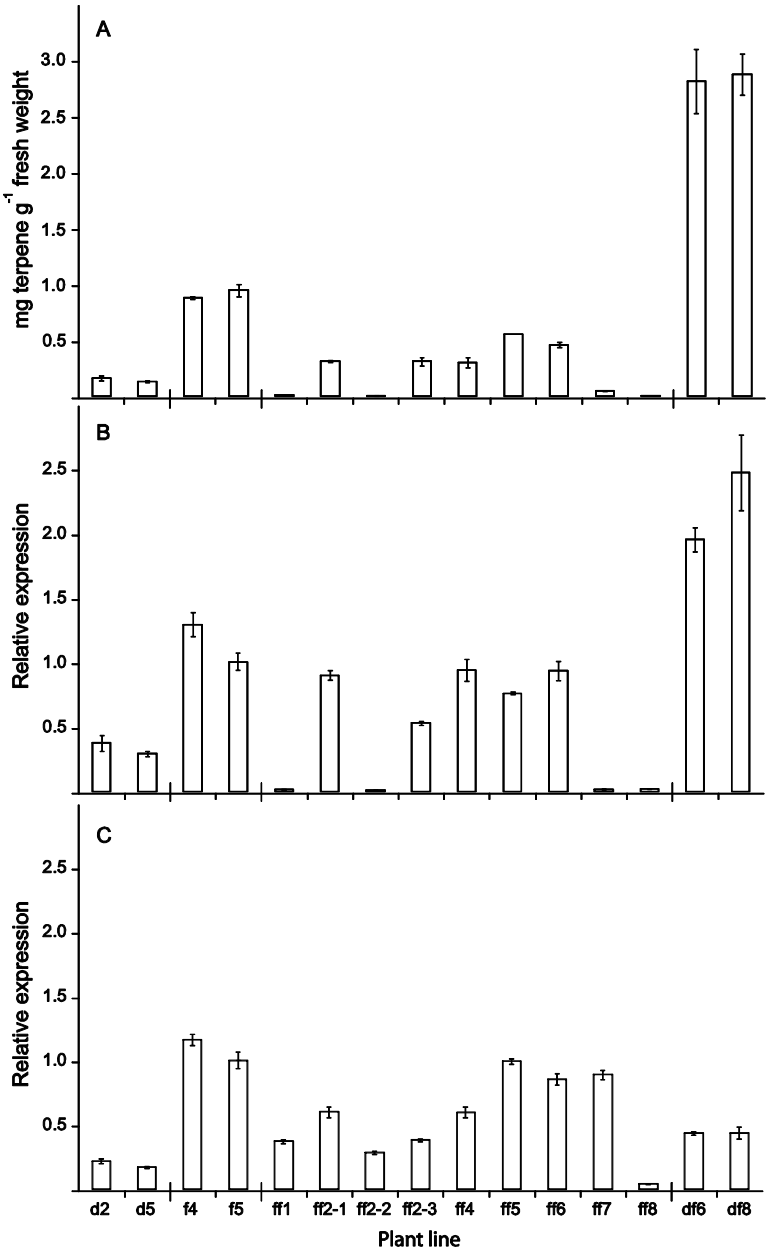


Figure 7 Correlation between relative transcript levels of *Ovtps*2 and *Ovtps*5 and γ-terpinene content in *O. vulgare* lines. (**A**) γ-Terpinene concentrations in selected *O. vulgare* lines. (**B**) Relative transcript levels of the γ-terpinene synthase *Ovtps*2 determined by qRT-PCR. (**C**) Relative transcript levels of the inactive putative terpene synthase *Ovtps*5 determined by qRT-PCR. Each bar represents mean values ±SE of three biological and three technical replicates except for individual plant lines ff2-1 to ff2-3 (n = 3). Plant line f5 was used as calibrator for all experiments in relative qRT-PCR.

Nevertheless, the better quantitative resolution of the qRT-PCR data allowed a correlation analysis between terpene synthase transcript levels and the terpene amounts *in planta*. Spearman's

correlation coefficient was calculated between the relative transcript levels and the major products of each of the terpene synthases (Table 3).

Table 3 Spearman's correlation coefficient r_s[1] values for the relationships between the amounts of various terpenes in *O. vulgare* lines and the relative or absolute transcript level of terpene synthases. (n = 39).

Terpene	Gene	Spearman's P	value	
α-terpinene	*Ovtps*2rel[2]	0.38	0.019	*
α-terpinene	*Ovtps*2abs[3]	0.43	0.006	**
γ-terpinene	*Ovtps*2rel	0.93	0.000	***
γ-terpinene	*Ovtps*2abs	0.89	0.000	***
thymol	*Ovtps*2rel	0.35	0.030	*
thymol	*Ovtps*2abs	0.38	0.019	*
γ-terpinene + *p*-cymene + thymol	*Ovtps*2rel	0.90	0.000	***
γ-terpinene + *p*-cymene + thymol	*Ovtps*2abs	0.86	0.000	***
(-)-germacrene D	*Ovtps*3rel	0.31	0.055	
sabinene	*Ovtps*5rel	-0.29	0.075	
sabinene	*Ovtps*5abs	-0.13	0.417	
γ-terpinene	*Ovtps*5rel	0.59	0.000	***
γ-terpinene	*Ovtps*5abs	0.53	0.000	***
(E)-β-caryophyllene	*Ovtps*6rel	0.302	0.061	

[1]The Spearman correlation coefficient quantifies the strength of the association between the variables and varies between -1 and +1. Correlations are significant if $P < 0.05$. [2]data from relative qRT-PCR. [3]data from absolute qRT-PCR.

For these sesquiterpene synthases, no correlation could be demonstrated. However, variation in sesquiterpene synthase transcript levels and sesquiterpene product formation among the plant lines studied are so low that it might be hard to detect any correlations. Therefore, a correlation between *Ovtps*3 and *Ovtps*6 transcript levels and their terpene products cannot be excluded with certainty. For the monoterpene synthase *Ovtps*2, a significant correlation was found ($r_s = 0.93$, $P < 0.001$) between relative and absolute transcript levels and the *in vitro* enzyme product *γ*-terpinene (Figure 7). A weak correlation ($r_s = 0.38$, $P < 0.05$) was also measured for *α*-terpinene. The three individual plants of line ff2 were plotted separately in Figure 7 due to the dramatic intra clonal variation in *γ*-terpinene content in this line. All other plant lines showed a rather low intra-clonal variation in both terpene content and transcript level and were therefore combined. The lines which lack the *Ovtps*2 transcript contain only trace amounts of the corresponding *γ*-terpinene terpene product. These traces detectable in lines ff1, ff2, ff7, and ff8

are most likely produced by *Ov*TPS1 or *Ov*TPS7 which both form small amounts of *γ*-terpinene upon expression *in vitro*.

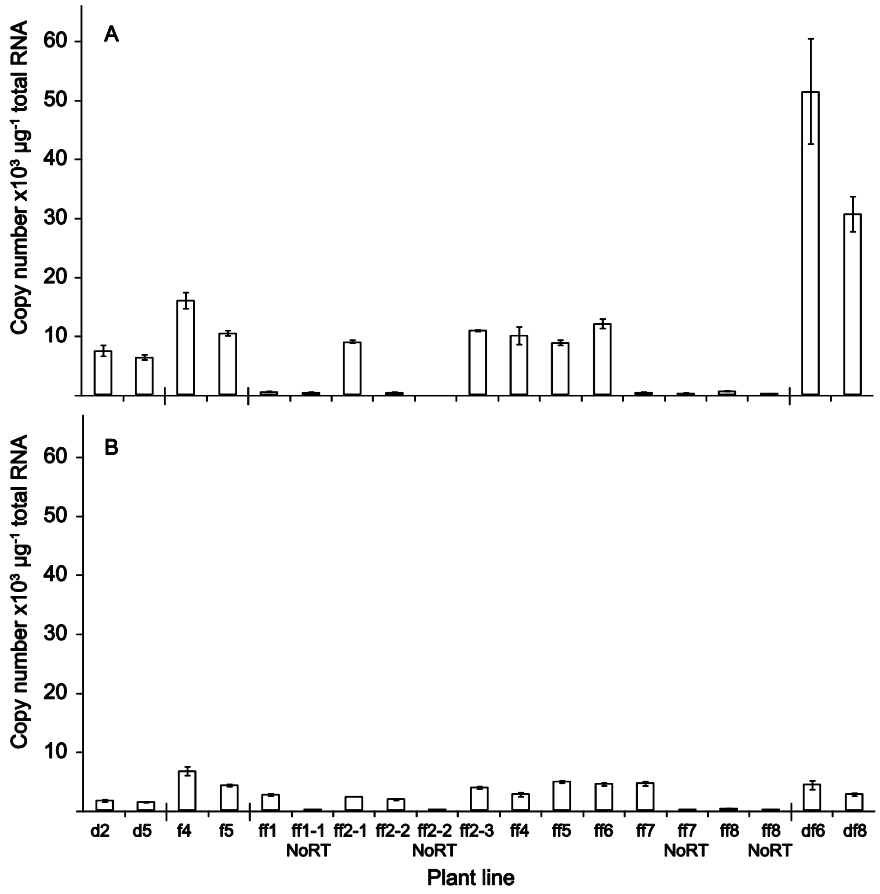


Figure 8 Absolute transcript levels of the *γ*-terpinene synthase *Ovtps*2 and the inactive putative terpene synthase *Ovtps*5 in *O. vulgare* lines. For both transcripts, the copy numbers per µg total RNA were determined by absolute qRT-PCR and normalized per mg fresh plant material of selected *O. vulgare* lines. (**A**) Absolute copy numbers of *Ovtps*2. (**B**) Absolute copy numbers for *Ovtps*5. Each bar represents mean values ±SE of three biological and three technical replicates except for individual plants ff2-1 to ff2-3 and non-RT (NoRT) controls (control with total RNA without reverse transcription) (n=3). No data were collected for the non-RT control of plant ff2-2 for *Ovtps*2.

To test whether *p*-cymene and thymol might be derived from the *Ovtps*2 product *γ*-terpinene as proposed (Poulose and Croteau, 1978a), we determined the correlation between the sum of these three terpenes and *Ovtps*2 transcript levels. This correlation was indeed strong (r_s = 0.90, $P < 0.001$) and supports the existence of such a biosynthetic pathway. However, only a weak correlation (r_s = 0.35, $P < 0.05$) could be established between *Ovtps*2 transcript level and thymol concentration alone, suggesting that if *γ*-terpinene is indeed converted to thymol, the later steps of this pathway are regulated independently of the terpene synthase step. Surprisingly, the transcript levels of *Ovtps5* which was inactive *in vitro*, also correlated with *γ*-terpinene production. However, this correlation was much weaker (r_s = 0.59, $P < 0.001$ than for *Ovtps*2. To further evaluate possible functions of *Ovtps*5 in oregano, absolute qRT-PCR similar to

(Palovaara and Hakman, 2008) was performed for both *Ovtps*2 and *Ovtps*5. The normalized copy numbers per µg total RNA for *Ovtps*5 ranged from 232 (line ff8) to 6,590 (line f4) which is relatively low compared to *Ovtps*2. For *Ovtps*2, copy numbers per µg total RNA ranged from 218 in line ff7 to 51,240 in line df6 (Figure 8). Copy number in non-RT controls was always lower but in a similar range as the low expression lines, e.g. line ff7 (Supplementary Material Table S4).

Most major terpenes of *O. vulgare* essential oil are produced by the terpene synthases *Ov*TPS1 through *Ov*TPS7

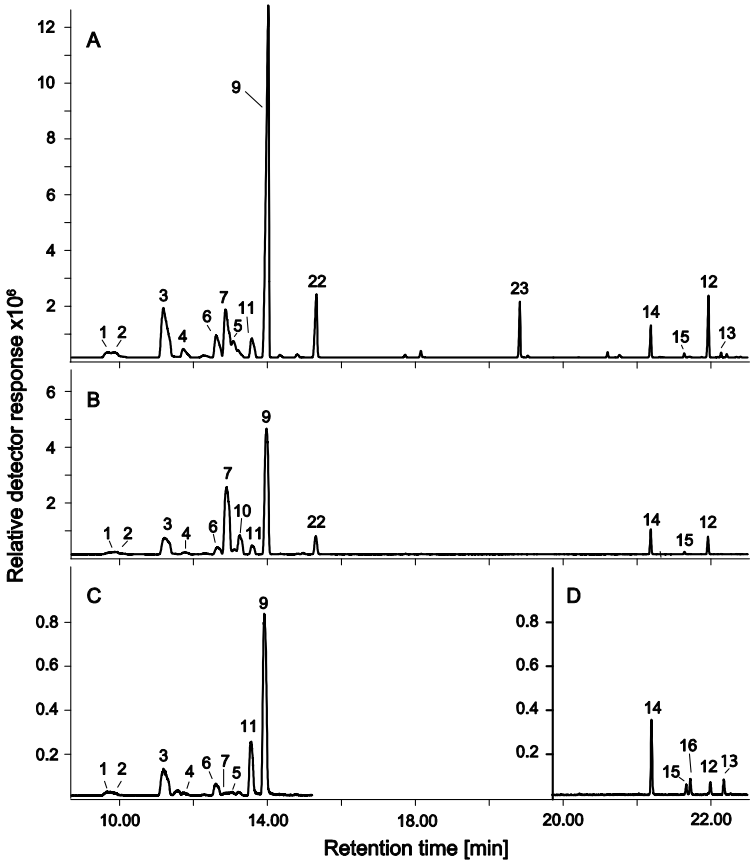


Figure 9 The terpene synthases *Ov*TPS1, *Ov*TPS2, *Ov*TPS3, *Ov*TPS4, *Ov*TPS6 and *Ov*TPS7 produce most major terpenes of the *O. vulgare* essential oil. The terpenes blends were separated by GC and analyzed by MS. The traces of the total ion chromatogram are shown. (**A**) Mono- and sesquiterpenes extracted from *O. vulgare* cultivar f02-04 (line f2). (**B**) Mono- and sesquiterpenes extracted from *O. vulgare* cultivar d06-01 (line d2). (**C**) Terpene products resulting from an *in vitro* assay containing *Ov*TPS1, *Ov*TPS2, and *Ov*TPS7. Equal volumes of bacterial extracts containing the heterologously expressed monoterpene synthases were mixed and assayed with GPP substrate and Mn^{2+} cofactor. (**D**) Terpene products resulting from an *in vitro* assay containing *Ov*TPS3, *Ov*TPS4, and *Ov*TPS6. Equal volumes of bacterial extracts containing the heterologously expressed sesquiterpene synthases were mixed and assayed with FPP substrate and Mg^{2+} cofactor.

The *in vitro* product spectra of the six active terpene synthases characterized correspond well to the full spectrum of mono- and sesquiterpenes in *O. vulgare* essential oil. To test the supposition that we had characterized essentially all the important terpene synthases involved in essential oil formation, we performed mixed assays combining all of the characterized monoterpene or sesquiterpene synthases. Assays containing all three monoterpene synthases (*Ov*TPS1, *Ov*TPS2 and *Ov*TPS7), in approximately equal protein levels along with GPP as a substrate and Mn^{2+} as a cofactor, produced a monoterpene profile almost identical to that of the plant oil (Figure 9a-c). Similar results were observed for a mixture containing all the sesquiterpene synthases (*Ov*TPS3, *Ov*TPS4 and *Ov*TPS6) with FPP as substrate and Mg^{2+} as cofactor, whose profile corresponded closely to the total sesquiterpene profile of *O. vulgare* essential oil (Fig. 9a, b, d).

2.5 Discussion

Six active monoterpene and sesquiterpene synthases were isolated from *O. vulgare*

Over sixty mono- and sesquiterpenes have been reported from the complex and highly variable essential oil of *Origanum vulgare* plants. Almost half of these compounds were extracted from the plant lines used in this study, including thymol, γ-terpinene, and *p*-cymene, which impart the characteristic oregano flavor (*Origanum vulgare* L.) and sabinene hydrate which is responsible for the flavor of marjoram (*Origanum majorana* L.) (Skoula and Harborne, 2002). To learn more about the biosynthesis of these terpenes and its regulation in *O. vulgare*, we investigated the terpene synthases of this species by isolating the members of this gene family, determining the enzymatic activity after heterologous expression in *E. coli* and correlating gene expression with the pattern of terpene accumulation in different *O. vulgare* lines. Terpene synthases convert the ubiquitous prenyl diphosphate intermediates, such as GPP and FPP, to monoterpene and sesquiterpene products, some of which undergo further oxidation. Here, we isolated seven terpene synthase genes, all of which gave active proteins after heterologous expression except *Ov*TPS5. *Ov*TPS1, *Ov*TPS2 and *Ov*TPS7 are considered monoterpene synthases because they convert the C_{10} substrate GPP to monoterpenes, but did not form any sesquiterpene products from the C_{15} substrate, FPP. In addition, these proteins show high sequence similarity to other monoterpene synthases and all contain a transit peptide for targeting to the plastid, a common organelle for monoterpene formation. The other characterized terpene synthases, *Ov*TPS3, *Ov*TPS4 and *Ov*TPS6 are considered sesquiterpene synthases. Even though most of these enzymes form monoterpenes from GPP as well as sesquiterpenes from FPP, their greater sequence similarity to other sesquiterpene synthases vs. monoterpene synthases, and the lack of any transit peptide (the cytosol is the most common site for sesquiterpene formation), suggest they function *in planta* as sesquiterpene synthases.

The terpene synthases described produce most of the terpene constituents of *O. vulgare* essential oil

The six active terpene synthases characterized here produce the majority of terpenes found in *O. vulgare*. One major terpene product not formed by these enzymes is *trans*-sabinene hydrate. Despite the reports of a sabinene hydrate synthase activity in sweet marjoram (*Origanum majorana* L., previously: *Majorana hortensis* Moench*.)* (Hallahan and Croteau, 1988, 1989), no enzyme responsible for its formation from GPP could be identified in the present study. Another compound not directly formed by the characterized terpene synthases is the aromatic monoterpene alcohol thymol, which is predicted to be synthesized from γ-terpinene (a product of *Ov*TPS2) via *p*-cymene (Poulose and Croteau, 1978a; Poulose and Croteau, 1978b). Analysis of

the terpene content of the inbred lines ff2 und ff7 (Figure 1) supports this hypothesis as no thymol is found in lines that lack γ-terpinene and *p*-cymene. γ-Terpinene is most likely converted to thymol by the action of one or more cytochrome P450 oxidases, catalyzing a hydroxylation similar to that described for (-)-*S*-limonene in menthol biosynthesis in *Mentha* sp. (Lupien et al., 1999; Haudenschild et al., 2000). Conversion of γ-terpinene to thymol might proceed via a *p*-cymene intermediate which was detected in minor amounts in *in vitro* assays of the γ-terpinene synthase *Ov*TPS2. However, these low levels of *p*-cymene may be instead due to spontaneous conversion of γ-terpinene into *p*-cymene (Granger et al., 1964).

The enzyme responsible for γ-terpinene formation *in vitro*, *Ov*TPS2, is likely to be a major terpene synthase activity in *O. vulgare in vivo*. Between 25.2 and 48.4 % of the total terpene content of this species consists of compounds that are products of this enzyme *in vitro*. The role of *Ov*TPS2 in γ-terpinene formation *in vivo* is also supported by the fact that the *Ovtps2* gene was only present in lines that produced γ-terpinene, In addition, the biochemical properties of this enzyme closely resemble those of a γ-terpinene synthase extracted from thyme leaves (Poulose and Croteau, 1978b). The K_m of *Ov*TPS2 for GPP was 8.71 µM, close to the K_m of 14 µM observed for the thyme γ-terpinene synthase enzyme. The pH optimum of both enzymes (6.8) was identical. The sequence of the *Ov*TPS5 enzyme is quite similar to that of *Ov*TPS2, but the fact that *Ov*TPS5 was completely inactive *in vitro* indicates that it is unlikely to contribute to γ-terpinene formation *in vivo*. This conclusion is also supported by the results from RNA hybridization and qRT-PCR. Especially in the plant lines ff1, ff2, ff7, and ff8, the correlation between transcript pattern and terpene content excludes *Ovtps*5 from an active involvement in the formation of γ-terpinene. We have not been able to identify a terpene synthase responsible for the production of sabinene hydrate. Perhaps the sequence of this enzyme is too different from other terpene synthases to be detected from sequence comparison. It is also possible that the sequence of the elusive gene(s) is identical to that of other terpene synthases within the binding sites of the utilized primers. This could result in the suppression of PCR products with the cDNA of the desired gene.

O. vulgare monoterpene and sesquiterpene synt hases have different phylogenetic origins

The terpene synthases of *Origanum vulgare* fall into two clades separating monoterpene and sesquiterpene synthases from each other (Figure 10). The *O. vulgare* monoterpene synthases are joined by many monoterpene synthases from other Lamiaceae. However, only two other sesquiterpene synthases from Lamiaceae are found in the sesquiterpene synthase clade, because

these are the only two other Lamiaceae sesquiterpene synthases identified to date, both from *Ocimum basilicum*.

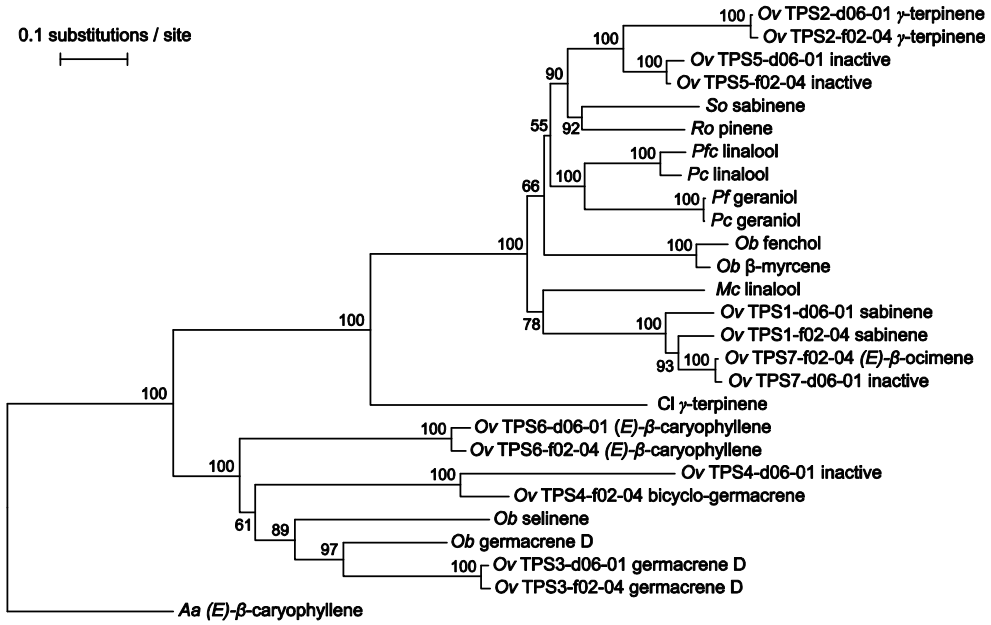


Figure 10 Dendrogram analysis of monoterpene and sesquiterpene synthases from *O. vulgare* and functionally related terpene synthases. The dendrogram was constructed using the neighbour-joining method. The name of the major product of each enzyme is given after the abbreviation of the species (or after the abbreviation of the gene and cultivar designation in case of the *O. vulgare* sequences). GenBank accession numbers are in parentheses. *Aa – Artemisia annua:* (*E*)-*β*-caryophyllene synthase (AF472361); *Cl – Citrus limon*: *γ*-terpinene synthase (AF514286); *Mc – Mentha citrata*: linalool synthase (AY083653); *Ob – Ocimum basilicum*: fenchol synthase (AY693648), myrcene synthase (AY693649), selinene synthase (AY693643), germacrene D synthase (AY693644); *Ov – Origanum vulgare*: TPS1-d06-01 (GU385980), TPS1-f02-04 (GU385979), TPS2-d06-01 (GU385978), TPS2-f02-04 (GU385977), TPS3-d06-01 (GU385976), TPS3-f02-04 (GU385975), TPS4-d06-01 (GU385974), TPS4-f02-04 (GU385973), TPS5-d06-01 (GU385972), TPS5-f02-04 (GU385971), TPS6-d06-01 (GU385970), TPS6-f02-04 (GU385969), TPS7-d06-01 (GU385968), and TPS7-f02-04 (GU385967 *Pf – Perilla frutescens*: geraniol synthase (DQ088667); *Pfc – Perilla frutescens* var. *crispa*: linalool synthase (AF444798); *Pc – Perilla citriodora*: linalool synthase (AY917193), geraniol synthase (DQ234300); *Ro – Rosmarinus officinalis*: pinene synthase (EF495245); *So – Salvia officinalis*: sabinene synthase (AF051901).

It is worth noting that the sesquiterpene synthase *Ov*TPS6, an (*E*)-*β*-caryophyllene synthase, displayed a low sequence identity with (*E*)-*β*-caryophyllene synthases from the Asteraceae, *Artemisia annua* (41.2 %) and the Cucurbitaceae, *Cucumis sativus* (44.0 %).

Two trends previously noted for other terpene synthases were also found to be true here. The monoterpene synthases of Lamiaceae share rather high amino acid identities (usually 55-90 % despite their different catalytic functions. Yet terpene synthases outside this family are much more divergent even though they might have the same catalytic function. For example, a *γ*-terpinene synthase from *Citrus limon* (included in Fig. 10) is clearly separated from *Ov*TPS2,

the γ-terpinene synthase of *O. vulgare,* with 37 % amino acid identity indicating that these genes are the result of repeated evolution (Gang, 2005). The phylogenetic analysis also indicates that *Ov*TPS1 and *Ov*TPS7 appear to result from gene duplication and neofunctionalization. *Ov*TPS2 and *Ov*TPS5 might also result from gene duplication followed by functional loss of *Ov*TPS5 due to inactivation accompanied by loss of the transit peptide.

For most of the characterized *O. vulgare* terpene synthases, only a single allele was apparently identified from the EST libraries, suggesting that the other allele of these genes is not transcribed or does not encode an active enzyme. This assumption is supported by our analyses of the selfed lines of *O. vulgare* cultivar f02-04. Since products of *Ov*TPS2 are completely absent in four of the studied lines, only one of the two *Ov*TPS2 alleles is likely to be active in the cultivar f02-04. For *Ov*TPS5, two alleles were found but none displayed any terpene synthase activity in the *in vitro* assay. The presence of catalytically inactive alleles appears to be typical for the terpene synthase gene family in other species as well (Köllner et al., 2004; Köllner et al., 2008). It is conceivable that *Ov*TPS2 and *Ov*TPS5 originate from a duplication event within the genome of *O. vulgare*. Subsequently, *Ov*TPS5 might have lost is function due to a frameshift mutation but still kept its expression pattern due to preservation of the promoter and other regulatory sequences. This is not consistent with the usual expectation that the regulatory sequences are more quickly altered than the structural gene. For example, the diversification of a pair of terpene synthases of strawberry resulted in altered spatial regulation (cytoplasmatic versus plastidic expression) but did not change the sequence of the structural gene (Aharoni et al., 2004). A high rate of diversification of genes involved in plant secondary metabolism has often been observed and both structural and regulatory changes might contribute to the large variety of terpene patterns found in plants (Köllner et al., 2004; Iijima et al., 2004b).

Terpene synthase gene expression determines terpene composition in *O. vulgare*

The isolated terpene synthase genes of *O. vulgare* appear to play a major role in controlling terpene composition in this species since the transcript levels of individual genes correlate closely with the amounts of the encoded enzyme products found in the essential oil.

The close correlation of γ-terpinene synthase expression and terpene composition indicates that transcript regulation of terpene synthase genes is the most important regulatory mechanism controlling terpene composition in *O. vulgare*. The results of the RNA hybridization assays suggest that this mechanism regulates the activity of other monoterpene synthases as well. In contrast, the low levels of sesquiterpenes in *O. vulgare* essential oil might not be regulated strictly on transcript level. Our correlation analysis of *Ovtps*3 and *Ovtps*6, however, was hampered by the limited variation of sesquiterpene content between the *O. vulgare* lines. Further support for the assertion of transcript level regulation at least for monoterpene formation

comes from the combined *in vitro* assays of the heterologously expressed terpene synthases. When approximately equal amounts of the expressed active terpene synthases were combined in the presence of GPP or FPP as substrates, the blends of monoterpenes or sesquiterpenes produced strongly resembled those of *O. vulgare* terpene blends with the absence of the monoterpene *trans*-sabinene hydrate and thymol (Figure 9). This suggests that terpene synthase expression levels directly control the composition of the essential oil, and provide no indication for operation of any further regulatory mechanisms like compartmentation or metabolite channeling at the site of *O. vulgare* terpene biosynthesis. However, control of total terpene yield may result from processes at other levels of organization. The *O. vulgare* cultivars used in this study, f02-04 and d01-06, showed major differences in the quantity of terpenes produced, with the quantity of cultivar f02-04 being approximately twice that of d01-06. This difference is likely caused by the fact that the leaf surface of f02-04 has a higher density of glandular trichomes, the sites of synthesis and storage of the monoterpenes and sesquiterpenes (Gershenzon et al., 1989; Turner et al., 1999).

Terpene synthases have value for molecular engineering of terpene biosynthesis in *O. vulgare*

Our results suggest several ways in which the terpene composition of *O. vulgare* could be modified by molecular methods to produce plants with greater value as culinary herbs or sources of pharmaceuticals. Given the regulatory importance of terpene synthases in this species, over-expression or suppression of these catalysts is a straightforward way of altering the concentrations of their products. Alteration of the product profile of a single terpene synthase requires modification of the reaction mechanism of the terpene synthase itself. This can be achieved by site-directed mutagenesis of the enzyme or breeding in other alleles of this gene. For a general increase in monoterpene production, it may be possible to overexpress enzymes of the plastidial methylerythritol phosphate pathway. For example, overexpression of deoxyxylulose phosphate reductoisomerase in mint resulted in a significant increase in monoterpene production (Mahmoud and Croteau, 2001). A prerequisite for all of these strategies is the development of a transformation system for *O. vulgare*. However, in the meantime, the terpene synthase genes can still be utilized as markers for directed breeding of *O. vulgare* varieties with higher value essential oil (Novak et al., 2008).

3 Chapter II

Cytochrome P450s participate in the biosynthesis of the phenolic monoterpenes, thymol and carvacrol, in oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.)

3.1 Abstract

Thymol and carvacrol are major aroma constituents of the essential oil of two culinary herbs, oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.). These phenolic monoterpenes have a broad range of biological activities acting as antimicrobial compounds, insecticides, anti-oxidants and pharmaceutical agents. A pathway for the biosynthesis of thymol from the mono-terpene *γ*-terpinene via an intermediate *p*-cymene was proposed in the late 1970s (Poulose and Croteau, 1978a) but has never been validated. Here, we demonstrate the involvement of cytochrome P450 monooxygenases in the conversion of *γ*-terpinene to thymol and carvacrol. We isolated eleven cytochrome P450 gene sequences from oregano, thyme and marjoram that were assigned to five gene names, *CYP71D178* through *CYP71D182*. The transcript levels of most of these genes are well-correlated with the occurrence of thymol and carvacrol. Heterologous expression of two of them in yeast resulted in active proteins catalyzing the formation of *p*-cymene, thymol and carvacrol from *γ*-terpinene. Since *p*-cymene itself was not accepted as a substrate, it is likely that *γ*-terpinene is directly converted to thymol and carvacrol with *p*-cymene as a side product. The properties and sequence motifs of these P450s are similar to those of well-characterized monoterpene hydroxylases isolated from mint.

3.2 Introduction

The Lamiaceae plant family contains many aromatic plants of great scientific and economic interest such as rosemary, sage, mint, marjoram, oregano and thyme. The aroma associated with these plants arises from the essential oil found in peltate glandular trichomes on the aerial parts of the plant. These glandular trichomes consist of highly specialized secretory cells in which the components of the essential oil are synthesized and subsequently accumulate in a subcuticular storage cavity (Gershenzon et al., 1989; Turner et al., 1999). The composition of the essential oils of oregano, thyme and marjoram is dominated by mono- and sesquiterpenes (Skoula and Harborne, 2002; Stahl-Biskup, 2002). These substances are responsible for the aroma and flavor of these herbs, and the extracted essential oils are used for the manufacturing of perfumes and cosmetics as well as for medicinal and pharmaceutical purposes as antimicrobial or antiseptic agents (Kintzios, 2002; Stahl-Biskup, 2002). Mono- and sesquiterpenes are also thought to help defend the plant against herbivores and pathogens (Gershenzon and Dudareva, 2007).

Two monoterpenes of the Lamiaceae that have attracted much attention are thymol and carvacrol which are often found in thyme and oregano. These two phenolic monoterpenes are especially known for their antiherbivore, antimicrobial, pharmaceutical and antioxidant activities (Isman, 2000; Hummelbrunner and Isman, 2001; Ultee et al., 2002; Sedy and Koschier, 2003; Braga et al., 2008). They are even used to treat bee hives against varroa mite without harming the bees (Floris et al., 2004).

The general outline of monoterpene biosynthesis is well known (Gershenzon, 1999; Wise and Croteau, 1999). First, the ubiquitous C_{10} intermediate, geranyl diphosphate (GPP), is converted by enzymes known as monoterpene synthases to cyclic or acyclic products. Then, oxidized monoterpenes are formed from these initial cyclic or acyclic products by reactions catalyzed frequently by cytochrome P450s monooxygenases (Gershenzon, 1999; Wise and Croteau, 1999). The oxidized monoterpenes, thymol and carvacrol, are most likely derived from one of the initial cyclic products, *γ*-terpinene, by oxidation (Poulose and Croteau, 1978a). In two cultivars of oregano (*Origanum vulgare* L.), *γ*-terpinene was recently demonstrated to be formed from GPP by a monoterpene synthase, *Ov*TPS2 (Crocoll et al., 2010). From *γ*-terpinene, the pathway for thymol formation is thought to proceed via the aromatic compound, *p*-cymene, as an intermediate (Fig. 1a) (Poulose and Croteau, 1978a). It seems conceivable that the conversions of *γ*-terpinene via *p*-cymene to thymol and carvacrol could be catalyzed by cytochrome P450s. Despite the interest in thymol and carvacrol as pharmaceuticals and plant defenses, no genes or enzymes responsible for the formation of these phenolic monoterpenes from *γ*-terpinene or *p*-cymene have been described to date.

Here, we report several lines of evidence supporting the involvement of cytochrome P450s in thymol and carvacrol biosynthesis. First, eleven new cytochrome P450 gene sequences are

described from oregano, thyme and marjoram which were assigned to five new cytochrome P450 genes, *CYP71D178* through *CYP71D182*. The expression levels of these genes, quantified by relative and absolute quantitative real-time PCR (qRT-PCR), were then correlated with thymol and carvacrol formation in different oregano (*Origanum vulgare* L.), thyme (*Thymus vulgaris* L.) and marjoram (*Origanum majorana* L.) plant lines. Finally, the expression of three of these P450 genes in yeast resulted activities that converted *γ*-terpinene to *p*-cymene as the major product, but also to thymol and carvacrol.

3.3 Materials and Methods

Plant Material

Oregano (*Origanum vulgare* L.) plants were propagated from stem cuttings and grown in the greenhouse with additional illumination by sodium lamps (16 h day) at 20-22 °C (day) and 18-20 °C (night), and relative humidity ranged from 30-55 %. Plants were potted in commercial soil (Tonsubstrat, Klasmann, Geeste / Groß-Hesepe, Germany) and watered every 2 to 3 days with tap water. Two cultivars of *Origanum vulgare* L. (cv. 'd06-01' and cv. 'f02-04') were selected from the collection of *Origanum* species of the botanical garden of Università di Bari, Italy. These were chosen for the presence of *γ*-terpinene, *p*-cymene, thymol, and carvacrol in the essential oil. An additional criterion was a low density of hairy trichomes which facilitates the extraction of glandular trichomes for RNA isolation. Herbarium specimens of the two genotypes are kept at the Herbarium of the Institute of Applied Botany, University of Veterinary Medicine, Vienna, Austria. For the cultivar d06-01, three clonal lines were chosen (designated d2, d5 and d8) and from the cultivar f02-04 four lines were chosen (designated f2 through f5). Additional lines were derived by selfing f02-04 plants (seven lines chosen designated ff1, ff2, ff4, ff5, ff6, ff7 and ff8) and crossing line d06-01 with f02-04 (four lines chosen designated df5 through df8).

Thyme (*Thymus vulgaris* L.) chemotypes T28 and L48 were kindly provided by J.D. Thompson (Thompson et al., 1998; Thompson, 2002). Thyme cultivar (cv. 'Tc'), oregano cultivar (cv. 'Ct') and marjoram (*Origanum majorana* L.) cultivar (cv. 'gT') were bought at local markets in Jena, Germany. These plants were chosen for their extreme differences in thymol and carvacrol contents compared to the above mentioned thyme chemotypes and oregano plant lines. All plants were grown in the greenhouse as described above.

Terpene extraction from leaves

For terpene extractions to compare with the measurement of gene expression by RNA hybridization, leaf material of the *O. vulgare* lines d2, d5, d8, f2-f5, df5-df8, ff2, ff4, ff6, ff7 was harvested in June. Young expanding leaves from five plants of each line were pooled and frozen in liquid N_2 immediately after harvest and ground to a fine powder with mortar and pestle. The powder (50-100 mg) was soaked in 1 ml ethyl acetate:pentane (2:1) containing an internal standard (menthol, 50 ng μl^{-1}) for 24 h at room temperature with constant rotation. The solution was cleared with activated charcoal for 5 min and dried over a column of 500 mg water-free Na_2SO_4. All extractions were performed in triplicate.

For terpene extractions to compare with the measurement of gene expression by qRT-PCR, leaf material of the lines d2, d5, f4, f5, df6, df8, ff1, ff2, ff4, ff5, ff6, ff7, ff8 was harvested in No-

vember, while plant material from the thyme cultivar 'Tc', oregano cultivar 'Ct' and marjoram cultivar 'gT' was harvested in April. Young expanding leaves from three plants of each line were individually harvested and extracted as described above. All extractions were done with three biological and three technical replicates except for thyme 'Tc', oregano 'Ct' and marjoram 'gT' where only one biological replicate was available.

GC-MS Analysis of volatiles

Products of terpene extractions from plants and from cytochrome P450 assays were identified by gas chromatography (Agilent Hewlett-Packard 6890, Agilent Technologies, Santa Clara, CA, USA) coupled to a mass spectrometer (Agilent Hewlett-Packard 5973, Agilent Technologies) or a flame ionization detector (FID). For analyses, 1 µl injections of pentane or ethyl acetate:pentane (2:1) extracts were made with an injector temperature of 230 °C. The terpenes were separated on a DB5-MS column: 30 m length, 0.25 mm inner diameter and 0.25 µm film (J&W Scientific, Santa Clara, CA, USA). The program had an initial temperature hold at 40, 50 or 65 °C for 2 min, first ramp 2-8 °C min^{-1} to 175 °C, second ramp 90 °C min^{-1} to 250 °C, final 3 min hold. Limonene enantiomers and hydroxylated products were further identified on a chiral column (HYDRODEX®-ß-3P, Macherey-Nagel, Düren, Germany): 25 m length, 0.25 mm inner diameter. The temperature program was 80 °C for 2 min, first ramp 2 °C min^{-1} to 165 °C, second ramp 50 °C min^{-1} to 200 °C, final 3 min hold. The GC-MS carrier gas was helium at 1 ml min^{-1} and the GC-FID carrier gas was hydrogen at 2 ml min^{-1}. All terpene products were identified by using Agilent Technologies software with the Wiley275.L, NIST98.L and Adams2205.L MS libraries, as well as by comparison of mass spectra and retention times with those of authentic standards (Sigma-Aldrich Chemicals, Steinheim, Germany). The amounts of the individual terpenes were determined by GC-FID by comparison with the peak areas of the internal standard. Spearman's rank correlation coefficient was calculated to determine the correspondence of the terpene amounts and transcript levels.

Identification of cytochrome P450 gene candidates

The first gene fragments for a cytochrome P450 were isolated by sequencing of 2,364 clones of a cDNA library of isolated peltate glandular trichome clusters from *Origanum vulgare* cultivar 'f02-04' For further details see (Crocoll et al., 2010). ESTs with similarities to the monoterpene hydroxylase CYP71D18 previously isolated from mint (Lupien et al., 1999) were chosen to design primers to isolate full length cDNA clones by RACE-PCR. For this procedure, primers were used with cDNA libraries of oregano cultivars d06-01 and f02-04 and the BD SMART RACE cDNA Amplification Kit (BD Bioscience Clontech, Mountain View, CA, USA). The components of the PCR reaction were: 0.8 µl Adv. Taq DNA Polymerase Mix (5 U $µl^{-1}$), 5 µl

10 x Adv. Taq PCR-buffer, 1 µl dNTPs (10 mM each), 5 µl universal primer mix and 1 µl gene-specific primer (10 pmol μl^{-1}), 0.5-1 µl cDNA and PCR grade water added to a final volume of 50 µl. The PCR was conducted with an initial denaturation at 94 °C for 2 min, followed by 30-35 cycles: denaturation at 94 °C for 30 s, annealing ranging from 54 °C to 57 °C for 30 s, and extension at 68 °C for 60 to 150 s, and a final step at 70 °C for 5 min. PCR fragments were analyzed by cloning into the pCR4-TOPO vector (TOPO TA cloning kit for sequencing, Invitrogen, Carlsbad, CA, USA) and subsequently sequenced. RACE-PCR was repeated several times to verify the correct 5' and 3' ends of the cDNAs. cDNA fragments from at least three independent RACE-PCR reactions were fully sequenced to prevent errors.

Further cytochrome P450 sequences were isolated from peltate glandular trichome cDNA libraries of oregano cultivar f02-04 (CYP71D179-f2) and *Thymus vulgaris* L. chemotypes L48 (CYP71D179v1-L48) and T28 (CYP71D180v1-T28) by RACE-PCR. Additional cytochrome P450 gene sequences were amplified from the cDNA of young expanding leaves of the oregano cultivar 'Ct' (CYP71D181-Ct1 and CYP71D180-Ct2), the thyme cultivar 'Tc' (CYP71D182-Tc1, CYP71D179v2-Tc2, CYP71D180-Tc3, CYP71D181-Tc4) and the marjoram cultivar 'gT' (CYP71D180v-gT1) employing primers for the 5' and 3' ends of the open reading frames of the already isolated cytochrome P450 sequences of CYP71D178, CYP71D179 and CYP71D180 (CYP71D178-182-fwd: 5'-GATGGATATTTCAATTTCATGGGT-3', CYP71D178-182-rev: 5'-ATTATGAGGTTGGATTGTGGATT-3').

The SeqMan program (Lasergene DNAStar V8.02, Madison, WI, USA) was used for contig assembly. Resulting contigs were compared to nr/nt nucleotide collection databases using the BLASTN search algorithm. All amino acid alignments were conducted using ClustalX (protein weight matrix: Gonnet series; gap opening: 10.00; gap extension: 0.20; delay divergent sequences: 30%) (Thompson et al., 1997). The resulting tree was constructed by the neighbor joining algorithm with a bootstrap sample of 1000 and visualized by Treecon 1.3b (Van de Peer and De Wachter, 1994).

RNA extraction from leaf material and RNA hybridization analysis

Whole leaf RNA was isolated from the same frozen, homogenized leaf material of *O. vulgare* that had been used for terpene extractions. The probe for CYP71D178 was amplified as ssDNA from a short fragment of 360 bp (bp 1079-1438) of its open reading frame. Hybridization and washing conditions were identical to those previously described (Crocoll et al., 2010). Washed blots were sealed in PVC bags, exposed to BioMax MS films (Kodak, Carestream Health, Rochester, NY, USA) with an intensifying screen for 15 h at -80 °C. Films were developed in a film developer (Konica medical film developer SRX-101A, Konica, Tokyo, Japan) according to the manufacturer's instructions.

RNA isolation from leaf material for qRT-PCR

Total RNA was extracted from homogenized leaf material harvested in November (*O. vulgare* plant lines d2, f5, ff4, ff7, ff8, df6) and April (thyme cv. 'Tc', oregano cv. 'Ct' and marjoram cv. 'gT'). The RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). To remove residual genomic DNA, RNA was treated with RNAse-free DNAse I (Qiagen, Hilden Germany). RNA samples were analyzed on an Agilent Bioanalyzer 2100 and RNA 6000 Nano Labchip using the Expert software (Agilent version B.02.02.SI258) to determine quality, integrity and rRNA ratios. RNA was quantified by spectrophotometry.

For cDNA synthesis, Superscript III reverse polymerase (Invitrogen, Carlsbad, California, USA) was used according to the manufacturer's instructions but with reverse transcription of 3-5 µg total RNA in a 2 x scaled-up reaction volume.

Transcript quantification by relative and absolute qRT-PCR

All experiments were performed on a Stratagene Mx3000P (La Jolla, California, USA) using SYBR green I with ROX as an internal loading standard. Each 25 µl reaction contained cDNA corresponding to 2.5 ng total RNA. Controls included non-RT controls (using 2.5 ng total RNA without reverse transcription to monitor for genomic DNA contamination) and non-template controls (water as template).

PCR thermocycles were run as follows: an initial 10 min at 95 °C followed by 40 cycles of 30 s at 95 °C, 1 min at 57 °C (*Ovtps*2) or 60 °C (all P450s), and 1 min at 72 °C. Fluorescence was read following each annealing and extension phase. All runs were followed by a melting curve analysis from 55-95 °C. The products of each primer pair were cloned and sequenced at least eight times to verify primer specificity. The linear range of template concentration to threshold cycle value (C_t value) was determined by performing a series of sixfold dilutions (1- to 1,296-fold) using cDNA from three independent RNA extractions analyzed in three technical replicates. Cytochrome P450 primer pairs were tested with the other P450s as potentially cross-hybridizing templates. Amplification efficiencies were identical for CYP71D179/182 primers with CYP71D178, CYP71D179 and CYP71D182 as templates. All primers were designed using BeaconDesigner (version 5.0; PremierBiosoft, Palo Alto, California, USA) and HPLC-purified (Invitrogen, Carlsbad, California, USA). Primers were designed for regions identical in all known alleles for the respective gene (for primer details see Supplementary Material, Table S6). Primer efficiencies for all primer pairs were calculated using the standard curve method (Pfaffl, 2001). The stability of reference gene expression in the different oregano plant lines was tested by comparing C_t values between all lines with cDNA corresponding to 2.5 ng total RNA. All amplification plots were analyzed with the MX3000P^{tm} software to obtain C_t values. For relative

qRT-PCR, elongation factor 1 alpha (*OvEF1alpha*) was employed as a housekeeping gene. Relative transcript values were calculated using plant line f5 as calibrator.

Absolute quantification of *Ovtps*2 and cytochrome P450 gene copy numbers in each cDNA sample was conducted using a standard curve. Results were normalized against 1 mg fresh weight of plant material. The standard curves were generated with purified plasmid containing the respective genes. A 10-fold dilution series from 10^1 to 10^6 copies was determined for each template. The copy numbers found in non-RT controls (RNA without reverse transcription to monitor for genomic contamination) ranged from zero to a maximum of 33 copies per µg RNA and mg fresh plant material.

Heterologous expression in *Saccharomyces cerevisiae*

Open reading frames of four cytochrome P450 sequences (*CYP71D178* from cv. 'd06-01' and 'f02-04', *CYP71D180*v1-T28, *CYP71D179*v1-L48 and *CYP71D181*) were cloned into the pESC-Leu2d vector (Ro et al., 2008) and transformed into the *S. cerevisiae* strains WAT11 and W(R) (Pompon et al., 1996) by chemical transformation as described in (Gietz and Woods, 2002). The nucleotide sequence of CYP71D181 was optimized for yeast codon usage and cloned into pESC-Leu2d by Geneart (Regensburg, Germany). pESC-Leu2d was chosen since it harbors a modified Leu promoter region resulting in higher plasmid copy numbers in yeast and therefore higher protein production. The *S. cerevisiae* strain WAT11 also expresses the *A. thaliana* cytochrome P450 reductase *ATR1* (At4g24520) and W(R) expresses high levels of the endogenous yeast P450 reductase. Single colonies were used to inoculate 30 ml overnight cultures in SC minimal medium at 28 °C and 160 rpm. SC minimal medium (without leucine): 6.7 g l^{-1} yeast nitrogen base (without amino acids with ammonium sulfate) (Sigma-Aldrich Chemicals, Steinheim, Germany), 100 mg l^{-1} of the amino acids adenine, arginine, cysteine, lysine, threonine, tryptophan, uracil; 50 mg l^{-1} of the amino acids aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, valine; 20 g l^{-1} D-glucose. The following day, 100 ml of YPDA full medium (10 g l^{-1} yeast extract, 10 g l^{-1} bactopeptone (BD, Le Pont de Claix, France), 74 mg l^{-1} adenine hemisulfate, 20 g l^{-1} D-glucose) was inoculated with overnight culture corresponding to one unit of OD_{600} (approx. 2 x 10^7 cells ml^{-1}) and the culture grown for 32 to 35 h at 28 °C and 160 rpm shaking. The cultures were then centrifuged for 5 min at 5000 x *g* at 16 °C, resuspended in 100 ml YPGA induction medium (10 g l^{-1} yeast extract, 10 g l^{-1} bactopeptone, 74 mg l^{-1} adenine hemisulfate, 20 g l^{-1} D-galactose) and grown for another 15-18 h at 25 °C and 160 rpm.

Preparation of microsomal protein

The extraction of microsomal cytochrome P450 was performed similar to that previously described (Urban et al., 1994). In brief, induced cultures were harvested by centrifugation at 7,500 x *g* for 10 min at 4 °C. The supernatant was decanted and the pellet resuspended in 30 ml of TEK buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM KCl). The suspension was again centrifuged for 10 min at 7,500 x *g* at 4 °C. The supernatant was discarded and the pellet carefully resuspended in 1 ml of TES buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 600 mM sorbitol, 10 g l^{-1} Bovine serum fraction V protein and 1.5 mM β-mercaptoethanol). The solution was transferred to a 50 ml conical tube, and the centrifugation tube washed with another 1 ml TES buffer which was added to the resuspension. Glass beads (0.45-0.50 mm diameter, Sigma-Aldrich Chemicals, Steinheim, Germany) were then added until their volume was up to the top of the cell suspension. Yeast cell walls were disrupted by 1 min shaking by hand, 5 x, with 1 min breaks and cooling on ice in between. The crude extract was recovered and glass beads were washed 4 times with 5 ml TES. The supernatant was pooled and centrifuged at 7,500 x *g* for 10 min at 4 °C to pellet the larger cell fragments. The supernatant was carefully recovered and centrifuged at 100,000 x *g* for 60 min at 4 °C in an ultracentrifuge. The microsomal protein fraction collected in the pellet was washed once with 2 ml TES and once with TEG (50 mM Tris-HCl, 1 mM EDTA, 30 % w/v glycerol). Finally, the pellet was resuspended in 2 ml TEG buffer and completely homogenized in a glass homogenizer (Potter-Elvehjem, Fisher Scientific, Schwerte, Germany). Aliquots were stored at -20 °C and used for protein assays, CO difference spectra measurements and quantification of total protein content. Amounts of cytochrome P450 in microsomal preparations were calculated from the reduced carbon monoxide difference spectra (A_{450} minus A_{490}) by using the differential absorption coefficient of 91 mM^{-1} (Omura and Sato, 1964).

Cytochrome P450 enzyme assays

Standard assays were conducted in 1.5 ml GC vials with 0.3 to 3 µg protein per 300 µl assay volume, 75 mM sodium phosphate buffer, pH 6.8, and 100 µM substrate (supplied in a maximum of 6 µl DMSO which had no detectable effect on the reaction). The reaction was initiated by the addition of 1mM NADPH. The assay mixture was incubated for 20 min at 28 °C with constant shaking on a thermo mixer (Eppendorf, Hamburg, Germany). Reaction was linear for 1 h under these conditions. Supplying NADPH at 1 mM at the beginning of the reaction was found to be sufficient to support catalysis without addition of a regeneration system. The reaction was stopped by freezing at -80 °C, and products were extracted with 200 µl pentane (containing 10 ng μl^{-1} nonyl acetate as internal standard) by 5 min shaking at 1400 rpm. Assay-pentane mixtures were then centrifuged for 10 min at 4,200 rpm and frozen at -80 °C for at least

2 h. The solvent layer was recovered and analyzed by GC-MS and GC-FID as described under GC-MS analysis of volatiles. Monoterpene substrates tested were: *γ*-terpinene, *p*-cymene, *α*-terpinene, (-)-*R*-*α*-phellandrene, (+)-*R*-limonene, (-)-*S*-limonene. Boiled controls and controls without substrate, without cofactor or with substrate and cofactor at time point zero were included in the experiments. The substrates, *γ*- and *α*-terpinene and (-)-*R*-*α*-phellandrene were found to be contaminated with small amounts of *p*-cymene. Spontaneous aromatization is known to occur with *γ*-terpinene to form *p*-cymene (Granger et al., 1964). This is also likely to happen with the reactive substrates, *α*-terpinene and (-)-*R*-*α*-phellandrene. To accurately measure enzymatic formation of *p*-cymene, assays were performed without microsomal protein and the amounts of *p*-cymene detected were subtracted from the assay results with protein. The need for oxygen was tested by treating the reaction mixture (without protein and NADPH) with argon for 20 min. Protein and NADPH were added afterwards with a syringe through the septum of the lid. pH optima were determined by testing with sodium phosphate as the standard buffer (gave higher activity than potassium phosphate or Tris-HCl) over a range from 5.5 to 8.5.

3.4 Results

CYP71D178 expression is correlated with thymol and carvacrol biosynthesis in oregano

The formation of thymol and carvacrol is thought to involve hydroxylation of γ-terpinene and *p*-cymene precursors. We searched an *Origanum vulgare* L. (oregano) cDNA library for sequences similar to those of known cytochrome P450 monoterpene hydroxylases. Sequencing of 2,364 random clones of a library of isolated peltate glandular trichome clusters from *O. vulgare* L. cultivar 'f02-04' resulted in a contig of 504 bp made up of three ESTs which was similar to the limonene-6-hydroxylase previously isolated from mint (Lupien et al., 1999).

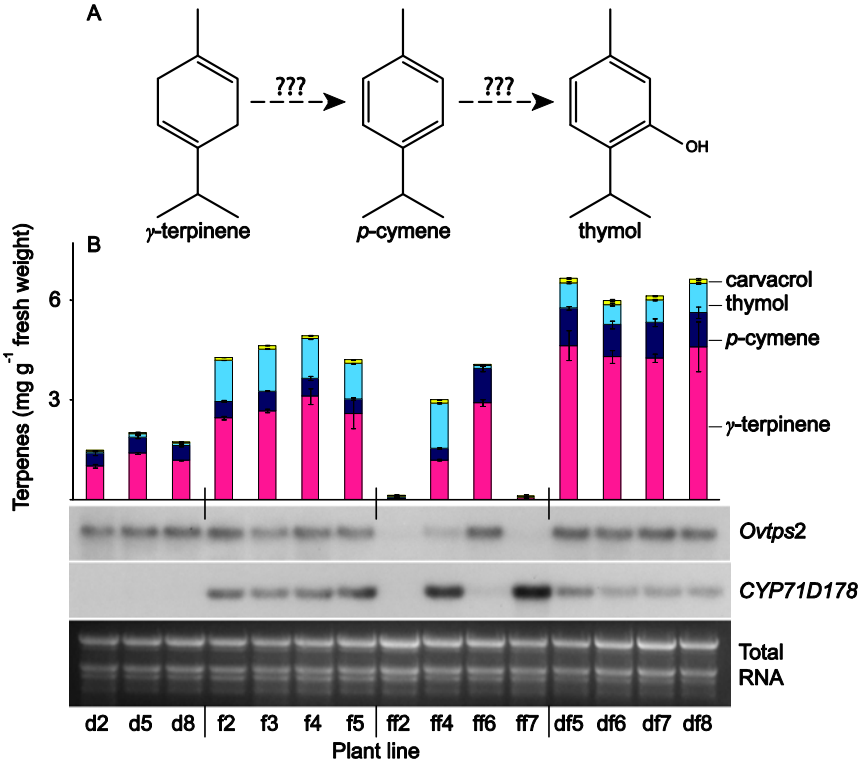


Figure 1 Proposed pathway to thymol and expression analysis of *Ovtps*2 and *CYP71D178 in planta* compared to essential oil contents of various oregano plant lines. (**A**) Possible pathway of thymol formation in thyme predicted by (Poulose and Croteau, 1978a). (**B**) The amounts for γ-terpinene, *p*-cymene, thymol and carvacrol are shown for 15 oregano lines: three clonal lines of the cultivar d06-01 (d2, d5, d8), four clonal lines of cultivar f02-04 (f2, f3, f4, f5), four lines from a selfing of the cultivar f02-04 (ff2, ff4, ff6, ff7), and four lines from a cross of both cultivars (df5, df6, df7, df8). Transcript levels of the terpene synthase *Ovtps*2 and *CYP71D178* in leaves of *O. vulgare* were measured by RNA hybridization analyses. The bottom panel shows an ethidium-bromide-stained agarose gel with total RNA as control for equal RNA loading.

This cytochrome P450-like gene fragment was used to design a probe for RNA hybridization to check whether expression of the full-length gene is correlated with thymol formation in an array of 15 oregano plant lines. The plant lines were derived from two oregano cultivars as described

in (Crocoll et al., 2010). Three of the lines, d2, d5, and d8, were clones of the cultivar d06-01 while the lines f2, f3, f4, and f5 were clones of the cultivar f02-04. In addition to these parental clones, four other lines were generated by selfing of cultivar f02-04 (ff2, ff4, ff6, and ff7) and four further lines from a cross between both parent cultivars (df5, df6, df7, and df8). When the results of the RNA hybridization blot were compared to the terpene content of the oregano plant lines, a strong correlation between transcript abundance and thymol content was evident (Fig. 1b). Thymol and to a lesser extent carvacrol were only found in the essential oil of plant lines which had transcripts for both this putative cytochrome P450 gene and the *γ*-terpinene synthase *Ovtps*2 which forms the substrate for the P450 enzyme. Plant line ff7 showed high transcript abundance of the P450 gene but only traces of thymol and carvacrol. However, this plant line almost completely lacked the predicted precursor *γ*-terpinene; hence no thymol or carvacrol could be formed (Fig. 1b).

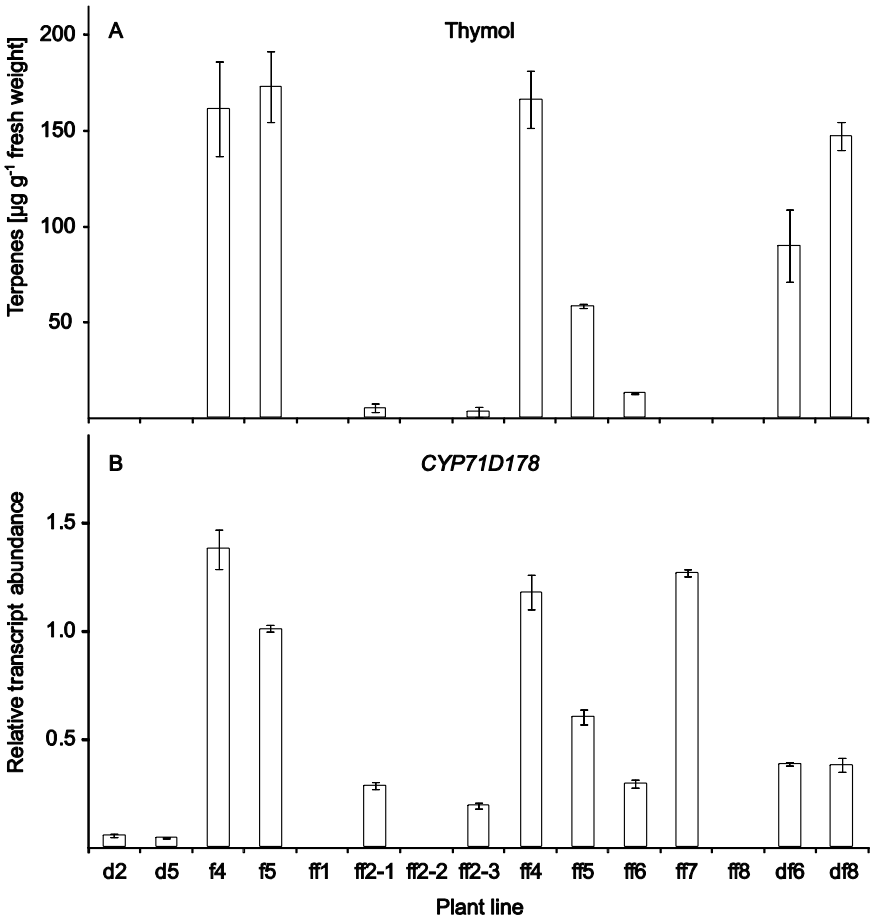


Figure 2 Correlation between (**A**) thymol content in selected *O. vulgare* lines and (**B**) relative transcript levels of *CYP71D178* determined by qRT-PCR. Each bar represents mean values ±SE of three biological and three technical replicates except for individual plant lines ff2-1 to ff2-3 (n = 3). Plant line f5 was used as calibrator for all relative qRT-PCR experiments.

The results from the RNA hybridizations were verified by relative qRT-PCR performed on 13 of the same oregano plant lines (d2, d5, f4, f5, ff1, ff2, ff4, ff6, ff7, ff8, df6, df8). The same

pattern was found as observed for the RNA hybridization analysis (Fig. 2). Only those lines containing transcript of the P450 gene had thymol in their essential oil except for plant line ff7. Despite the high transcript levels in ff7, neither thymol and carvacrol nor *p*-cymene could be detected. *γ*-Terpinene was found in trace amounts only. Spearman's correlation coefficient (r_s) was calculated for the relationship of thymol content to transcript abundance for the P450 gene both with and without plant line ff7. Both calculations gave a positive and significant correlation of $r_s = 0.69$ ($P < 0.001$) (with line ff7) and of $r_s = 0.91$ ($P < 0.001$) (without line ff7).
The complete open reading frame for this putative cytochrome P450 from oregano was isolated from cDNA libraries of cultivars 'f02-04' and 'd06-01' by RACE-PCR. The nucleotide sequences were identical from both cultivars, and the gene was named *CYP71D178* according to the P450 nomenclature (Nelson et al., 1996; Nelson, 2009).

A group of ten cytochrome P450 sequences with high identity to *CYP71D178* was isolated from thyme, oregano and marjoram

A search for similar sequences in oregano led to the isolation of another cytochrome P450 gene from oregano cultivar f02-04. The sequence had high similarity to *CYP71D178* and was named *CYP71D179*. Two additional sequences with similarity were isolated from thyme (*Thymus vulgaris* L.) by colony screening and RACE-PCR. These sequences were kindly provided by Julia Asbach. One gene was annotated as *CYP71D180* and the other included in *CYP71D179*. Amino acid identity between two sequences of more than 97% leads to inclusion under the same gene name, according to the rules of cytochrome P450 nomenclature (Nelson et al., 1996). The thyme sequences were isolated from two different chemotypes of *T. vulgaris*: chemotype T28, which contains thymol and carvacrol in a 1:2 ratio although the larger part of its essential oil consists of *γ*-terpinene and *p*-cymene, and chemotype L48 which is dominated by linalool but contains thymol in amounts comparable to oregano plant line df6 (Julia Asbach, personal communication).
Since *CYP71D178* through *180* had identical 5' and 3' sequences, we designed primers for both ends and used them to isolate additional cytochrome P450 sequences from cDNA of three commercial oregano, thyme and marjoram cultivars. These had been selected for their high diversity in thymol and carvacrol content and named accordingly. The oregano cultivar 'Ct' contains very high amounts of carvacrol (6.3 mg g^{-1} fresh weight) and minor amounts of thymol (0.06 mg g^{-1} fresh weight). In contrast, the thyme cultivar 'Tc' contains mainly thymol (1.6 mg g^{-1} fresh weight) and very little carvacrol (0.05 mg g^{-1} fresh weight). The marjoram cultivar 'gT' is devoid of thymol and carvacrol but contains the putative precursor *γ*-terpinene (0.38 mg g^{-1} fresh weight) and small amounts of the predicted intermediate *p*-cymene (0.05 mg g^{-1} fresh weight) (Supplementary Material, Table S5).

Seven cytochrome P450 gene sequences were isolated from these commercial cultivars, four from thyme cultivar 'Tc', two from oregano 'Ct' and one from marjoram 'gT'. Four of these additional seven sequences were assigned to *CYP71D179* and *CYP71D180* according to the cytochrome P450 nomenclature (Nelson et al., 1996; Nelson, 2009). The other three sequences were assigned to new gene names, two to *CYP71D181* and one to *CYP71D182*. A list of all these cloned P450s and their origin can be found in Table 1.

Table 1 Names of cytochrome P450 genes isolated in the course of this study with their plant and cultivar origins. Reference sequences for comparing the five named cytochrome P450s are underlined. Potential alleles from one plant species are designated with a suffix, v1 or v2, for *CYP71D179* and *CYP71D180*.

Plant species 'cultivar'	CYP name	cultivar gene #
Oregano 'd06-01'	*CY71D178*[a]	d1
Oregano 'f02-04'	*CY71D178*[a]	f1
	CY71D179	f2
Oregano 'Ct'	*CY71D181*	Ct1
	CY71D180	Ct2
Thyme 'T28'	*CY71D180*v1	T28
Thyme 'L48'	*CY71D179*v1	L48
Thyme 'Tc'	*CY71D182*	Tc1
	*CY71D179*v2	Tc2
	*CY71D180*v2	Tc3
	CY71D181	Tc4
Marjoram 'gT'	*CY71D180*	gT1

[a]Both *CYP71D178* nucleotide sequences from oregano cultivars d06-01 and f02-04 are identical.

An amino acid sequence alignment of all clones resulted in four clusters each representing one of the CYP71D sequences, except that CYP71D179 and CYP71D182 were placed together (Fig. 3). Between clusters, sequence identity was less than 80 %. All proteins shared similarity with the limonene-6-hydroxylase, CYP71D18, from mint. All eleven cytochrome P450 sequences, CYP71D178 to CYP71D182, shared amino acid identities of 73 to 76 % with the mint monoterpene hydroxylase while two clearly separated clusters were evident: one containing CYP71D178, CYP71D179 and CPY71D182 and the other CYP71D180 and CYP71D181. Within clusters, sequence identities were 97 % or higher. Therefore, the sequences are designated as the same gene, according to P450 nomenclature standards (Nelson et al., 1996; Nelson, 2009). For a typical 500 amino acid cytochrome P450 protein, 97 % identity translates to less than 15 amino acid differences. Most sequences within each cluster differed by less than 6 amino acids, except the CYP71D182 which has 11 amino acid differences from CYP71D179 and was therefore considered a different gene and named accordingly (David R. Nelson, personal communication). Sequences coming from plants of the

same species but different chemotypes, are designated with a suffix, e.g. v1, v2 in CYP71D179 and CYP71D180 (Table 1).

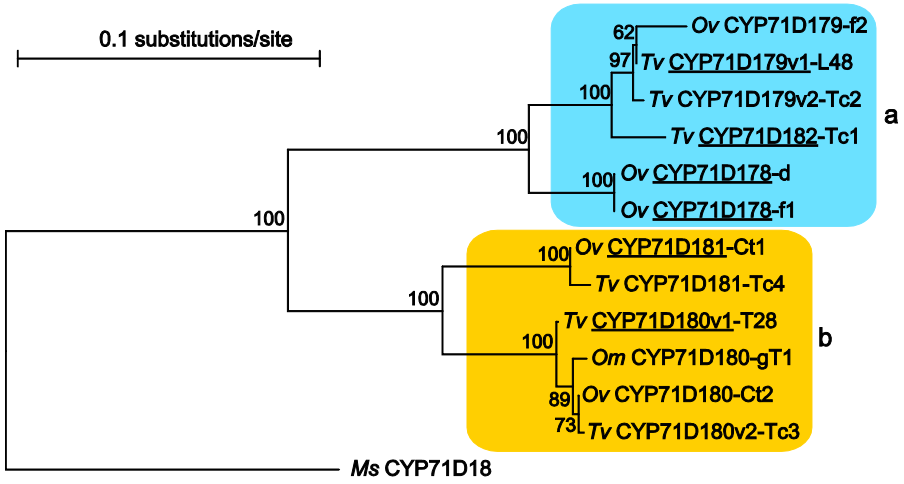


Figure 3 Dendrogram analysis of cytochrome P450s from oregano, thyme and marjoram compared to one member of the same cytochrome P450 subfamily CYP71D from mint, CYP71D18 which encodes a previously characterized monoterpene hydroxylase (Colby et al., 1993; Lupien et al., 1999; Haudenschild et al., 2000). Two subgroups are shown. Subgroup (a) contains CYP71D178, CYP71D179 and CYP71D182. Subgroup (b) consists of CYP71D180 and CYP71D181. The dendrogram was constructed using the neighbour-joining method. The CYP name of each cytochrome P450 is given after the abbreviation of the species of origin and followed by the internal numbering within one cultivar: *Ov* – *Origanum vulgare*, *Tv* – *Thymus vulgaris*, *Om* – *Origanum majorana*; d – oregano cv. d06-01, f –oregano cv. f02-04, Ct – oregano cv. 'Ct', Tc – thyme cv. 'Tc', gT – marjoram cv. 'gT'.

CYP71D178-182 share sequence motifs and substrate recognition sites

For each CYP gene, a single reference sequence (indicated in Table 1 and Fig. 3) was chosen for a more detailed analysis of the amino acid sequences. Sequences were compared to the limonene-6- and the limonene-3-hydroxylases from mint (CYP71D18 and CYP71D13), and common sequence motifs for cytochrome P450s were identified (Fig. 4). The most conserved P450 motif (PFxxGxRxCxG) represents the heme binding loop and is often considered a 'signature' sequence for P450 proteins (Feyereisen, 2005). This motif is responsible for the characteristic 450 nm absorption of the Fe^{II}-CO complex of cytochrome P450 (Mansuy and Renaud, 1995).

Possible substrate recognition sites (SRS) were deduced from alignments with other P450s with known or modeled SRS (Gotoh, 1992; Rupasinghe et al., 2003) (Fig. 4). Substrate recognition sites SRS1 and SRS4 through SRS6 show considerable sequence conservation and were therefore easy to recognize without structural modeling (Mansuy and Renaud, 1995). On the other hand, SRS2 and SRS3 contain no conserved amino acid residues (Rupasinghe et al., 2003) but their location can be estimated since cytochrome P450s share a high degree of secondary

and tertiary structural homology in which secondary structure elements are found in similar locations.

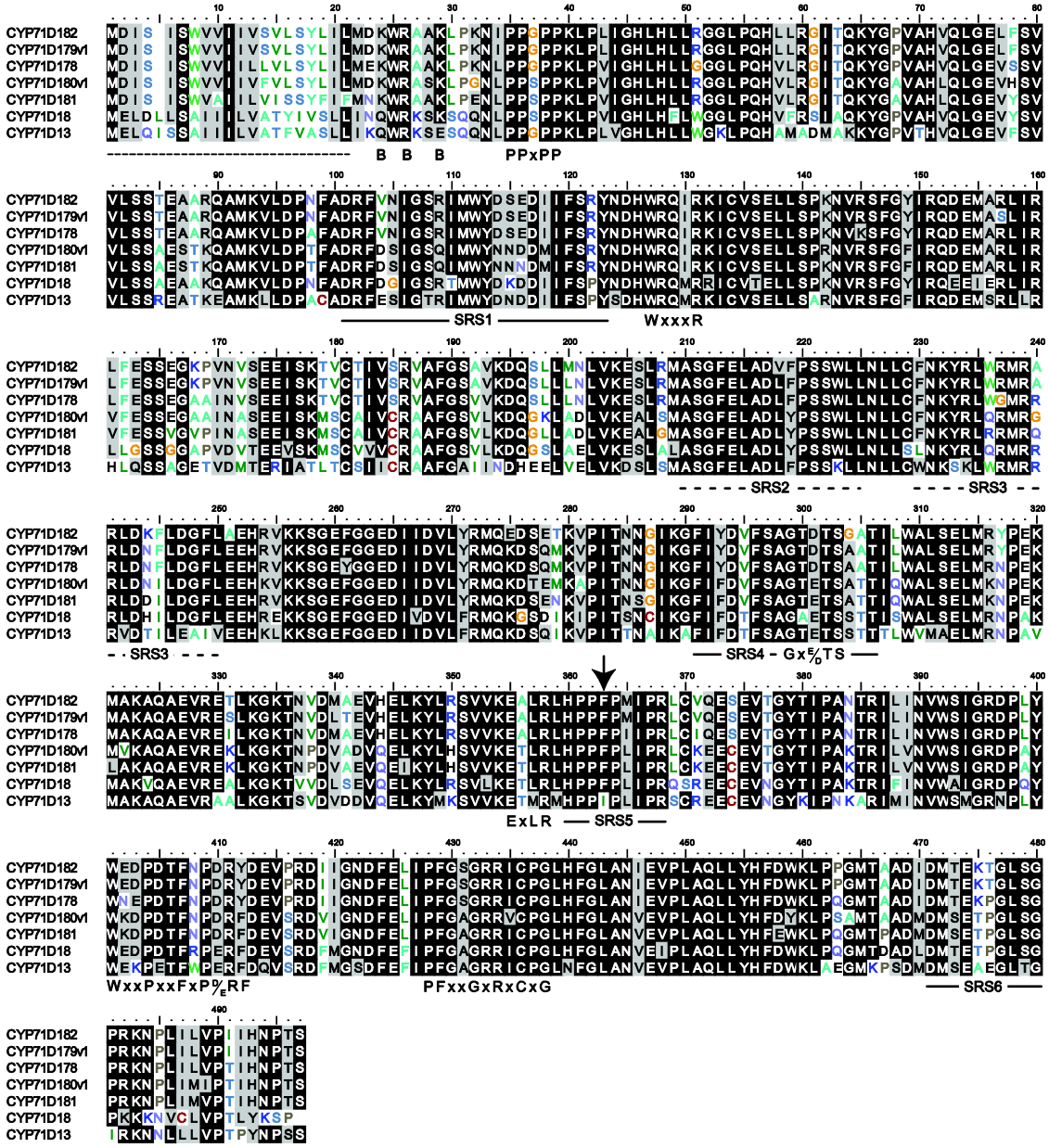


Figure 4 Amino acid alignment of reference sequences for all five named cytochrome P450s, CYP71D178-CYP71D182. Mint CYP71D13 and CYP71D18 are included for comparison. Common sequence motifs of cytochrome P450s are shown: the P450 'signature' sequence PFxxGxRxcxG; WxxxR motif; ExLR motif; proline rich hinge (PPxPP); the membrane anchor is underlined with a dotted line. Putative substrate recognition sites are underlined and named from SRS1 to SRS6. SRS2 and SRS3 are likely found within the markings of the broader dotted regions. The arrow indicates an amino acid residue which is responsible for catalytic differences in CYP71D13 and CYP71D18.

Approximately 50 % of the amino acids are found in α-helices and 15 % in β-sheets (Ravichandran et al., 1993; Halkier, 1996). SRS2 is located at the end of helix F and SRS3 at the beginning of helix G (Gotoh, 1992). The possible locations of SRS2 and SRS3 in our eleven

cytochrome P450 sequences were predicted based on protein secondary structure determined via the SWISS-MODEL workspace (Jones, 1999; Arnold et al., 2006). SRS2 most likely lies in the region between amino acids residues 210-225 and SRS3 between residues 230-250. Among the five cytochrome P450 sequences compared, there are many amino acid similarities in the SRS regions. For example, for CYP71D180 and CYP71D181 all amino acids in SRS4-SRS6 are identical and these are the same as those in the limonene-6-hydroxylase from mint. CYP71D178, CYP71D179 and CPY71D182 also share amino acids at most sites in these regions too, but these are sometimes different from those in CYP71D180 and CYP71D18. CYP71D178 has some unique amino acid substitutions in SRS5 and SRS6 compared to all other sequences, but is identical to CYP71D179 and CYP71D182 in SRS1 and to CYP71D179 in SRS4. The arrow in Figure 4 indicates a phenylalanine residue in SRS5 which is replaced by isoleucine in the mint limonene-3-hydroxylase (CYP71D13). This single amino acid was found to be responsible for the regiospecificity of (-)-*S*-limonene hydroxylation, either at carbon position C6 (CYP71D18) or at C3 (CYP71D13) (Schalk and Croteau, 2000). All cytochrome P450 sequences in the present study have a phenylalanine at this position but differ at a site two amino acids downstream. CYP71D180 and CYP71D181 have a methionine at this position, CYP71D178 an isoleucine and CYP71D179 and CYP71D182 share a leucine together with both mint limonene hydroxylases. Within the five designated P450s, the individual sequences share the same residues in the putative substrate recognition sites as their reference sequences (Supplementary Material, Fig. S2) besides the following exceptions. CYP71D179-f2 differs in a few amino acids in SRS1, SRS4 and SRS5 compared to the reference CYP71D179v1. CYP71D181-Ct1 and Tc4 differ in SRS1 at position 105 where CYP71D181-Tc4 contains a phenylalanine residue in exchange for an isoleucine found in all other P450s of this study. CYP71D180 from marjoram contains a histidine residue at position 122 instead of a tyrosine as in all other P450s.

CYP71D178-182 are differentially transcribed in various oregano, thyme and marjoram lines

To determine the role of the P450 genes in thymol and carvacrol formation, we performed a comparison between terpene content and transcript abundance for *CYP71D178* through *182* using an array of nine oregano plant lines, six used in the initial RNA hybridizations (d2, f5, ff4, ff7, ff8 and df6) and three commercial cultivars: oregano (cv. ‘Ct’), thyme (cv. ‘Tc’) and marjoram (cv. ’gT’). The same plant material was used for RNA and terpene extractions. Absolute qRT-PCR was performed to avoid possible problems caused by variable expression levels of housekeeping genes among different plant species. In previous studies, absolute qRT-PCR was found to produce the same expression pattern as relative qRT-PCR but with the

advantage of absolute transcript numbers to compare between genotypes (Palovaara and Hakman, 2008; Crocoll et al., 2010).

The five cytochrome P450s were represented by a set of four primer pairs designed for the consensus sequences of each of the P450 sequence clusters in Fig. 3. However, due to the high sequence identity among the four clusters, cross binding of primers could not be avoided in all cases. For example, primers designed for *CYP71D178* bound exclusively to *CYP71D178*, but for *CYP71D179* this was not possible. *CYP71D180* primers were designed to bind exclusively to *CYP71D180* sequences, but it seemed likely that *CYP71D181* primers would also amplify *CYP71D180* since the reverse primer was 100 % identical for these sequences (Table 2).

In order to elucidate primer specificity, individual gene fragments amplified by PCR with the P450 primer pairs were sequenced from selected plant lines. All fragments amplified with *CYP71D178*, *CYP71D180* and *CYP71D181* primer pairs belonged to the respective P450 genes. However, PCR with *CYP71D179* primers resulted in considerable cross-reaction depending on the plant line used. In the oregano plant lines d2 to df6, the majority of the fragments amplified with CYP71D179 primers belonged to *CYP71D179* (75 %) and the rest to *CYP71D178* (25 %). In oregano cultivar 'Ct', only *CYP71D179* was amplified; no fragments for a gene similar to *CYP71D182* were found in this cultivar. In thyme cv. 'Tc', the majority of the fragments amplified with *CYP71D179* primers actually belonged to *CYP71D182* (69 %) with lesser amounts of *CYP71D179* (25 %) and *CYP71D178* (6 %). Therefore, no clear resolution of *CYP71D179* and *CYP71D182* expression levels was possible.

Table 2 Specificity of primer pairs used for absolute qRT-PCR with cytochrome P450 genes from oregano, thyme and marjoram. The data are based on *in silico* predictions that were confirmed in the course of the actual analysis, with the exception that *CYP71D181* primers were not found to amplify any *CYP71D180* fragments (listed in parentheses).

	Primer pair			
P450	*CYP71D178*	*CYP71D179* / *182*	*CYP71D180*	*CYP71D181*
CYP71D178	++	++	--	--
CYP71D179	--	++	--	--
CYP71D182	--	++	--	--
CYP71D180	--	--	++	(-+)
CYP71D181	--	--	--	++

++ = binding of both primers, -- = no primer binding, -+ = binding of reverse primer

The results of the qRT-PCR analysis showed that *CYP71D178* and *CYP71D179* / *182* were expressed in the original oregano plant lines d2 to df6, but *CYP71D180* and *181* were not. *CYP71D178* was transcribed in relatively low copy numbers (~26,000 copies in plant line ff4) compared to *CYP71D179* / *182* which showed much higher transcript abundance especially in

thyme cultivar ‘Ct’ (147,000 copies) (Fig. 5). (Complete copy numbers can be found in Table S7, Supplementary material.) *CYP71D180* was transcribed only in the oregano cultivar ‘Ct’ and thyme cultivar ‘Tc’ with 381 in very low copy numbers. *CYP71D181* was expressed exclusively in oregano cultivar ‘Ct’ with 85,000 copies.

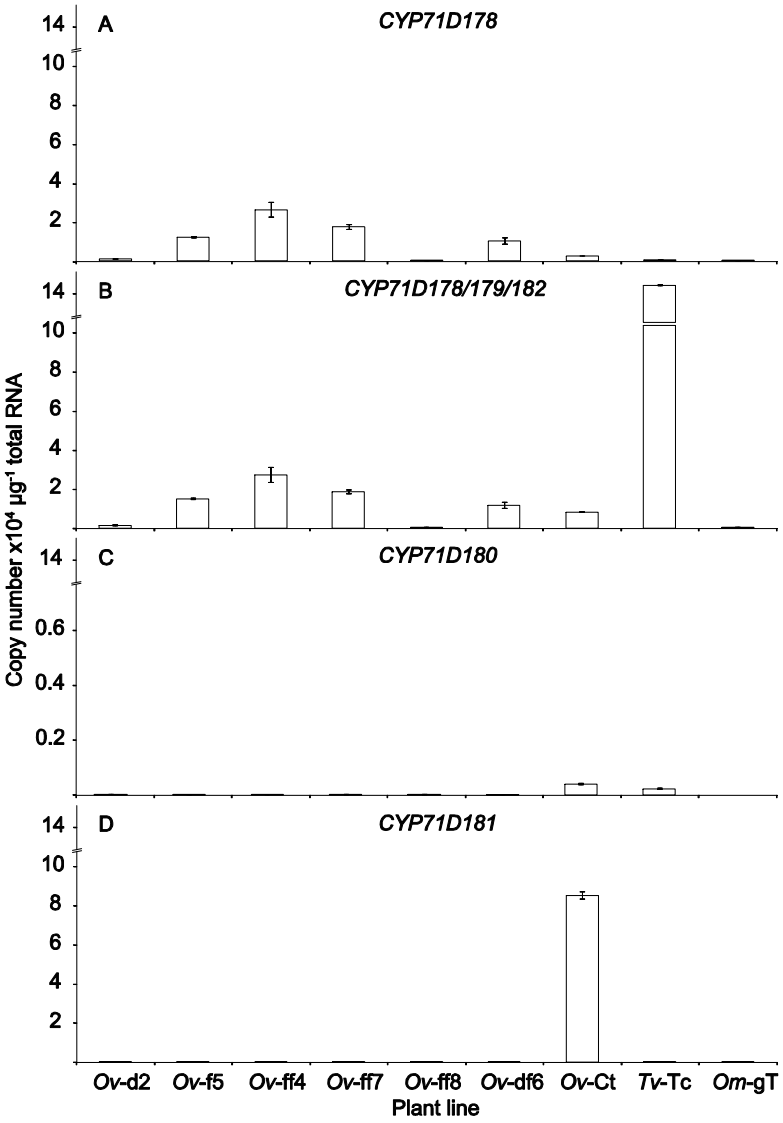


Figure 5 Absolute transcript levels of *CYP71D178* through *CYP71D182* in six *O. vulgare* lines (d2, f5, ff4, ff7, ff8, df6) and three commercial cultivars, oregano cultivar ‘Ct’, thyme cultivar ‘Tc’ and marjoram cultivar ‘gT’. (**A**) Absolute copy numbers of *CYP71D178*. (**B**) Absolute copy numbers for the three P450s *CYP71D178*, *CYP71D179* and *CYP71D182*. All three P450s were amplified with the same efficiency. (**C**) Absolute copy numbers of *CYP71D180*. (**D**) Absolute copy numbers of *CYP71D181*. For all transcripts, copy numbers per µg total RNA were determined by absolute qRT-PCR and normalized per mg fresh plant material. Each bar represents mean values ±SE of three biological and three technical replicates except for oregano cv. ‘Ct’, thyme cv. ‘Tc’ and marjoram cv. ‘gT’ which had only three technical replicates.

Correlation of *CYP71D178-182* transcript levels to thymol and carvacrol con tent suggests specific biosynthetic roles for these genes

CYP71D179 / 182 showed highest transcript abundance in thyme cultivar 'Tc' (Fig. 5) where thymol was found in highest amounts (Fig. 6a), and thus may encode a protein involved in thymol biosynthesis. By contrast, the *CYP71D181* transcript was found in high amounts almost exclusively in oregano cultivar 'Ct' (Fig. 5) where carvacrol was in highest abundance (Fig. 6), and may be involved in carvacrol biosynthesis. *CYP71D180* transcript might also be involved in carvacrol biosynthesis because of its presence in low levels in oregano 'Ct' as well as in thyme 'Tc', which has traces of carvacrol. The oregano plant lines, d2 to df6, had thymol but virtually no detectable carvacrol. However, their thymol levels were relatively low compared to those of the thyme cultivar 'Tc'.

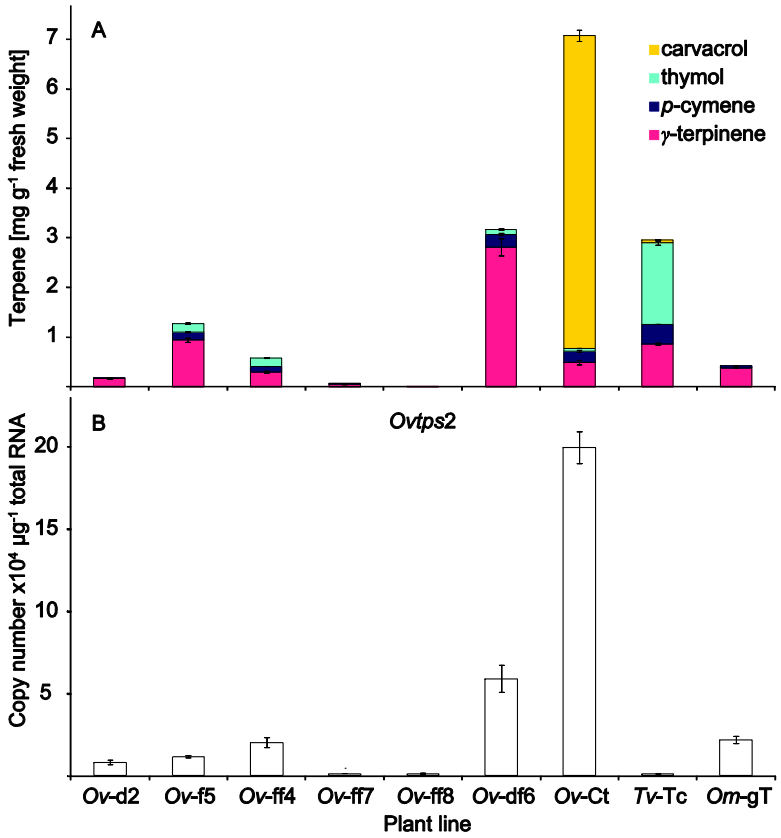


Figure 6 (**A**) Terpene contents of plant lines from Figure 5 and (**B**) Expression data for γ-terpinene synthases from absolute qRT-PCR. Absolute copy numbers for transcript of *Ovtps*2 are shown. Each bar represents mean values ±SE (n = 9 except for oregano cv. 'Ct', thyme cv. 'Tc' and marjoram cv. 'gT' n=3).

Nevertheless, the presence of *CYP71D178* and *CYP71D179 / 182* transcripts in these oregano lines suggests that these genes are also involved in thymol biosynthesis. Virtually none of these transcripts were detected in the lines ff8 or the marjoram cultivar 'gT', plants which both lacked thymol and carvacrol. Surprisingly, line ff7 had high levels of transcript, despite containing only

traces of these phenolic monoterpenes. Co-expression of two or three cytochrome P450s in the same plant could indicate that the pathway to thymol and carvacrol involves more than one cytochrome P450.

In oregano and marjoram, γ-terpinene, is formed by closely related enzymes

*Ovtps*2 has been previously isolated from *Origanum vulgare* and described as a γ-terpinene synthase (Crocoll et al., 2010). To check whether similar genes might be responsible for γ-terpinene formation in the three oregano, thyme and marjoram cultivars, 'Ct', 'Tc' and 'gT', we compared the transcript levels (determined by qRT-PCR with primers designed for *Ovtps*2) with γ-terpinene content in these plant lines. The data showed there is a good association between the presence of transcripts and γ-terpinene content in 7 of 8 lines (Fig. 6). In marjoram, a closely related gene with high sequence identity seems to be present and responsible for γ-terpinene formation. However, the exception is thyme, where in the cultivar 'Tc' no expression of a corresponding gene was found (Fig. 6b) despite an abundance of γ-terpinene. A different gene might be responsible for γ-terpinene formation in thyme. In fact, in a related project, a γ-terpinene synthase was isolated and characterized from thyme chemotype T28 (Julia Asbach, unpublished results). The gene sequence of this *Tvtps*1 is 90 % identical to that of the oregano *Ovtps2*. Recently, we isolated a second γ-terpinene synthase, *Ovtps*8, from oregano cultivars d06-01 and f02-04. The newly designated *Ovtps*8 gene sequence is 99.6 % identical to *Tvtps*1 from thyme and after heterologous expression in *E. coli* gave active enzyme producing γ-terpinene from geranyl diphosphate as substrate (data not shown). Transcript levels for these genes have not yet been determined.

Heterologous expression of *CYP71D178*, *CYP71D180*v1 and *CYP71D181* in *S. cerevisiae* demonstrated the formation of *p*-cymene, thymol and carvacrol from γ-terpinene and α-terpinene

To characterize the enzyme catalytic activities of the identified P450, five of the sequences (*CYP71D178*, *CYP71D179*, *CYP71D179*v1, *CYP71D180*v1 and *CYP71D181*) were expressed in *S. cerevisiae* strains modified to express high levels of the endogenous yeast cytochrome P450 reductase or the P450 reductase 1 from *A. thaliana* (*AtR1*). Three of the genes showed enzyme activity after extraction of microsomal protein from the yeast cultures: *CYP71D178*, *CYP71D180*v1 and *CYP71D181*. Assays were carried out under linear conditions with 100 µM substrate and 1 mM NADPH as cofactor. γ-Terpinene was accepted as substrate by all three proteins to form mostly *p*-cymene, the predicted intermediate in thymol biosynthesis (Poulose and Croteau, 1978a) (Fig. 7a). Small amounts of carvacrol were also detected in CYP71D180v1 and CYP71D181 enzyme assays, and small amounts of both thymol and carvacrol were

detected in CYP71D178 assays, (Fig. 7b). The formation of *p*-cymene may be artifactual since spontaneous conversion of *γ*-terpinene into *p*-cymene is known (Granger et al., 1964) and was observed in assays with microsomal protein from empty vector controls (Fig. 8) as well as with the two other substrates, *α*-terpinene and (-)-*R*-*α*-phellandrene (data not shown).

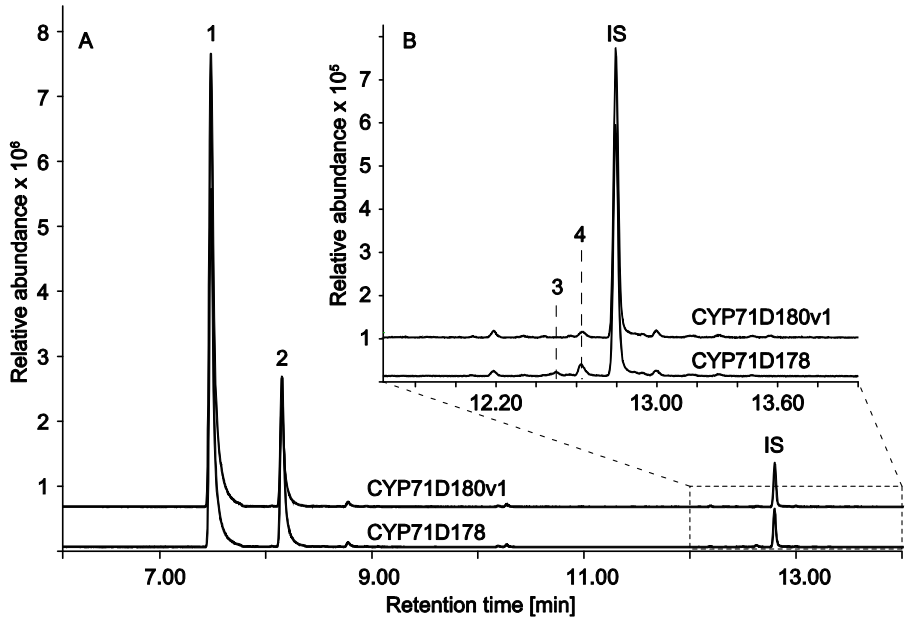


Figure 7 Products of CYP71D178 and CYP71D180v1 measured *in vitro* after incubation with *γ*-terpinene in the presence of 1 mM NADPH. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) Major product of CYP71D178 and CYP71D180v1: 1, *p*-cymene; 2, the substrate *γ*-terpinene. (**B**) Minor products of CYP71D178 and CYP71D180v1: 3, thymol and 4, carvacrol. Nonyl acetate was used as internal standard (IS) for quantification.

When *p*-cymene was offered as a substrate, it was not accepted by any of the three active enzymes (Fig. 9a). However, other cyclohexanoid monoterpene dienes like *α*-terpinene (Fig. 9b) and (-)-*R*-*α*-phellandrene (data not shown), were both converted by CYP71D178 and CYP71D180v1 into *p*-cymene. CYP71D181 converted *α*-terpinene into *p*-cymene and carvacrol (data not shown). *α*-Terpinene may also be an enzyme substrate *in vivo* since this comound was present in amounts ranging from 6 (plant line ff8) to 157 µg g^{-1} fresh weight (plant line df6). (-)-*R*-*α*-Phellandrene was present in the commercial oregano, thyme and marjoram plants. Exact values for all terpenes can be found in Table S5, Supplementary Material. *γ*-Terpinene content ranged from 8 µg g^{-1} fresh weight (plant line ff8) up to 2868 µg g^{-1} fresh weight in plant line df8.

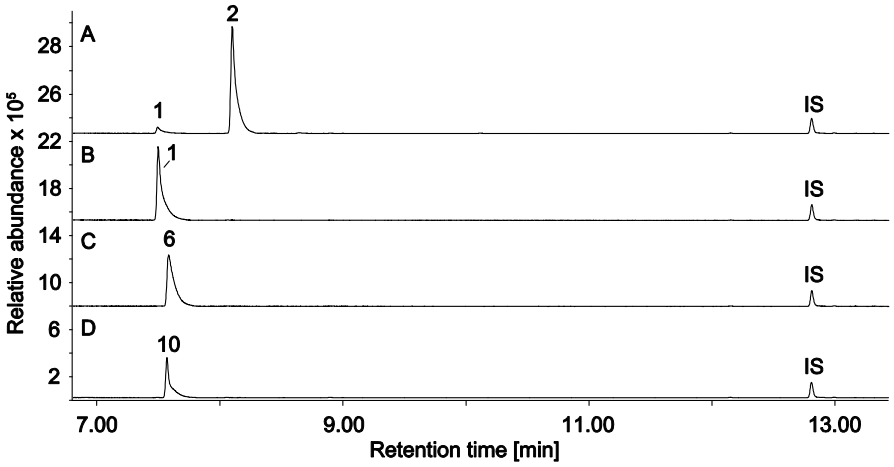


Figure 8 Empty vector control assay results after incubation with substrate in the presence of 1 mM NADPH. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) Low formation of 1, *p*-cymene from 2, *γ*-terpinene. (**B**) No product formation from 1, *p*-cymene. (**C**) No product formation from 6, (-)-*S*-limonene. (**D**) No product formation from 10, (+)-*R*-limonene. Nonyl acetate was used as internal standard (IS) for quantification.

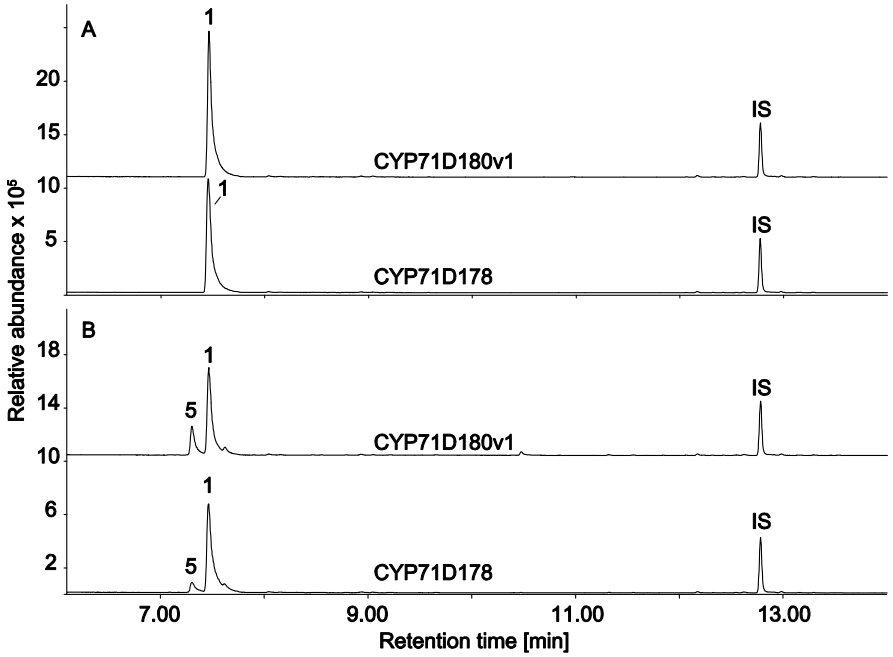


Figure 9 Products of CYP71D178 and CYP71D180v1 measured *in vitro* after incubation with *p*-cymene or *α*-terpinene in the presence of 1 mM NADPH. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) No product formation by CYP71D178 and CYP71D180v1 with *p*-cymene as substrate. (**B**) Products from *α*-terpinene as substrate: 1, *p*-cymene; 5, *α*-terpinene. Nonyl acetate was used as internal standard (IS) for quantification.

Enzymatic properties of CYP71D178 and CYP71D180v1 differ only slightly from those of other P450 monoterpene hydroxylases

The pH optima for cytochrome P450s are usually on the basic side of neutrality (Mihaliak et al., 1993). However, the pH optimum for γ-terpinene conversion by CYP71D178 was determined between pH 6.8 and 7.0 with half maximal activities at 5.8 and 8.5. The pH optimum was identical for CYP71D181 but with a narrower range of half maximal activities at 6.4 and 8.0. For CYP71D180v1, the optimum was more acidic at pH 6.4, but there was minimal activity loss when the pH was raised to 6.8. Activities were drastically reduced at pH 6.

The apparent K_m values for *p*-cymene formation from γ-terpinene were determined to be 37 µM for CYP71D178 and 40 µM for CYP71D180v1. These are in the same range as for the limonene hydroxylases from mint, which have a K_m of 20 µM for the substrate (-)-*S*-limonene (Karp et al., 1990). In the present study, conversion rates were V = 975 ng mg protein^{-1} h^{-1} for CYP71D178 and V = 1658 ng mg protein^{-1} h^{-1} for CYP71D180v1. k_{cat} was determined for CYP71D180v1 as 1.24 s^{-1}, but not for the other proteins as no clear CO-difference spectra could be measured for these to calculate the exact amounts of active P450 enzyme in the microsomal preparations (Table 3).

Table 3 Apparent K_m values and catalytic efficiencies for CYP71D178 and CYP71D180v1. k_{cat} could only be calculated for CYP71D180v1 since no clear CO difference spectra could be measured for CYP71D178 and CYP71D181 to calculate molar protein amounts.

	CYP71D178	CYP71D180v1	
Substrate	K_{mapp} [µM]	K_{mapp} [µM]	k_{cat} [s^{-1}]
γ-terpinene	37.2	40.3	1.24
(+)-*R*-limonene	n.d.	14.1	0.08
(-)-*S*-limonene	n.d.	0.11	0.15

n.d. = not determined

The regiospecificity of hydroxylation of the substrate limonene differs between CYP71D178 and CYP71D180 / CYP71D181

We investigated the ability of the CYP71D178, CYP71D180v1 and CYP71D181 enzymes to utilize limonene, the substrate of the very similar CYP71D13 / 18 hydroxylases characterized from mint species. In mint, (-)-*S*-limonene is either hydroxylated by CYP71D13 in peppermint to the C3-oxidized product, (-)-*trans*-isopiperitenol or in spearmint by CYP71D18 to the C6-oxidized product, (-)-*trans*-carveol. The positions of limonene hydroxylation (C3 vs. C6) are relevant to the present study since thymol is a product of C3 hydroxylation (Fig. 11), while carvacrol is hydroxylated at a position (C2) which corresponds to that of C6 in limonene.

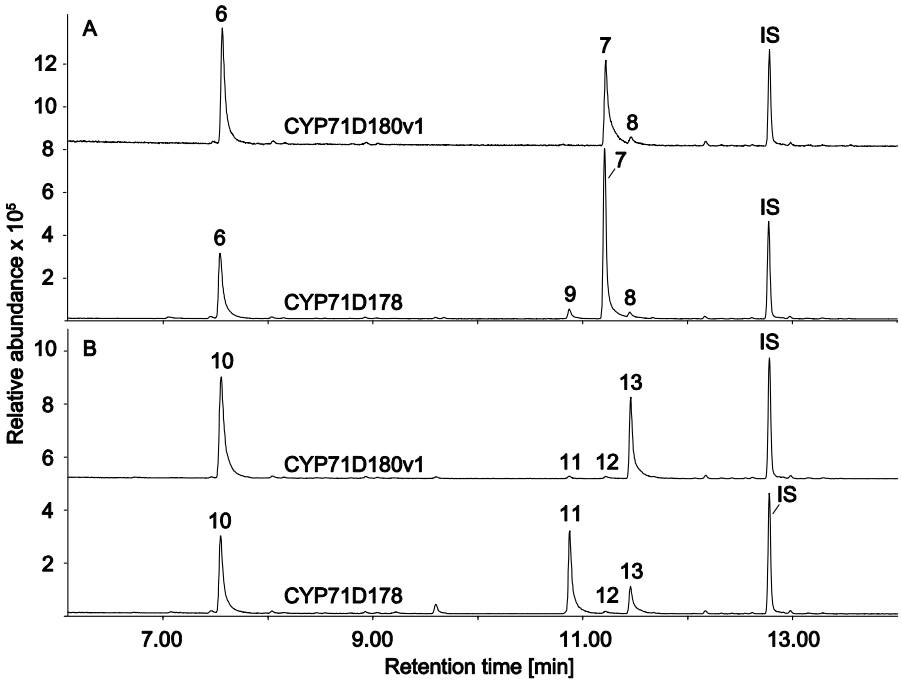


Figure 10 Products of CYP71D178 and CYP71D180v1 measured *in vitro* after incubation with limonene in the presence of 1 mM NADPH. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) Products formed by CYP71D178 and CYP71D180v1 from (-)-*S*-limonene (6): 7, (-)-*trans*-carveol, 8, (-)-*cis*-carveol, 9, (-)-*trans*-isopiperitenol. (**B**) Products formed by CYP71D178 and CYP71D180v1 from (+)-*R*-limonene (10): 11, (+)-*trans*-isopiperitenol, 12, (+)-*trans*-carveol, 13, (+)-*cis*-carveol. Nonyl acetate was used as internal standard (IS) for quantification.

All three enzymes accepted (-)-*S*-limonene as substrate and converted it into (-)-*trans*-carveol, a C6-oxidation product (Figs. 10a and 11). On the other hand, administration of the enantiomeric (+)-*R*-limonene resulted in a more complex pattern of results. (+)-*R*-Limonene was previously shown to be converted by the mint 3-hydroxylase, CYP71D13, to the corresponding 3-oxygenated product, (+)-*trans*-isopiperitenol, while the mint 6-hydroxylase, CYP71D18, converted the substrate to a mixture dominated by the corresponding 6-oxygenated product, (+)-*cis*-carveol (Wüst et al., 2001; Wüst and Croteau, 2002). Interestingly, the P450s in this study did not follow this pattern. (+)-*R*-Limonene was converted to a C3-oxygenated product, (+)-*trans*-isopiperitenol, by CYP71D178, but to a C2(C6)-oxygenated product, (+)-*cis*-carveol by CYP71D180v1 (Fig. 10b) and CYP71D181 (data not shown). With both, limonene and γ-terpinene, the stereospecificity of the enzyme are similar. While CYP71D178 produces both, thymol and carvacrol, (both C3 and C2 oxidation products), CYP71D180v1 and CYP71D181 produce carvacrol only (C2 oxidation product). The enzyme activities correspond well to the transcript patterns and the essential oil composition among different plant lines (Figs. 5, 6). CYP71D178 transcript was correlated with thymol (C3) content while CYP71D180 and CYP71D181 transcripts were correlated with carvacrol (C2) content. Although these enzymes

are able to use the substrate limonene, no mint cytochrome P450 has been reported to use *p*-cymene, *γ*-terpinene, *α*-phellandrene or *α*-terpinene as substrates (Karp et al., 1990). The products formed from all monoterpene substrates tested are listed in Table 4.

The apparent K_m values of CYP71D180v1 for the enantiomeric limonenes were lower than for *γ*-terpinene. For (-)-*S*-limonene, the apparent K_m value was only 0.11 µM, but the velocity of the reaction was 4.5-fold lower compared to *p*-cymene formation from *γ*-terpinene (V = 365 ng mg^{-1} protein h^{-1}). For (+)-*R*-limonene apparent K_m was 14 µM with an 8-fold slower conversion rate (V = 205 ng mg^{-1} protein h^{-1}). k_{cat} values for limonene substrates were 15-fold and 9-fold lower than for *γ*-terpinene (Table 3). The limonenes are unlikely to be substrates for CYP71D178 and CYP71D180v1 *in vivo* because limonene occurs only in very low amounts as the (+)-*R*-enantiomer (6 to 60 µg g^{-1} fresh weight) in the oregano plant lines studied. No limonene was found in the commercial oregano, thyme or marjoram cultivars. And none of the possible limonene metabolites, including carveol and isopiperitenol, could be identified in any of the investigated plants (Table S5, Supplementary Material).

Table 4 Major *in vitro* products formed from monoterpene substrates by CYP71D178, CYP71D180v1 and CYP71D181 from oregano and thyme compared to CYP71D13 and CYP71D18 from mint. Chemical structures can be found in Figure 11.

	γ-terpinene	*α*-terpinene	(+)-*R*-limonene	(-)-*S*-limonene	(-)-*R*-*α*-phellandrene
CYP71D178	*p*-cymene, thymol, carvacrol	*p*-cymene	(+)-*trans*-isopiperitenol, (+)-*cis*-carveol	(-)-*trans*-carveol, (-)-*trans*-isopiperitenol	*p*-cymene
CYP71D180v1	*p*-cymene, carvacrol	*p*-cymene	(+)-*cis*-carveol	(-)-*trans*-carveol	*p*-cymene
CYP71D181	*p*-cymene, carvacrol	*p*-cymene, carvacrol	(+)-*cis*-carveol	(-)-*trans*-carveol	*p*-cymene
CYP71D13[a]	nc	nc	(+)-*trans*-isopiperitenol	(-)-*trans*-isopiperitenol	nc
CYP71D18[a]	nc	nc	(+)-*cis*-carveol (-)-*trans*-isopiperitenol	(-)-*trans*-carveol	nc

nc = no conversion, [a] Data from (Karp et al., 1990; Wüst et al., 2001; Wüst and Croteau, 2002)

3.5 Discussion

Thymol and carvacrol formation is widespread throughout the plant kingdom

The pathway of thymol formation has been studied since the early 1960s when Yamazaki and coworkers identified thymol as a terpenoid biosynthetic product despite the fact that it is aromatic (Yamazaki et al., 1963). At the end of the 1970s, experiments were performed in which the radioactively labeled monoterpenes, *γ*-terpinene and *p*-cymene, were fed to thyme (Poulose and Croteau, 1978a). Based on the results, it was postulated that the biosynthesis of thymol starts with *γ*-terpinene as initial monoterpene substrate and proceeds via the aromatic *p*-cymene as an intermediate (Fig. 1a). Thymol and its chemical relative carvacrol are found not only in thyme and oregano, but also in many other plant species in different families, e.g. in horsemint or bee balm (*Monarda* sp.), savory (*Satureja* sp.) and *Thymbra* sp. (all from the Lamiaceae); in ajwain or bishop's weed (*Trachyspermum ammi* [L.] Sprague, Apiaceae), and in the so-called Mexican oregano (*Lippia* sp., Verbenaceae) (Craveiro et al., 1981; Matos et al., 2000; Catalan, 2002; Gwinn et al., 2010; Stashenko et al., 2010). Most species which produce thymol and / or carvacrol also contain *γ*-terpinene and *p*-cymene. The co-occurrence of *γ*-terpinene and *p*-cymene with either one or both phenolic monoterpenes hints towards a common mechanism for thymol and carvacrol biosynthesis which might have evolved independently several times in the plant kingdom.

Thymol and carvacrol formatio n depends on the presence of *γ*-terpinene in oregano, marjoram and thyme

The *γ*-terpinene synthase *Ov*TPS2 was found to be the enzyme responsible for *γ*-terpinene biosynthesis in oregano (Crocoll et al., 2010). In the closely related marjoram, an ortholog of *Ovtps*2 seems to be responsible for *γ*-terpinene formation (Fig. 5). In thyme, the monoterpene synthase, *Tv*TPS1, was identified as the major *γ*-terpinene synthase (Julia Asbach, unpublished data). *Tvtps*1 is most closely related to *Ovtps*8 in oregano (99.5 % sequence identity on both, nucleotide and amino acid level) but only 90 % identical to *Ovtps*2. Transcript levels for *Tvtps*1 have been measured in different thyme chemotypes and were found to strongly correlate with *γ*-terpinene content (Julia Asbach, personal communication). Whether or not *Ovtps*8 has any importance in *γ*-terpinene formation in oregano needs to be tested in the future. Since no ESTs with similarity to *Ovtps*8 were found in a cDNA library made from oregano cultivar f02-04 (Crocoll et al., 2010), it might not be actively involved in *γ*-terpinene biosynthesis.

The identified cytochrome P450s are involved in the production of thymol or carvacrol

The gene expression pattern of *CYP71D178* in oregano was well-correlated with thymol content in different sets of oregano lines (Figs. 1b and 2). The only exception was oregano line ff7 which showed a high transcript accumulation but only traces of thymol and carvacrol. However, this line lacked *γ*-terpinene which is a precursor of phenolic monoterpene formation.

Nevertheless, it remained unclear whether only one cytochrome P450 enzyme is sufficient for the formation of thymol or carvacrol since this conversion involves two formal oxidation steps. Hence, additional P450 sequences were sought from oregano, thyme and marjoram cultivars using primers from *CYP71D178*. Eleven sequences were found classified into five different P450s that gave a clear differentiation into two subgroups a) and b) (Fig. 3).

The two subgroups of P450s could conceivably represent the two different catalytic activities, one for thymol and one for carvacrol. This hypothesis was tested by absolute qRT-PCR on a group of oregano plant lines as well as three oregano, thyme and marjoram cultivars with varying thymol and carvacrol content. Although qRT-PCR was partly hampered by the high nucleotide sequence identity between the different P450s, a clear differentiation of expression pattern was observed for three of the five cytochrome genes (Fig. 5). The high copy numbers of *CYP71D181* correlated with the very high carvacrol content in oregano cultivar 'Ct', and the expression levels of *CYP71D179* and *CYP71D182* correlated with thymol content. Plants in which transcripts of any of the P450s could not be detected produced no thymol and carvacrol. These correlations suggested that expression of these CYP71D P450s is regulated at the transcript level in these plants.

Other studies have reported on the roles of cytochrome P450s in monoterpene metabolism. Geraniol is hydroxylated by a geraniol-10-hydroxylase in *Catharanthus roseus* (Meijer et al., 1993), and sabinene hydroxylase catalyzes the formation to sabinol in sage (Karp et al., 1987). Several examples are known of limonene hydroxylases which catalyze hydroxylations at different carbon atoms in this cyclic monoterpene. In spearmint, a 6-hydroxylase forms (-)-*trans*-carveol from (-)-*S*-limonene, whereas in peppermint this substrate is hydroxylated by a 3-hydroxylase to form (-)-*trans*-isopiperitenol, an intermediate in the menthol biosynthesis pathway (Karp et al., 1990; Lupien et al., 1995). Recently, a cytochrome P450 candidate for the hydroxylation of (-)-*S*-limonene was reported for *Perilla* which is able to catalyze the hydroxylation at three different positions to either form (-)-*trans*-isopiperitenol, (-)-*trans*-carveol or (-)-perillyl alcohol (Mau et al., 2010). In caraway, another limonene-6-hydroxylase catalyzes the reaction of the other enantiomer, (+)-*R*-limonene, into (+)-*trans*-carveol (Bouwmeester et al., 1998). Another important P450 in monoterpene modification is the

menthofuran synthase from mint which uses as a substrate pulegone, an intermediate in the menthol biosynthesis pathway (Bertea et al., 2001).

CYP71D178, CYP71D180v1 and CYP71D181 form thymol or carvacrol and *p*-cymene from *γ*-terpinene

Although five P450s were expressed in yeast, microsomal preparations gave enzymatically active protein for only three enzmyes, CYP71D178, CYP71D180v1 and CYP71D181. Despite the high substrate specificity reported for plant biosynthetic P450s (Schuler, 1996) all three enzymes accepted a variety of different monoterpenes as substrates, including *γ*-terpinene, *α*-terpinene, (-)-*R*-*α*-phellandrene, (+)-*R*-limonene and (-)-*S*-limonene, which are all cyclo-hexanoid monoterpenes with two double bonds (Fig. 11). Given the proposed role of *γ*-terpinene in the pathway to thymol and carvacrol, *γ*-terpinene was expected to be the natural substrate. All three active P450 enzymes converted *γ*-terpinene to one of these phenolic monoterpenes in small amounts: CYP71D180 and CYP71D181 formed carvacrol and CYP71D178 formed both thymol and carvacrol. But the main product in all cases was *p*-cymene. This aromatic monoterpene was suggested to be an intermediate of thymol biosynthesis, between *γ*-terpinene and thymol (Fig. 1a) (Poulose and Croteau, 1978a). However none of the enzymes converted *p*-cymene to thymol, carvacrol or any other product. The original pathway prediction suggested two separate steps of oxidation, one from *γ*-terpinene to the aromatic *p*-cymene and the second from *p*-cymene to the phenol, thymol (or carvacrol) (Fig. 1a). Both could be catalyzed by P450-type enzymes. The hydroxylation of aromatic rings in the second step is a widespread reaction for P450s which typically follows the so-called NIH shift mechanism (Jerina and Daly, 1974). If there are two separate P450-catalyzed steps and the enzymes investigated here are assumed to make *p*-cymene *in planta*, perhaps one of the not yet characterized CYP71D P450s is responsible for the second step from *p*-cymene to thymol or carvacrol. The co-expression of two or more CYP71D P450s in most of the plant lines studied (Fig. 6) supports such a two-enzyme scenario.

On the other hand, both oxidations from *γ*-terpinene to thymol (or carvacrol) could be catalyzed by a single P450 without release of the intermediate. It would also be possible that *p*-cymene cannot access the active site to act as an intermediate. Cytochrome P450s are known to catalyze multiple oxidations on a single substrate (Halkier et al., 1995; Bak et al., 1998; Ro et al., 2005). In such a scenario, *p*-cymene might be a bound intermediate or simply an *in vitro* artifact. The formation of *p*-cymene from *γ*-terpinene is known to occur non-enzymatically in the presence of oxygen (Granger et al., 1964) and might be enhanced by binding to P450s. *p*-Cymene is also a major product formed from the substrates *α*-terpinene and (-)-*R*-*α*-phellandrene. These

monoterpenes were found to form *p*-cymene spontaneously even more rapidly than γ-terpinene which might be due to the fact that the double bonds are conjugated (Fig. 11).

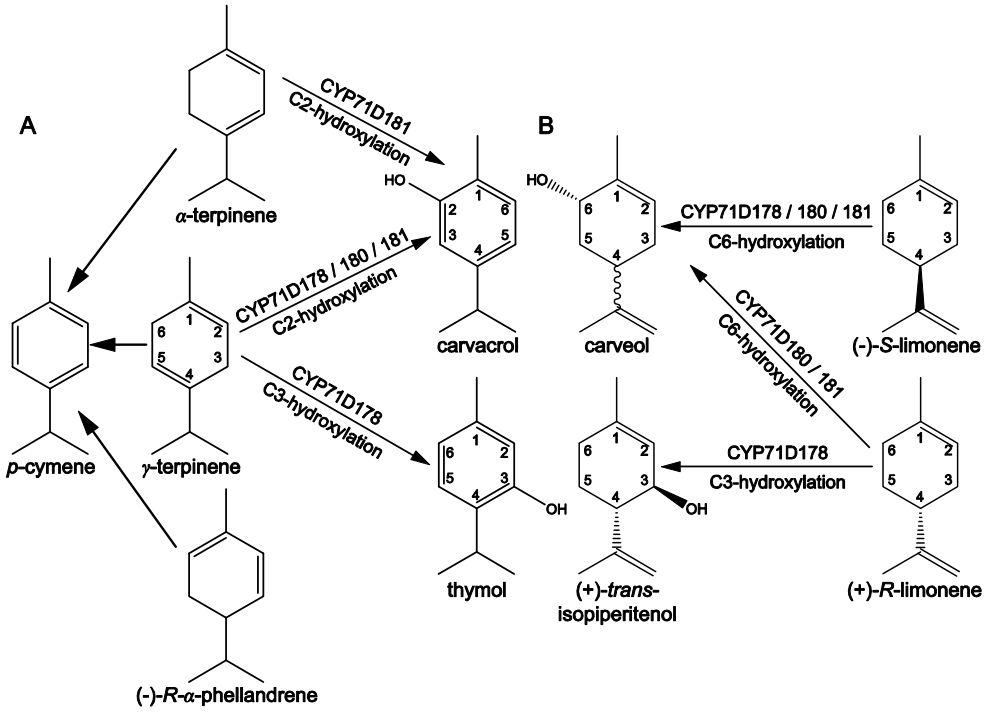


Figure 11 Chemical structures of all substrates and major products formed by CYP71D178, CYP71D180v1 and CYP71D181 *in vitro*. (**A**) All three enzymes form mainly *p*-cymene from γ-terpinene, α-terpinene and (-)-*R*-α-phellandrene as indicated by the arrows. CYP71D178 forms thymol and carvacrol from g-terpinene while CYP71D180v1 and CYP71D181 from only carvacrol. CYP71D181 forms carvacrol also from a-terpinene. (**B**) Major products formed from (-)-*S*-limonene and (+)-*R*-limonene. Mainly carveol is formed from (-)-*S*-limonene by all three enzymes. CYP71D180v1 and CYP71D181 hydroxylate (+)-*R*-limonene only at carbon C6 while CYP71D178 catalyzes a hydroxylation at C3. Hydroxyl groups at carbon C2 in carvacrol corresponds to that designated as C6 in carveol. The numbering differs because of different substituent priorities in the *p*-cymene vs. limonene carbon skeleton.

Given the similarity of CYP71D178-182 to the mint limonene hydroxylases CYP71D13 and 18, which produce allylic alcohols, the intermediate in a two-step reaction might be an allylic oxidation product of γ-terpinene. In this case, aromatization would constitute the second step. Interestingly, the two mint limonene hydroxylases exhibit different regiospecificities for their native (-)-*S*-limonene substrate, with CYP71D13 forming a C3 alcohol and CYP71D18 forming a C6 alcohol (equivalent to the C2 position on the thymol/carvacrol carbon skeleton). The formation of thymol and carvacrol could conceivably also involve regiospecific C2 and C3 oxidations of γ-terpinene catalyzed by separate P450s of the CYP71D family. Since CYP71D178 converted (+)-*R*-limonene to a C3 oxidation product and CYP71D180 and 181 to a C6 (C2) oxidation product (Figs. 7 and 11), CYP71D178 and other enzymes of high sequence similarity in group "a" (Fig. 3), CYP71D179 and CYP71D182, are potential thymol synthases

and that the group "b" enzymes, CYP71D180 and CYP71D181, are carvacrol synthases. Further characterization of CYP71D178-182 with a different expression system is necessary to confirm their catalytic properties.

The properties of oregano and thyme CYP 71D enzymes resemble those of P450 limonene hydroxylases from mint

The oregano and thyme P450s share many structural and biochemical characteristics with the limonene hydroxylases of mint, which are the most closely-related P450s on the basis of amino acid similarity that have been characterized to date. The shared characteristics include the potential substrate recognition sites SRS1-SRS6. These sites also differ among CYP71D178 through CYP71D182 which might explain some of the biochemical properties observed *in vitro*. In SRS5 (Fig. 4), it was reported that a single amino acid substitution (F361I) converts the regiospecificity of CYP71D18 from a C6- to a C3-hydroxylase (Schalk and Croteau, 2000). At the corresponding positions, CYP71D178 through D182 all bear a phenylalanine (F) like the mint C6-hydroxylase so this position cannot be responsible for regiospecific differences in catalysis. However, only two amino acids downstream of this position, there is a marked difference among the enzymes with CYP71D178 containing an isoleucine residue, CYP71D179 and 182 containing a methionine residue, and CYP71D180 and 181 (as well as the mint limonene hydroxylases, CYP71D13 and CYP71D18) containing a leucine at this position. An exchange in one of these residues might lead to a different regiospecificity.

The biochemical properties of the oregano and thyme CYP71D enzymes are slightly different from what was reported for CYP71D13 and CYP71D18 from mint (Karp et al., 1990) (Lupien et al., 1999). The pH optima for the mint P450s was reported to be around pH 7.4 whereas CYP71D178, CYP71D180v1 and CYP71D181 showed more acidic pH optima around 6.4 to 6.8. Although the K_m values of CYP71D180v1 for the limonene substrates were found to be much lower than for γ-terpinene, the velocity of conversion was much more rapid with γ-terpinene (Table 3). Since only very low amounts of limonene was found in the investigated plant lines (14- to 147-fold lower than γ-terpinene concentrations), γ-terpinene is expected to be the natural substrate. Still, it would be interesting to see whether elevated amounts of limonene would lead to the formation of carveol or isopiperitenol in these plants. For *Lippia* species, different chemotypes were described, some of which contain γ-terpinene together with *p*-cymene, thymol and/or carvacrol whereas others contain limonene together with carveol, carvone and piperitenone (Mesa-Arango et al., 2009; Escobar et al., 2010). To our knowledge no oregano or thyme chemotypes with elevated amounts of limonene or any of the latter products were described so far. A lemon-like thyme chemotype has been described that contains mainly geraniol, nerol and citral but no limonene or related products.

More information on the CYP71D178-182 enzymes and their role in thymol and carvacrol biosynthesis requires more precise characterization of their enzyme properties. Due to the presence of possible artifacts in yeast expression, overexpression in the model plant *A. thaliana* was tried. The results will be discussed in chapter III.

4 Chapter III

Thymol and carvacrol formation from γ-terpinene by CYP71D178 and CYP71D180 from oregano and thyme over-expressed in *A. thaliana*

4.1 Abstract

Thymol and carvacrol are biologically active compounds found in the essential oils of the two culinary herbs oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.). Several studies have shown an activity of thymol and carvacrol towards a broad range of organisms: microbes, fungi, insects and mammals. Also, important pharmaceutical and antioxidant functions are associated with these two phenolic monoterpenes.

A pathway for the formation of thymol in thyme had been predicted in the late 1970's to start with the monoterpene γ-terpinene and proceed via the aromatic *p*-cymene on the way to thymol and carvacrol (Poulose and Croteau, 1978a). Recently, different cytochrome P450s, CYP71D178 through CYP71D182, from thyme and oregano have been described to be involved in thymol and carvacrol biosynthesis and characterized from a yeast heterologous expression system. These enzymes were found to form thymol and / or carvacrol *in vitro* directly from γ-terpinene but could not use *p*-cymene as substrate (chapter II). However, they did produce *p*-cymene as a major product, so their catalytic roles are still uncertain. A second P450 might be present that converts *p*-cymene to thymol and carvacrol.

Here, we demonstrate direct thymol and carvacrol formation from γ-terpinene by two of these P450s, CYP71D178 and CYP71D180, after over-expression in the model plant *Arabidopsis thaliana*. Transgenic plants were fed with different monoterpene substrates including γ-terpinene and *p*-cymene. Thymol and carvacrol were formed by the transgenic plants from γ-terpinene whereas *p*-cymene feeding resulted in the formation of two other monoterpene alcohols, *p*-cymene-8-ol and cuminol (*p*-cymene-7-ol). Further experiments with structurally similar monoterpenes such as α-terpinene, (-)-*R*-α-phellandrene, (-)-*S*-limonene and (+)-*R*-limonene revealed that these P450s have broad substrate specificities, but are also capable of hydroxylation and aromatization of other monoterpenes. Thus, it is conceivable that the formation of thymol and carvacrol is catalyzed by single P450s directly from γ-terpinene via a two-step oxidation, whereas *p*-cymene is a side product resulting from premature release of the substrate from the active site.

The majority of the hydroxylated products formed by transgenic *Arabidopsis* plants were not released as free volatiles but bound as glycosides. This might be due to a detoxification mechanism to prevent cell damage.

4.2 Introduction

Plant cytochrome P450 monooxygenases (P450s) play important roles in metabolism by detoxifying xenobiotics and the biosynthesis of defense compounds against herbivores and pathogens (Schuler, 1996; Ohkawa et al., 1999; Werck-Reichhart et al., 2002). The impact of P450s becomes clear when one realizes the large sizes of P450 gene families in plants. In *Arabidopsis thaliana*, 246 putative P450 genes and 26 pseudogenes were annotated which represent approximately 1 % of its gene complement (Paquette et al., 2000; Werck-Reichhart et al., 2002; Schuler and Werck-Reichhart, 2003; Nelson et al., 2004), and in rice the number of 457 P450s is even more impressive (Schuler and Werck-Reichhart, 2003). Biosynthetic P450s catalyze steps in pathways leading to a vast range of plant compounds like lignin intermediates, sterols, furanocoumarins, flavonoids and terpenes (Bolwell et al., 1994; Schuler, 1996).

Cytochrome P450s are very important to generate some of the enormous structural diversity of terpenoid secondary metabolites in plants (Ro et al., 2005). Triterpenes, diterpenes, sesquiterpenes and monoterpenes are all substrates for these enzymes (Bolwell et al., 1994). For example, in diterpene resin acid biosynthesis in loblolly pine, multisubstrate, multifunctional P450s catalyze an array of consecutive oxidation steps with several different alcohol and aldehyde intermediates (Ro et al., 2005). In chicory, P450s catalyze the hydroxylation of sesquiterpenes (*Cichorium intybus* L.) (de Kraker et al., 2003). Monoterpenes are also substrates for P450s including acyclic monoterpenes, as in the 10-hydroxylation of geraniol in *Catharanthus roseus* (Meijer et al., 1993), and the hydroxylation of the bicyclic monoterpenes, sabinene and pinene (Karp et al., 1987) . The largest diversity of hydroxylations by cytochrome P450s is reported for limonene, and several P450s have been identified from mint, caraway and *Perilla* that use this cyclohexanoid monoterpene as a substrate (Karp et al., 1990; Bouwmeester et al., 1998; Mau et al., 2010). The regiospecificity of limonene hydroxylation can determine the downstream fate of the products in biosynthetic pathways. In mint, P450 mediated hydroxylation of (-)-*S*-limonene occurs either at position C3 or C6 (Lupien et al., 1995). Hydroxylation at C6 results in the formation of carveol which leads to the accumulation of carvone in spearmint whereas hydroxylation at C3 forms isopiperitenol in peppermint which is subsequently transformed into the commercially valuable menthol (Karp et al., 1990; Lupien et al., 1995; Croteau et al., 2005).

Recently, five cytochrome P450s, CYP71D178 through CYP71D182 have been isolated from oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.). They are involved in the biosynthesis of the two aromatic monoterpenes, thymol and carvacrol (chapter II). These two phenolic monoterpenes are especially known for their antiherbivore, antimicrobial, insecticidal, pharmaceutical and antioxidant activities (Isman, 2000; Hummelbrunner and Isman, 2001; Ultee et al., 2002; Sedy and Koschier, 2003; Floris et al., 2004; Braga et al., 2008).

Microsomal preparations from yeast over-expressing three of the described P450s, *CYP71D178*, *CYP71D180*v1 and *CYP71D181*, were found to form thymol and carvacrol directly from the monoterpene olefin, *γ*-terpinene. These results do not support an earlier prediction that the pathway to thymol and carvacrol starts with *γ*-terpinene and proceeds via the aromatic *p*-cymene as an intermediate (Poulose and Croteau, 1978a). Nevertheless, the *in vitro* conversion into thymol and carvacrol was rather low, and the major product formed by these P450s was *p*-cymene. This was regarded as side product which might arise as an artifact of the *in vitro* assay conditions. Arguments for the direct conversion were supported by the capability of all three P450s to form hydroxylated products from (-)-*S*- and (+)-*R*-limonene with the same regio-specific positioning of the hydroxyl groups as found in thymol and carvacrol (chapter II).

Here, we report the results from using *Arabidopsis thaliana* to over-express these P450s. In addition, we generated plants co-expressing *CYP71D178* and a gene encoding a *γ*-terpinene synthase. We also fed *γ*-terpinene directly to plant lines over-expressing *CYP71D178* and *CYP71D180*v1. The results support the prediction of a direct conversion of *γ*-terpinene to thymol and carvacrol by single cytochrome P450s in oregano and thyme. We therefore propose a new pathway for thymol and carvacrol formation directly from *γ*-terpinene, and that *p*-cymene is a side product formed by premature release from the active site of the cytochrome P450s involved.

4.3 Materials and Methods

Plant culture

Arabidopsis thaliana Col-0 wild type (WT) and transgenic plants were grown on soil in a controlled environment chamber (21°C, 55% relative humidity, and 100 µmol m^{-2} s^{-1} photosynthetically active radiation) for 4-5 weeks. The photoperiod was 10:14 hr light:dark for plants grown for monoterpene feeding experiments and 16:8 hr light:dark for plants grown for transformation or seed production.

Generation of transgenic plants

The complete open reading frames of γ-terpinene synthase *Ovtps*2-f0204 (GU385977) (Crocoll et al., 2010) and the cytochrome P450s, *CYP71D178* and *CYP71D180*v1 were transformed into *A. thaliana* Col-0. Expression constructs were created with Gateway technology (Invitrogen, Carlsbad, CA, USA). Validated sequences were subcloned from the pCR4-TOPO vector into the pDONR207 vector with BP Clonase II (Invitrogen, Carlsbad, CA, USA). Subsequently, the constructs were cloned into the pB2GW7,0 expression vector (Karimi et al., 2002) with LR Clonase II (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. *Agrobacterium tumefaciens* strain GV3101 transformed with the constructs or the empty pB2GW7,0 vector was used to transform *A. thaliana* ecotype Col-0 by vacuum infiltration (Bechtold et al., 1993). For selection of positive transformants, seeds were germinated on soil and watered 4 times with PESTANAL™ solution (200 g l^{-1} glufosinate ammonium, Hoechst Schering AgrEvo, Düsseldorf, Germany). Transformed plants surviving this treatment were further selected for cytochrome P450 over-expression by RNA hybridization analysis. Whole leaf RNA was isolated from two expanding leaves. Leaves were harvested and frozen with liquid nitrogen prior to homogenization and RNA extraction. Total RNA (7 µg) was run on a denaturing RNA gel (1 % (w/v) agarose, 10 % (v/v) NorthernMax 10 x denaturing gel buffer, Ambion, Austin, TX, USA), for approx. 70 min at 100 V. The RNA was blotted onto a Hybond-XL membrane (Amersham, Piscataway, NJ, USA) by capillary transfer using NorthernMax 10 x running buffer (Ambion, Austin, TX, USA) overnight. RNA was UV-crosslinked to the membrane two times with a fluence of 120 mJ cm^{-2}. For both prehybridization and hybridization, UltraHyb buffer (Ambion, Austin, TX, USA) was used. Probes were labeled with α[^{32}P]dATP using the Strip-EZ PCR Kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. The probe for CYP71D178 was amplified as ssDNA from a short fragment of 360 bp (bp 1079-1438) from its open reading frame. Probes were hybridized overnight at 42 °C. Membranes were washed twice with 3 x SSC at 42 °C, once with 1 x SSC at 42 °C, once with 0.1 x SSC at 42 °C, once with 0.1 x SSC at 50 °C and depending on the signal

strength once with 0.1 x SSC at 55-68 °C. All SSC buffers contained 0.1 % SDS. Washed blots were sealed in PVC bags and blots were exposed to storage phosphor screens (Amersham Bioscience, Uppsala, Sweden) for 4 to 5 h and analyzed with a Storm 840 scanner (Molecular Dynamics, Sunnyvale, CA, USA). *A. thaliana* plants showing high expression rates of *CYP71D178* or *CYP71D180*v1 were chosen for subsequent seed production. Independent homozygous lines with a single T-DNA insertion were chosen based on the segregation of PESTANAL^{TM} resistance. In brief, seeds were sterilized and germinated on MS medium (2 % (w/v) sucrose, 0.9 % (w/v) agar, and 30 µg ml^{-1} PESTANAL^{TM}). Two independent lines were chosen from each transformation event. *Ovtps*2-f0204: f7 and f39; *CYP71D178*: f5-6 and f5-8; *CYP71D180*v1: T28-12 and T28-22; pB2GW7,0: Vec-3. Plant crossings and monoterpene feeding were done with T3 generation plants.

Crossing of transgenic *A. thaliana*

T3-generation plants were grown under long day conditions (16 h light, 60 % humidity). The pistil of unopened flowers was dissected and stamina completely removed. Pollen was collected from a second plant expressing the other gene of interest and transferred to the dissected pistils. Crossings were done in both directions. All other flowers were removed. Seeds were collected after approximately 10 days and sown on soil. *Arabidopsis* plants were grown until flowering stage and flowers were checked for their volatile emission by solid phase micro extraction (SPME). SPME fibers (PDMS-100, Polydimethylsiloxane, Supelco, Bellefonte, PA, USA) were exposed for 30 min to open flowers.

Administration of volatile terpenes to transgenic *A. thaliana* plants

Transgenic *A. thaliana* plants (T3 generation) were grown under short day conditions with 8 hours light and 60 % humidity. Feeding experiments were performed in two different ways. In a first trial, the plants were separated by line. Two plants from each of two transgenic plant lines of *CYP71D178* (lines f5-6 and f5-8) and *CYP71D180*v1 (lines 12 and 22) and one vector control line (pB2GW7,0 line 3) and Col-0 wild type were put in 1 l glass beakers with sealed lids. Monoterpenes were added on 3 consecutive days with 100 µl (~ 75 mg) monoterpene each day, except for (-)-*R*-*α*-phellandrene which was only fed in amounts of 50 µl (~ 37.5 mg) per day. These preliminary experiments were used to pick over-expressing lines of both P450s for subsequent experiments and to test the vector control and wild type *Arabidopsis* plants for their hydroxylation activities (Results are shown in Supplementary Material, Figures S4-S9). Based on the results of these first trials, transgenic lines *CYP71D178* line f5-6 and *CYP71D180*v1 line 22 were chosen for a second trial. Representatives of these lines and control plants (pB2GW7,0

line 3 and Col-0 WT) were put in 3 l glass beakers with sealed lids. Pots with three plants of each line were put in one glass beaker and 100 µl (~ 75 mg) pure monoterpenes were added on five consecutive days to supply fresh, unmodified substrate. In both trials, monoterpenes were applied in 5 ml beakers set next to the plants for evaporation. Approximately 65 mg of γ-terpinene and each of the other monoterpenes (15-20 % of the amount added to the beakers) was volatilized over the course of the five days and thus available for conversion by the plants. Transgenic and wild type *Arabidopsis* plants were 4 to 5 weeks old when the experiments were started. Complete rosettes were harvested 24 h after the last monoterpene application, immediately frozen in liquid nitrogen and freeze dried for 2 days.

Extraction and determination of glycosidically bound monoterpene alcohols from *A. thaliana*

For quantification of glycosidically bound monoterpene alcohols freeze-dried leaf material was ground to a fine powder in a paint shaker (Kliebenstein et al., 2001b) and approx. 20 mg were extracted with 1 ml 80 % methanol containing β-n-octyl-glucopyranoside as an internal standard (5 µg ml^{-1}) to control for the efficiency of hydrolysis. Leaf material was extracted over night with constant shaking at room temperature. The solution was centrifuged at 16,100 x *g* for 10 min. Then, 750 µl of the supernatant were transferred into 1.5 ml glass vials and reduced under nitrogen at 35 °C to complete dryness. The dried plant extracts were resuspended in 300 µl citric acid buffer, pH 5.2. Next, 100 µl of the same buffer supplied with 10 mg ml^{-1} β-glucosidase from *Aspergillus niger* (Sigma-Aldrich, Steinheim, Germany) were added. The solution was carefully covered with 200 µl pentane containing nonyl acetate as internal standard (10 ng $µl^{-1}$) and incubated in a water bath at 37 °C for 24 h. The reaction was stopped by 2 min shaking at 1400 rpm. The vials were centrifuged for 10 min at 4200 x *g* to divide solvent phases and frozen at -80 °C for > 2 hours. The pentane layer was recovered and transferred to a new vial for analysis on GC-MS and GC-FID for product identification and quantification. A range of 1.5-282.5 µg (0.06-1 %) of the various substrates administered were hydroxylated and converted to glycosides (Supplementary Material, Table S9).

Products from initial feeding experiments were collected by solid phase micro extraction (SPME). SPME fibers (PDMS-100, Polydimethylsiloxane, Supelco, Bellefonte, PA, USA) were exposed for 5 min to *A. thaliana* plants fed for 24 h with monoterpenes and bound volatiles were analyzed by GC-MS. Alternatively, methanol extracts from *A. thaliana* plants fed with monoterpenes were dried as described above and treated with β-glucosidase (1 mg) for 2 hours. SPME fibers were exposed for 20 min to volatiles released from these assays and bound terpene volatiles were analyzed by GC-MS.

Analysis of terpene volatiles by gas chromatography

Products of β-glucosidase assays were identified by gas chromatography (Agilent Hewlett-Packard 6890, Agilent Technologies, Santa Clara, CA, USA) coupled to a mass spectrometer (Agilent Hewlett-Packard 5973, Agilent Technologies) or a flame ionization detector (FID). For analyses, 2 µl of pentane extracts were injected with an injector temperature of 230 °C. Alternatively, a SPME fiber exposed to volatiles as described above was introduced into the injector. The terpenes were separated on a DB5-MS column: 30 m length, 0.25 mm inner diameter and 0.25 µm film (J&W Scientific, Santa Clara, CA, USA). Different GC-programs were used with initial temperature from 40-65 °C for 2 min, first ramp 2-6 °C min^{-1} to 175 °C, second ramp 90 °C min^{-1} to 250 °C, final 3 min hold. Limonene enantiomers and hydroxylated products thereof were further identified on a chiral column (HYDRODEX®-ß-3P: 25 m length, 0.25 mm inner diameter (Macherey-Nagel, Düren, Germany); GC-program: 80 °C for 2 min, first ramp 2 °C min^{-1} to 165 °C, second ramp 50 °C min^{-1} to 200 °C, final 3 min hold. GC-MS carrier gas: helium at 1 ml min^{-1}; GC-FID carrier gas: hydrogen at 2 ml min^{-1}. All terpene products were identified by using Agilent Technologies software with the Wiley275.L, NIST98.L and Adams2205.L MS libraries, as well as by comparison of mass spectra and retention times with those of authentic standards (Sigma-Aldrich Chemicals, Steinheim, Germany). The amounts of the individual terpenes were determined by GC-FID.

Statistical analysis was performed with SigmaStat (Version 2.03, Systat Software, Erkrath, Germany).

4.4 Results

A. thaliana plants expressing *CYP71D178* and a *γ*-terpinene synthase produce low amounts of *p*-cymene

To determine the catalytic abilities of the cytochrome P450, CPY71D178 from oregano (*Origanum vulgare* L.), the gene was over-expressed in *Arabidopsis thaliana*. Plants over-expressing *CYP71D178* were then crossed with a line over-expressing the *γ*-terpinene synthase *Ovtps*2 (GU385977*)* from oregano to supply the necessary substrate, which is not found in *A. thaliana*. All genes were constitutively expressed under the control of a 35S promoter. Transgenic *Arabidopsis* plants expressing *Ovtps*2 produced *γ*-terpinene and also *p*-cymene almost exclusively from the flowers as volatiles (Fig. 1a). Terpene emission from leaves was hardly detectable (data not shown). The simultaneous expression of *CYP71D178* and *Ovtps*2 in *A. thaliana* resulted in the exclusive formation of *p*-cymene (Fig. 1b). We could not detect any volatile or glycosidically bound hydroxylated products from plants co-expressing *CYP71D178* and *Ovtps*2.

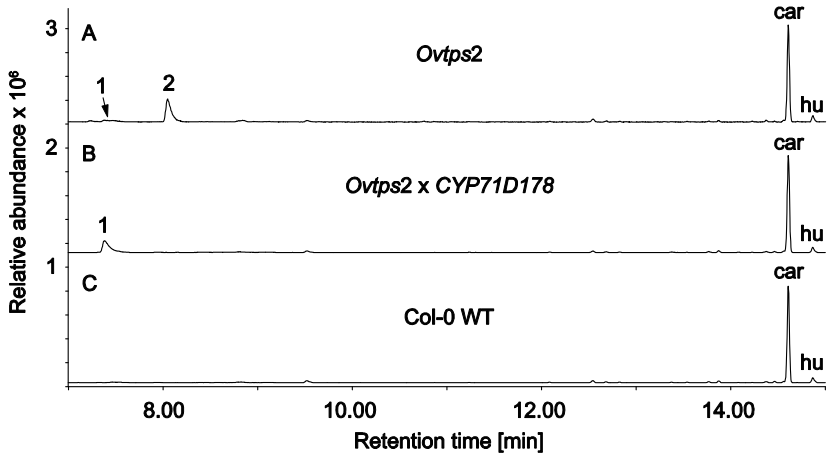


Figure 1 GC-MS traces of volatiles released from flowers of transgenic *A. thaliana* expressing *CYP71D178* and a *γ*-terpinene synthase show release of *p*-cymene. (**A**) Volatiles released from plants over-expressing only the *γ*-terpinene synthase *Ovtps*2: 1, *p*-cymene; 2, *γ*-terpinene; car, *(E)-β*-caryophyllene; hu, *α*-humulene. (**B**) Volatiles released from plants co-expressing both *CYP71D178* and the *γ*-terpinene synthase, *Ovtps*2: 2, *p*-cymene; car, *(E)-β*-caryophyllene; hu, *α*-humulene. (**C**) Volatiles released from WT *A. thaliana* plants: car, *(E)-β*-caryophyllene; hu, *α*-humulene. (*E)-β*-caryophyllene and *α*-humulene are released constitutively by *Arabidopsis* flowers and were used as a reference for terpene formation by transgenic plant lines. Volatiles were collected by SPME.

Feeding of different monoterpenes to transgenic *A. thaliana* plants over-expressing *CYP71D178* or *CYP71D180*v1 results in hydroxylated products bound as glycosides

As another approach to supplying γ-terpinene as substrate for the over-expressed P450s, monoterpenes were fed to transgenic *Arabidopsis* plants over-expressing *CYP71D178* from oregano or *CYP71D180*v1 from thyme in closed glass vessels via the surrounding air (Fig. 2c). Various monoterpenes were used as substrates: γ-terpinene as the predicted initial substrate, *p*-cymene as the potential intermediate and α-terpinene and (-)-*R*-α-phellandrene as structurally similar substrates to γ-terpinene. In addition, (+)-*R*-limonene and (-)-*S*-limonene were tested since these gave different hydroxylation products from *in vitro* assays with CYP71D178, CYP71D180v1 and CYP71D181 (chapter II). Moreover, both limonene substrates were used as references in order to qualify the amounts of products formed in comparison to the amounts fed. Both limonene enantiomers are converted into hydroxylated products by both P450s with high efficiency *in vitro* (chapter II). All tested monoterpenes have similar structures and were assumed to have a similar efficiency in permeating plant membranes to encounter the over-expressed cytochrome P450s.

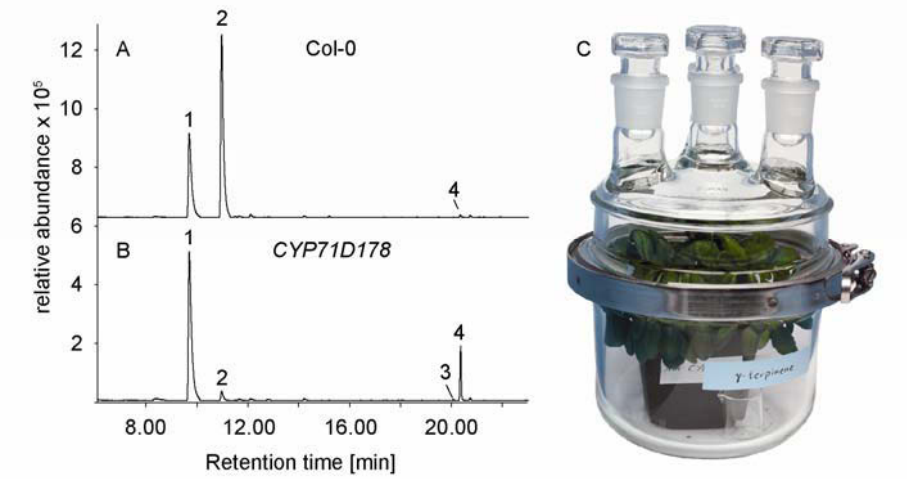


Figure 2 GC-MS traces of volatiles released from leaves of transgenic *A. thaliana* over-expressing *CYP71D178* after feeding with 50 µl γ-terpinene for 24 h in a closed glass vessel. SPME measurements of (**A**) Volatiles released from Col-0 wild-type control plants. (**B**) Volatiles released from transgenic *A. thaliana* plants: 1, γ-terpinene; 2, *p*-cymene; 3, thymol; 4, carvacrol. (**C**) Experimental setup for monoterpene feeding to *A. thaliana* plants. Four to five week old plants were put into closed glass vessels and monoterpenes were applied into a glass beaker next to the plants.

Initial feeding tests with γ-terpinene indicated that, as in the *in vitro* assays with microsomal protein produced in *S. cerevisiae*, *p*-cymene was released as the major volatile product. Only small amounts of thymol and carvacrol were released from transgenic *A. thaliana* plants. Wild type *A. thaliana* Col-0 plants fed with γ-terpinene released *p*-cymene too, and trace amounts of carvacrol (Fig. 2a). Nevertheless, carvacrol and thymol were identified from plants over-

expressing *CYP71D178* in much larger proportions than those observed from wild type *A. thaliana* plants (Fig. 2b).

p-Cymene is spontaneously formed from γ-terpinene, α-terpinene and (-)-*R*-α-phellandrene upon contact with oxygen (Granger et al., 1964). After 24 h, 30 to 40 % of the volatile γ-terpinene and 50 to 70 % of the volatile (-)-*R*-α-phellandrene were found to have been converted into *p*-cymene when put in a glass vessel (Supplemental Material, Figures S3, S4). Therefore, it was not possible to determine the exact conversion of these substrates to *p*-cymene by the introduced cytochrome P450s. In the presence of transgenic *A. thaliana* plants over-expressing the oregano and thyme P450s, more than 90 % of volatile γ-terpinene was transformed into *p*-cymene.

We suspected that some thymol and carvacrol formed by the P450s could have been stored as glycosides since linalool synthase over-expression in petunia resulted in the formation of S-linalyl-β-D-glucopyranoside (Lücker et al., 2001) rather than in the release of volatile linalool. Therefore, methanol extracts from transgenic *A. thaliana* plants that had been fed with γ-terpinene were subjected to β-glucosidase treatment. This resulted in the release of thymol, carvacrol and thymoquinone from *A. thaliana* plants over-expressing *CYP71D178* which were quantified by GC-FID (Fig. 3). Thymoquinone was found to arise from further hydroxylation of thymol by *A. thaliana* by direct feeding of wild-type plants with thymol (Data not shown). The parent thymol and carvacrol glycosides could also be detected and identified by LC-MS.

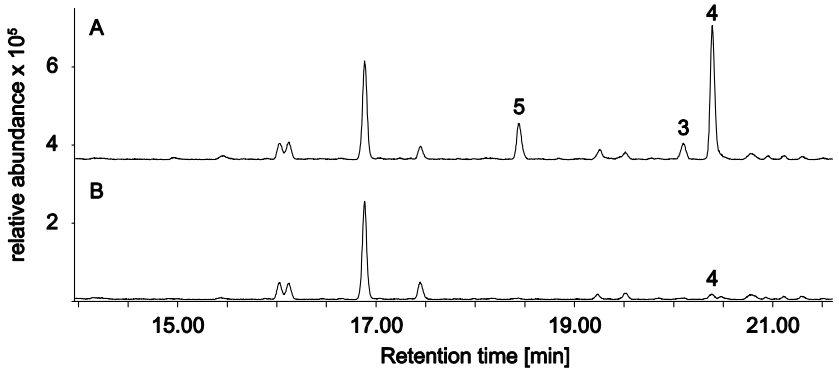


Figure 3 Detection of monoterpene glycosides in transgenic *A. thaliana* expressing *CYP71D178*. (**A**) GC-MS traces of volatiles from leaf extracts of plants fed γ-terpinene for 24 h. The extract was treated with β-glucosidase. Volatiles were collected by SPME: 3, thymol; 4, carvacrol; 5, thymoquinone. (**B**) As above, except extract was not treated with β-glucosidase: 4, carvacrol.

Transgenic *A. thaliana* plants over-expressing *CYP71D178* or *CYP71D180*v1 form thymol and carvacrol from γ-terpinene

Thymol and carvacrol were both readily detected as glycosides from *A. thaliana* over-expressing the thyme and oregano CYP71D P450s. While thymol was found in low amounts

only in *CYP71D178* plant lines, carvacrol was found in higher amounts in these lines and in those over-expressing *CYP71D180v1* (Fig. 4). Both control lines (vector control and wild type plants) contained only small amounts of carvacrol, cuminol (*p*-cymene-7-ol) and *p*-cymene-8-ol. Other minor products can be found in Table S8, Supplementary Material.

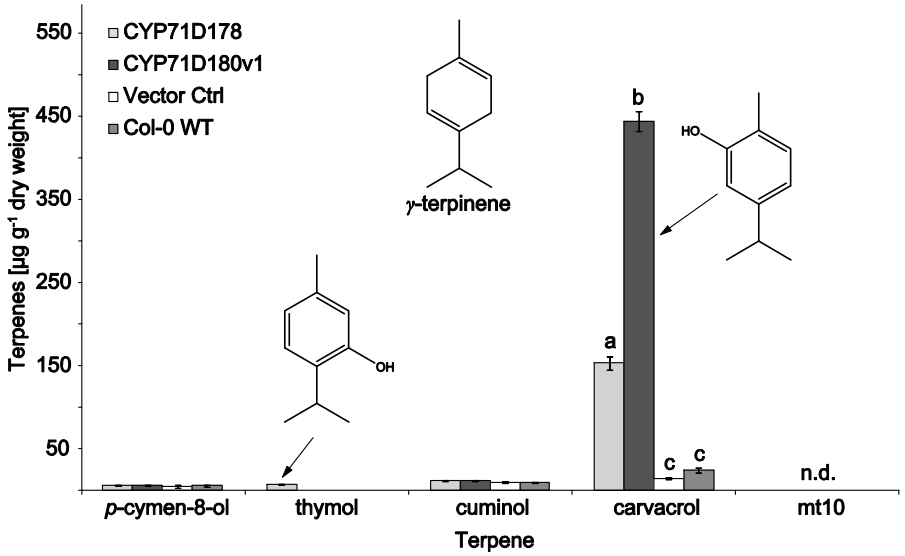


Figure 4 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *γ*-terpinene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. Structures of substrate, *γ*-terpinene, and important products, thymol and carvacrol, are shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound. n.d. = not detectable.

Other cyclohexanoid monoterpenes are al so metabolized by the P450s, bu t not *p*-cymene

Feeding of *p*-cymene to transgenic, vector control and wild-type lines resulted in the formation of *p*-cymene-8-ol or cuminol as principal products, but thymol and carvacrol were not detected (Fig. 5). Other cyclohexanoid, non-aromatic monoterpenes were metabolized by P450 over-expression lines. The monoterpene *α*-terpinene, an isomer of *γ*-terpinene with one double bond in an adjacent position in the ring was converted to carvacrol by CYP71D180v1, but CYP71D178 did not produce significantly more carvacrol than the control lines (Fig. 6). (-)-*R*-*α*-Phellandrene, another cyclohexanoid monoterpene diene, was converted solely to an unknown monoterpene product (Fig. 7) with an *m/z* of 152, that is likely a hydroxylated monoterpene without an aromatic ring system (Fig. 8).

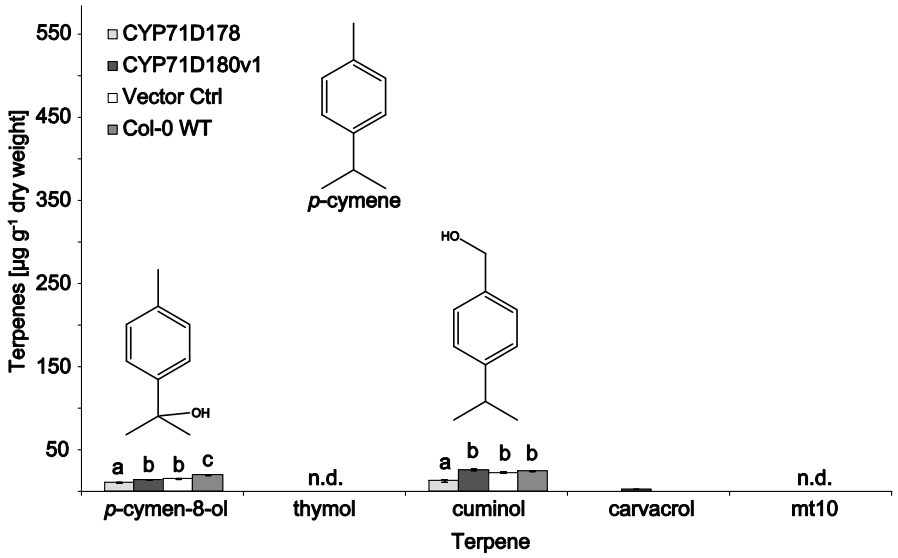


Figure 5 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *p*-cymene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. Structures of substrate, *p*-cymene, and major products, *p*-cymene-8-ol and cuminol, are shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound. n.d. = not detectable.

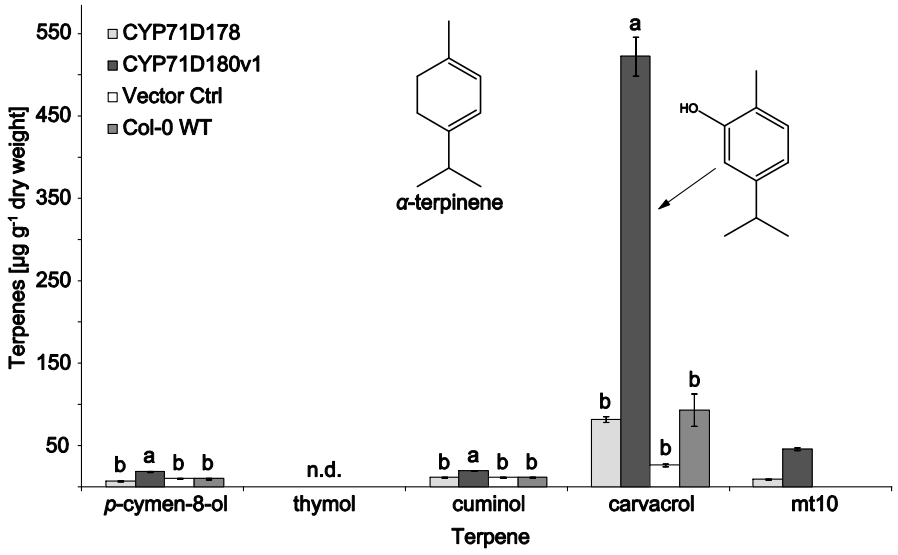


Figure 6 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *α*-terpinene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. Structures of substrate, *α*-terpinene, and the major products, carvacrol, are shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound. n.d. = not detectable. mt10 = unidentified monoterpene.

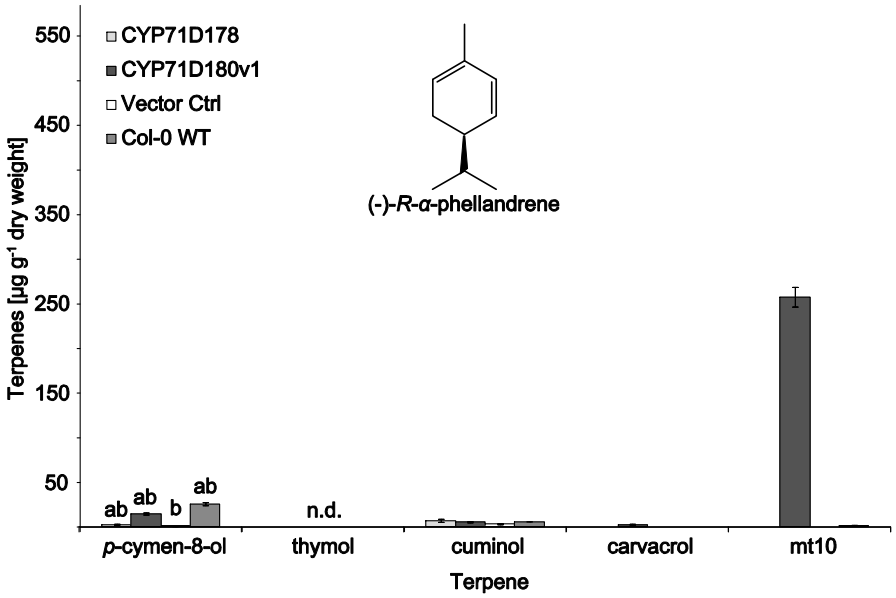


Figure 7 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with (-)-*R*-*α*-phellandrene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. The structure of the substrate, (-)-*R*-*α*-phellandrene, is shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound. n.d. = not detectable. mt10 = unidentified monoterpene.

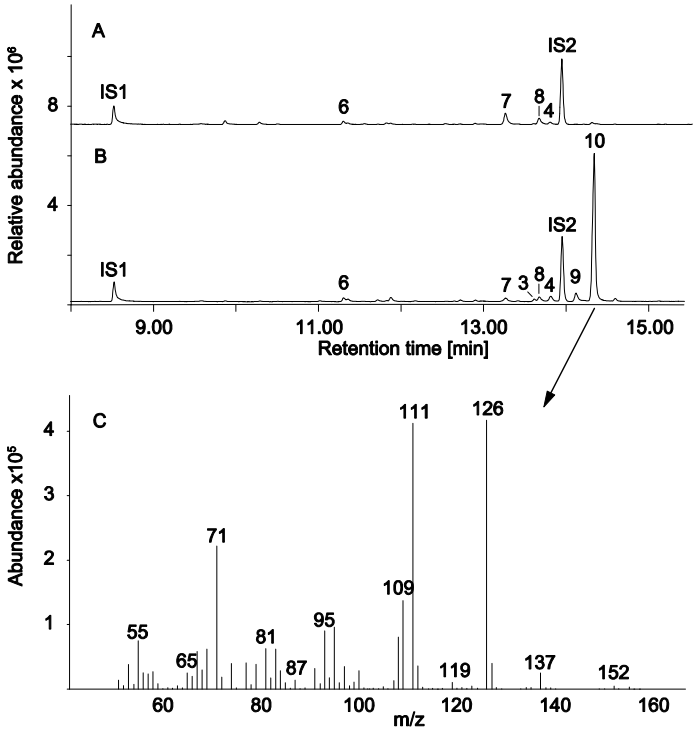


Figure 8 GC-MS-chromatogram of CYP71D180 plant line extract after β-glucosidase treatment. (**A**) GC-MS traces of volatiles released from extracts of *A. thaliana* Col-0; IS1, internal standard 1; 6, p-cymene-8-ol; 7, unidentified monoterpene; 8, cuminol; 4, carvacrol; IS2, internal standard 2. (**B**) GC-MS traces of volatiles released from extracts of *A. thaliana* over-expressing *CYP71-D180*v1; IS1, internal standard 1; 6, p-cymene-8-ol; 7, unidentified monoterpene; 3, thymol; 8, cuminol; 4, carvacrol; IS2, internal standard 2; 9, unidentified monoterpene; 10, unidentified monoterpene (mt10). (**C**) MS-Spectrum of peak 10 represents an unidentified monoterpene alcohol.

Limonene is converted into the allylic alcohols, carveol and isopiperitenol

The feeding of limonene enantiomers resulted in the same regiospecific production of hydroxylated monoterpenes as was already observed *in vitro* with microsomal protein from yeast expression (chapter II). Plant line CYP71D178 accumulated mainly (-)-*trans*-carveol from (-)-*S*-limonene feeding, but also some (-)-*trans*-isopiperitenol (Fig. 9). (+)-*trans*-Isopiperitenol was formed from (+)-*R*-limonene (Fig. 10). CYP71D180v1 plant lines formed almost exclusively (-)-*trans*-carveol from both limonene enantiomers (Figs. 9 and 10). *A. thaliana* wild type and vector control lines formed mainly limonene-10-ol or perillyl alcohol from both enantiomers which were also present in similar amounts in the P450 over-expressing plant lines. The overall conversion with the limonene substrates was slightly higher than with the other substrates tested. The high amounts of carveol present in vector control and wild type *A. thaliana* after feeding especially with (-)-*S*-limonene probably result from cross-contamination. Separate feeding of these lines resulted only in very low carveol formation (Supplementary Material, Figs. S8, S9)

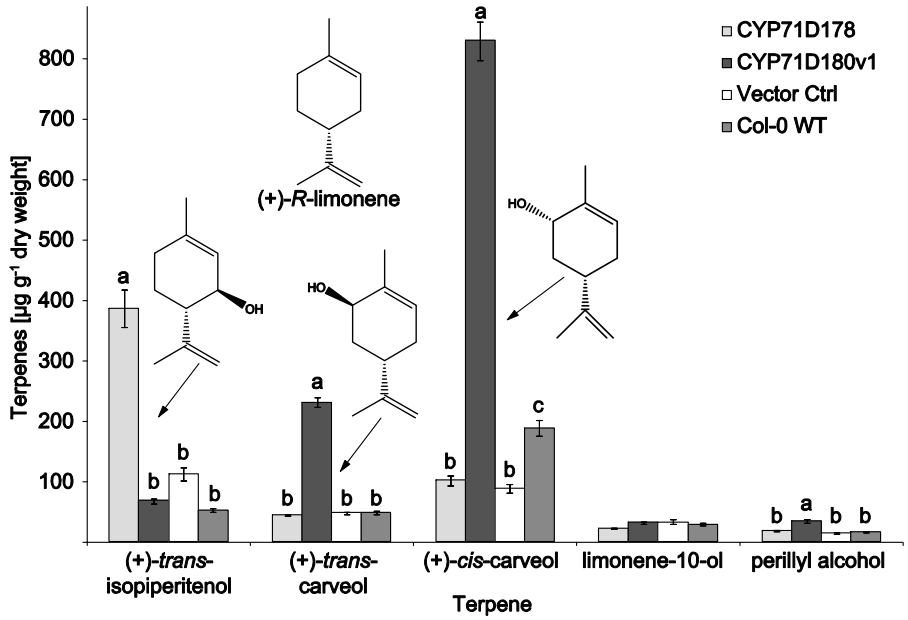


Figure 9 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with (+)-*R*-limonene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. Structure of the substrate, (+)-*R*-limonene, and major products, (+)-*trans*-isopiperitenol, (+)-*trans*-carveol and (+)-*cis*-carveol, are shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound.

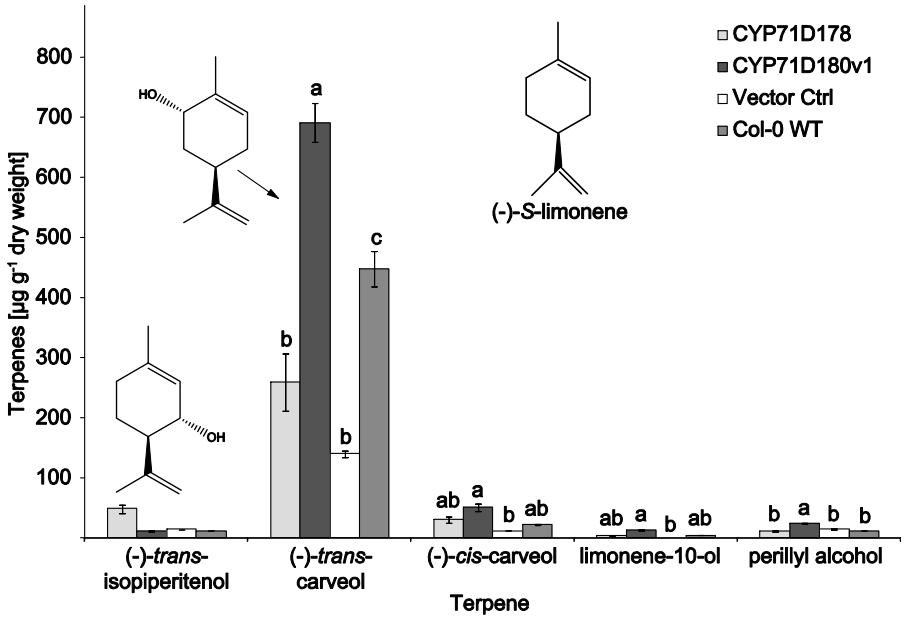


Figure 10 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with (-)-*S*-limonene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* (Col-0 WT) plants. Structure of the substrate, (-)-*S*-limonene, and major products, (-)-*trans*-isopiperitenol and (-)-*trans*-carveol, are shown. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values ± standard error (n = 9). Mean values were tested for significant differences ($P < 0.05$) by One-Way ANOVA followed by Tukey's test for all pairwise comparisons. Significant differences are indicated by different letters. Statistics were performed separately for each compound.

4.5 Discussion

CYP71D178 and CYP71D180 convert γ-terpinene into thymol and carvacrol

The heterologous expression of three cytochrome P450s from oregano and thyme, CYP71D178, CYP71D180v1 and CYP71D181 in yeast had already hinted towards a direct conversion of γ-terpinene into thymol and carvacrol. This hypothesis had been supported by correlation of the expression of these genes with thymol and carvacrol accumulation, by trace production of thymol and carvacrol upon administration of γ-terpinene, and by the regiospecificity of hydroxylation with (+)-*R*-limonene as substrate at either carbon C3 or C6. The positions of the hydroxyl groups resembled very much those of thymol and carvacrol (chapter II) (Fig. 11). In the present study we could confirm these findings by feeding γ-terpinene to transgenic *A. thaliana* over-expressing *CYP71D178* or *CYP71D180*v1. *CYP71D178* over-expressers formed both thymol and carvacrol (accumulated as glycosides) while *CYP71D180*v1 over-expressors formed only carvacrol from γ-terpinene (Fig. 4).

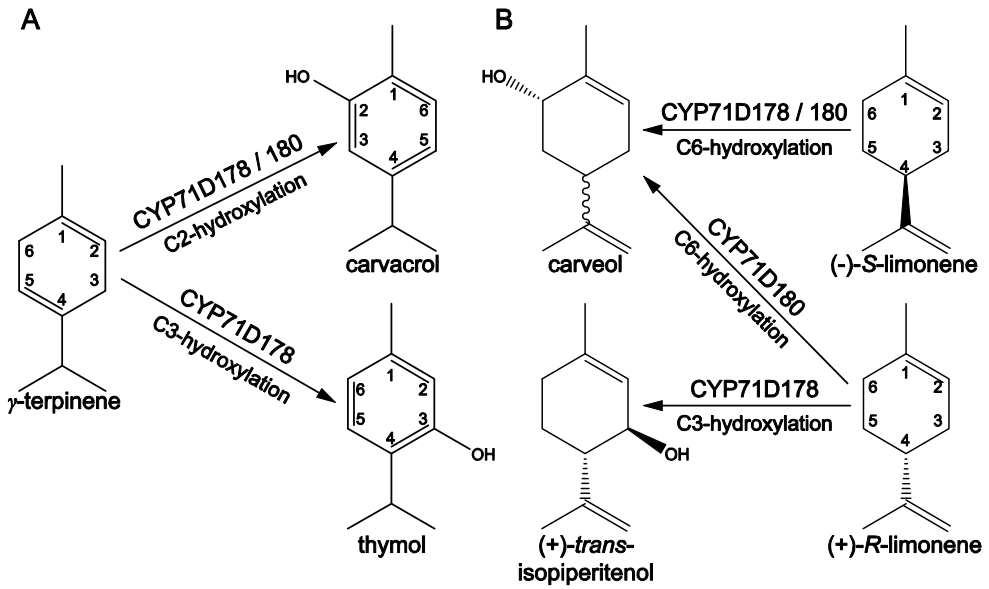


Figure 11 Summary of products formed from γ-terpinene and limonene fed to *A. thaliana* lines transformed with CYP71D genes. (**A**) Products formed from γ-terpinene: CYP71D178 catalyzes C2- and C3-hydroxylations to carvacrol or thymol while CYP71D180v1 forms only carvacrol. (**B**) Products formed from (+)-*R*- and (-)-*S*-limonene: CYP71D178 catalyzes C3- or C6-hydroxylations depending on the limonene enantiomer while CYP71D180v1 catalyzes only C6-hydroxylations from both enantiomers.

The aromatic monoterpene, *p*-cymene, was found as the major product from γ-terpinene when these P450s were expressed in yeast and microsomal extracts assayed *in vitro*. Here, a similar pattern was found on feeding γ-terpinene to transgenic *A. thaliana* over-expressing *CYP71D178* or *CYP71D180*v1. However, *p*-cymene was also formed by wild type and vector control

Arabidopsis plants. Given these results and the previously described spontaneous formation upon contact with oxygen (Granger et al., 1964), we conclude that the formation of *p*-cymene *in vitro* and *in vivo* is very likely an artifact of enzyme catalysis in a heterologous system.

The fact that CYP71D178 and 180 convert γ-terpinene to thymol and carvacrol, a process requiring two formal oxidations, is not unusual for a member of the P450 family. Catalysis of multi-step oxidations is well known for a number of cytochrome P450s. In diterpene resin acid biosynthesis in loblolly pine, multisubstrate, multi-step P450s catalyze an array of consecutive oxidation steps with several different alcohol and aldehyde intermediates (Ro et al., 2005). Another example for multifunctional P450s is the well studied pathway of cyanogenic glucoside formation where the two P450s each catalyze a series of multi-step oxidations on the way from amino acids to α-hydroxynitriles (Halkier et al., 1995; Bak et al., 1998).

p-Cymene is not an intermediate in thymol and carvacrol formation by oregano and thyme CYP71D P450s

The aromatic monoterpene *p*-cymene was originally suggested as intermediate in the pathway of thymol and carvacrol biosynthesis proposed for thyme over thirty years ago (Poulose and Croteau, 1978a). The initial substrate, γ-terpinene, was predicted to be oxidized to *p*-cymene which in a second step is hydroxylated to form either thymol or carvacrol (Fig. 1a). This was a likely pathway since cytochrome P450s are often responsible for the hydroxylation of aromatic rings (Jerina and Daly, 1974), which are present in many pathways for plant secondary products such as furanocoumarins, anthocyanins, flavonoids and many more (Schuler, 1996).

In the work reported here, we were unable to find any proof for the intermediacy of *p*-cymene in the conversion of γ-terpinene to thymol and carvacrol. In the feeding experiments described in this chapter with the *A. thaliana* lines over-expressing *CYP71D178* and *CYP71D180*v1 and the *in vitro* assays conducted with CYP71D178, 180v1 and 181 protein expressed in yeast (chapter II), it was found that *p*-cymene itself was not converted to form thymol and carvacrol in any amounts above background levels (Supplementary Material, Fig. S5). Instead, *p*-cymene formed two other products at much higher rates, cuminol and *p*-cymene-8-ol, which carry hydroxyl groups at carbon positions C7 and C8 outside the aromatic ring. These two derivatives were most probably formed by enzymes from *Arabidopsis thaliana* itself, possibly by one of the 246 P450s found in the *Arabidopsis* genome (Paquette et al., 2000; Werck-Reichhart et al., 2002; Schuler and Werck-Reichhart, 2003). Some of these P450s might be capable of catalyzing the hydroxylation of the aromatic *p*-cymene to thymol possibly by an NIH shift mechanism (Guroff et al., 1967).

Thymol formation from *p*-cymene was reported to be catalyzed by several human P450 enzymes (Meesters et al., 2009). To insure that this conversion would not have been missed in

our own system, one of the P450s from this study, the human enzyme, CYP2D6, was purchased and tested for its activities. Unfortunately, we could not confirm thymol formation from *p*-cymene. Instead we identified carvacrol formation from *γ*-terpinene and cuminol formation from *p*-cymene rather than thymol (data not shown).

CYP71D178 and CYP71D180 give different profiles of thymol and carvacrol formation

While *CYP71D178* over-expressing *A. thaliana* plants formed both thymol and carvacrol, *CYP71D180* over-expressers produced only carvacrol from *γ*-terpinene (Fig 4). Such differences in the regiospecificity of hydroxylation reactions are well known for cytochrome P450s (Schuler, 1996) and have been especially well described for several P450s hydroxylating limonene in mint, caraway and *Perilla* (Karp et al., 1990; Bouwmeester et al., 1998; Mau et al., 2010).

The oregano and thyme P450s, CYP71D178, CYP71D180v1 and CYP71D181, also showed such a difference in the regiospecificity of the hydroxylation from (-)-*S*- and (+)-*R*-limonene *in vitro* (chapter II). In the present study with *A. thaliana* over-expression lines, we found the same pattern. *CYP71D178* over-expressing plant lines form mainly (-)-*trans*-carveol from (-)-*S*-limonene and (+)-*trans*-isopiperitenol from (+)-*R*-limonene whereas *CYP71D180*v1lines form only carveols from either substrate (Figs. 9, 10). The position of the hydroxyl-group in carveol at carbon C6 is identical to C2 in carvacrol whereas the C3-hydroxylation in isopiperitenol is identical in its position to the hydroxyl group in thymol (Fig. 11).

The reason for these differences in hydroxylation position might be related to differences found in the substrate recognition sites (SRS) of these enzymes as described in chapter II. CYP71D180 and CYP71D181 belong to one subgroup of the P450s studied while CYP71D178, CYP71D179 and CYP71D182 belong to a second subgroup (chapter II). For the two mint limonene hydroxylase P450s, CYP71D13 and CYP71D18, it was shown that only one amino acid residue in SRS5 is responsible for a different regiospecificity of limonene hydroxylation, C3 vs. C6 (Schalk and Croteau, 2000). The SRS5 of oregano and thyme CYP71D P450 sequences shows a similar pattern. All CYP71D180 and CYP71D181 sequences contain a leucine residue at amino acid position 364 whereas all CYP71D179 and CYP71D182 sequences contain a methionine. CYP71D178 is different from all other sequences with an isoleucine at the same position. Other major amino acid differences between the two oregano and thyme CYP71D P450 subgroups are found within the other SRS (chapter II). Whether these differences, especially in SRS5 are responsible for the formation of either thymol or carvacrol and isopiperitenol or carveol needs further investigations, e.g. site directed mutagenesis combined with structural modeling of these P450s. *In vitro* CYP71D180 and CYP71D181 had

shown identical hydroxylation patterns with both limonene substrates. Whether CYP71D179 and CYP71D182 show similar or different hydroxylation abilities compared to CYP71D178 is currently under investigation.

Based on these data we suggest that the two subgroups found in the amino acid alignment of the five CYP71D P450s from oregano and thyme shown in chapter II represent also a functional difference. CYP71D178, CYP71D179 and CYP71D182 are proposed to be thymol synthases, while CYP71D180 and CYP71D181 are proposed to be carvacrol synthases.

Thyme and oregano CYP71D P450s have broad substrate specificities

The narrow product specificity of hydroxylation at only two positions in the cyclohexanoid ring is paired with broad substrate specificity. The two CYP71D P450s studied in *A. thaliana* accepted five different cyclohexanoid monoterpenes as substrates. All accepted substrates contain two double bonds, at least one of which is within the cyclohexanoid ring. The position of the double bonds within the ring had an effect on the product. While *γ*-terpinene was hydroxylated at either C2 or C3 by CYP71D178, *α*-terpinene was exclusively hydroxylated at C2 by both enzymes. Moreover, the rate of formation of carvacrol by CYP71D180 was higher with *α*-terpinene as substrate. Thus, although *γ*-terpinene is the native substrate for thymol formation, carvacrol can be formed from *α*-terpinene as well.

The presence of two double bonds is also an important prerequisite for the formation of aromatic alcohols though this feature does not always lead to an aromatic product as seen with (-)-*R*-*α*-phellandrene. This cyclohexanoid monoterpene with two double bonds in the ring was converted by CYP71D180 into what is most likely an allylic alcohol, rather than a phenolic product as indicated by the *m/z* of 152 (Figs. 7, 8). The position of the hydroxyl-group within this compound might be the same as found in carveol and carvacrol.

The reaction mechanism of the oregano and thyme CYP71D P450s probably involves an allylic alcohol intermediate

Instead of *p*-cymene as an intermediate for thymol or carvacrol formation by the oregano and thyme CYP71D P450s, the mechanism of these reactions might involve an allylic alcohol intermediate formed from *γ*-terpinene which is then followed by a second oxidation resulting in aromatization (Fig. 12).

An initial allylic hydroxylation is supported by the high sequence similarity (> 73 % at the amino acid level) of the oregano and thyme CYP71D P450s with CYP71D13 and CYP71D18 from mint, which both carry out allylic hydroxylation of the cyclohexanoid monoterpene, limonene. Sequence similarity is especially high in the potential substrate recognition sites (chapter II). Moreover, CYP71D178 and CYP71D180 catalyze the hydroxylation of both

limonene substrates into products with identical regiospecificity to that seen in thymol and carvacrol (Fig. 11). Following the first allylic hydroxylation, a second oxidation may take place to form a ketone. Multi-step oxidations are a characteristic feature of P450s as stated above (Halkier et al., 1995; Bak et al., 1998; Ro et al., 2005). The proposed α,β-unsaturated ketone is inherently unstable and should aromatize to thymol or carvacrol via a keto-enol tautomerism.
The artifactual formation of *p*-cymene might also be explained by details of the reaction mechanism. P450 oxidation of γ-terpinene would be expected to be initiated by abstraction of a hydrogen radical (Meunier et al., 2004; Shaik et al., 2005). If the resulting *γ*-terpinene radical species were released from the active site, it might spontaneously oxidize to *p*-cymene even more readily than *γ*-terpinene itself upon contact with oxygen (Granger et al., 1964).

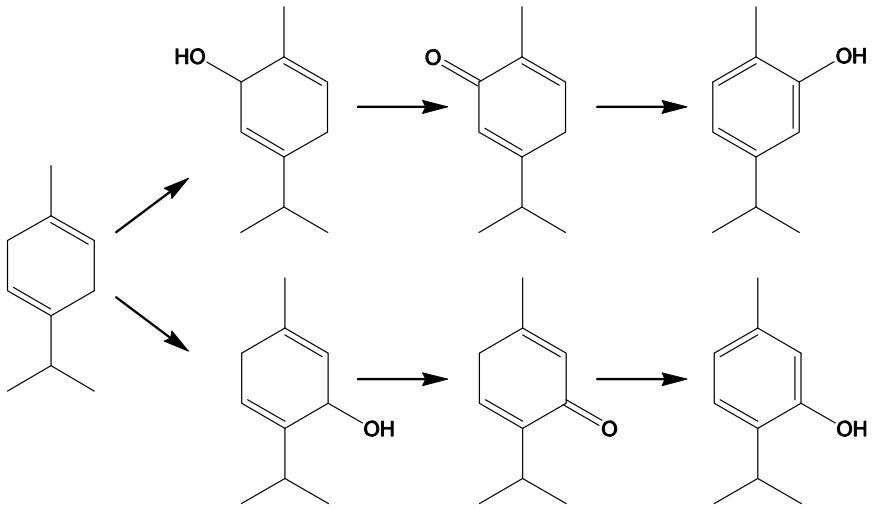


Figure 12 Possible steps in the reaction mechanisms for thymol and carvacrol formation from *γ*-terpinene by CYP71D P450s. The mechanism of thymol and carvacrol formation from *γ*-terpinene might involve an initial formation of an allylic alcohol just like in the most closely-related characterized P450, the mint limonene hydroxylase CYP71D18. This intermediate might be subsequently subject to a second oxidation to convert the alcohol to a ketone. Finally, the aromatic end products, thymol or carvacrol, are formed by a keto-enol tautomerism.

A. thaliana readily glycosylates monoterpene alcohols to detoxify these xenobiotics

In the course of this study, *A. thaliana* was found to convert hydroxylated monoterpenes very efficiently to glycosides. Free monoterpene alcohols were not detected in the plant or as volatiles in the headspace. A similar phenomenon was previously described for petunia over-expressing a linalool synthase from *Clarkia breweri* where the linalool (a monoterpene alcohol) formed was completely bound as glycosides (Lücker et al., 2001). The formation of glycosides from thymol, carvacrol and other monoterpene alcohols might be detoxification reactions to prevent cell damage by these compounds which are known to have strong anti-herbivore and

anti-microbial activities (Isman, 2000; Hummelbrunner and Isman, 2001; Ultee et al., 2002; Sedy and Koschier, 2003; Floris et al., 2004; Braga et al., 2008).

Interestingly, thymol- and carvacrol-glycosides have also been reported to occur in oregano at levels of 80-300 μg g^{-1} fresh weight (Skoula and Harborne, 2002; Stahl-Biskup, 2002). The amounts of free thymol and carvacrol, however, which are stored in the glandular trichomes, are 30 to 400 times higher than these glycosidically bound forms (Stahl-Biskup, 1993, 2002). Observations of the glycoside content in relation to the filling of the glandular trichomes indicated that glycosides are formed when the storage capacity of the subcuticular space is reached and is thought to be a protection mechanism to prevent cell damage, especially membrane destruction, by excess lipophilic volatiles such as phenols or alcohols from destroying membranes (Stahl-Biskup, 1993, 2002).

A. thaliana lacks specialized storage compartments for lipophilic secondary metabolites, like glandular trichomes, resin ducts and secretory cavities, and therefore needs to employ a different strategy to prevent autotoxicity. Glycosylation is such a mechanism which is involved in inactivation or detoxification of xenobiotics and other harmful components (Vogt and Jones, 2000; Meßner et al., 2003; Gachon et al., 2005). The conjugation of plant metabolites to sugar moieties is performed by enzymes from family 1 of glycosyltransferases. They are known to occur in conjunction with cytochrome P450s as part of a detoxification sequence (Pedras et al., 2001). *Arabidopsis* contains more than 100 glycosyltransferases but the *in planta* functions are established for only about 10 % (Yonekura-Sakakibara, 2009). One or more of these enzymes is probably responsible for the glycosylation of the hydroxylated products formed by CYP71D178 and CYP71D180.

5 General Discussion

The phenolic monoterpenes, thymol and carvacrol, are two highly bioactive compounds found in oregano and thyme as well as in a great diversity of other plant species. But despite the fact that thymol and carvacrol are widespread plant secondary compounds with some importance as plant defenses, flavors and pharmaceuticals, not much is known about their biosynthesis.
The pathway of thymol formation has been studied since the early 1960s when Yamazaki and coworkers identified thymol as a terpenoid biosynthetic product despite the fact that it is aromatic (Yamazaki et al., 1963). At the end of the 1970s, experiments were performed in which the radioactively labeled monoterpenes, *γ*-terpinene and *p*-cymene, were fed to thyme (Poulose and Croteau, 1978a). Based on the results, it was postulated that the biosynthesis of thymol starts with *γ*-terpinene as initial monoterpene precursor and proceeds via the aromatic *p*-cymene as an intermediate.

In this thesis, the biosynthetic pathway of thymol and carvacrol was investigated at different levels. First, the genes encoding enzymes of mono- and sesquiterpene biosynthesis were studied in oregano at both the genetic and the biochemical levels (chapter I). Terpene synthases provide products which can be further oxidized by cytochrome P450s. In chapter II five novel cytochrome P450s from oregano, thyme and marjoram were investigated for their role in the formation of thymol and carvacrol. Finally, a new pathway for the formation of thymol and carvacrol is proposed based on the results from over-expression of two cytochrome P450s in the model plant *Arabidopsis thaliana* (chapter III). Together, these chapters provide the first genetic and biochemical evidence for a two step pathway of thymol and carvacrol formation in the Lamiaceae species, oregano and thyme.

The essential oils of oregano, thyme and marjoram consist of a complex and highly variable mixture of over a hundred mono- and sesquiterpenes. About one third of these compounds were extracted from oregano plant lines used for the investigations in chapter I. Within these, thymol, carvacrol, *γ*-terpinene and *p*-cymene, impart the characteristic flavor associated with oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.), while sabinene hydrate represents the flavor of marjoram (*Origanum majorana* L.) (Skoula and Harborne, 2002). To learn more about the biosynthesis of these terpenes and its regulation in these plants, terpene synthases of *O. vulgare* were investigated by isolating members of this gene family, determining their enzymatic activity after heterologous expression in *E. coli* and correlating gene expression with the pattern of terpene accumulation in different *O. vulgare* lines (chapter I).

The terpene synthases of *Origanum vulgare* fall into two clades in which the monoterpene and sesquiterpene synthases are separated from each other (Fig. 1). The *O. vulgare* monoterpene synthases are joined by many monoterpene synthases from other Lamiaceae. However, only two other sesquiterpene synthases from Lamiaceae are found in the sesquiterpene synthase clade, because these are the only two other Lamiaceae sesquiterpene synthases identified to date, both from sweet basil (*Ocimum basilicum*).

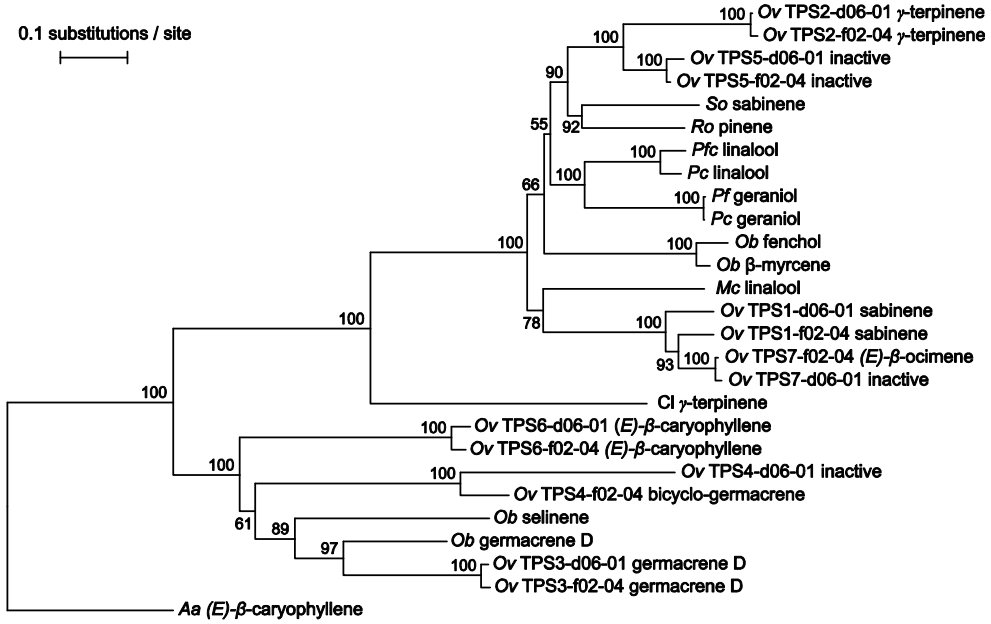


Figure 1 Dendrogram analysis of monoterpene and sesquiterpene synthases from *O. vulgare* and functionally related terpene synthases. The dendrogram was constructed using the neighbour-joining method. The name of the major product of each enzyme is given after the abbreviation of the species (or after the abbreviation of the gene and cultivar designation in case of the *O. vulgare* sequences). *Aa – Artemisia annua:* (*E*)-*β*-caryophyllene synthase; *Cl – Citrus limon*: *γ*-terpinene synthase; *Mc – Mentha citrata*: linalool synthase; *Ob – Ocimum basilicum*: fenchol synthase, myrcene synthase, selinene synthase, germacrene D synthase; *Ov – Origanum vulgare*: TPS1-d06-01, TPS1-f02-04, TPS2-d06-01, TPS2-f02-04, TPS3-d06-01, TPS3-f02-04, TPS4-d06-01, TPS4-f02-04, TPS5-d06-01, TPS5-f02-04, TPS6-d06-01, TPS6-f02-04, TPS7-d06-01, and TPS7-f02-04. *Pf – Perilla frutescens*: geraniol synthase; *Pfc – Perilla frutescens* var. *crispa*: linalool synthase; *Pc – Perilla citriodora*: linalool synthase, geraniol synthase; *Ro – Rosmarinus officinalis*: pinene synthase; *So – Salvia officinalis*: sabinene synthase.

Two trends previously noted for other terpene synthases were also found to be true in oregano. The monoterpene synthases of Lamiaceae share rather high amino acid identities (usually 55-90 %) despite their different catalytic functions. Yet terpene synthases outside this family are much more divergent even though they might have the same catalytic function. For example, a *γ*-terpinene synthase from *Citrus limon* (included in Fig. 1) is clearly separated from *Ov*TPS2, the *γ*-terpinene synthase of *O. vulgare,* with 37 % amino acid identity indicating that these genes are

the result of repeated evolution (Gang, 2005). The phylogenetic analysis also indicates that some terpene synthases appear to result from gene duplication and neofunctionalization or loss of function.

For most of the characterized *O. vulgare* terpene synthases, only a single allele was apparently identified from the EST libraries, suggesting that the other allele of these genes is not transcribed or does not encode an active enzyme. This assumption is supported by the analyses of the selfed lines of *O. vulgare* cultivar f02-04. Since products of *Ov*TPS2 are completely absent in four of the studied lines, only one of the two *Ov*TPS2 alleles is likely to be active in the cultivar f02-04. For another monoterpene synthase, *Ov*TPS5, two alleles were found but none displayed any terpene synthase activity *in vitro*. The presence of catalytically inactive alleles appears to be typical for the terpene synthase gene family in other species as well (Köllner et al., 2004; Köllner et al., 2008). Since *Ov*TPS2 and *Ov*TPS5 are highly identical, it is conceivable that they originate from a duplication event within the genome of *O. vulgare*. Subsequently, *Ov*TPS5 might have lost is function due to a frameshift mutation, but still kept its expression pattern due to preservation of the promoter and other regulatory sequences. This is not consistent with the usual expectation that the regulatory sequences are more quickly altered than the structural gene. For example, the diversification of a pair of terpene synthases of strawberry resulted in altered spatial regulation (cytoplasmatic versus plastidic expression), but did not change the sequence of the structural gene (Aharoni et al., 2004). A high rate of diversification of genes involved in plant secondary metabolism has often been observed and both structural and regulatory changes might contribute to the large variety of terpene patterns found in plants (Köllner et al., 2004; Iijima et al., 2004b).

Terpene synthase gene expression determines terpene composition

The isolated terpene synthase genes of *O. vulgare* appear to play a major role in controlling terpene composition in this species since the transcript levels of individual genes correlate closely with the amounts of the encoded enzyme products found in the essential oil.

The close correlation of *γ*-terpinene synthase expression and terpene composition indicates that transcript regulation of terpene synthase genes is the most important regulatory mechanism controlling terpene composition in *O. vulgare*. The results of RNA hybridization assays and qRT-PCR suggest that this mechanism regulates the activity of other monoterpene synthases as well. In contrast, the low levels of sesquiterpenes in *O. vulgare* essential oil might not be regulated strictly on a transcript level. However, the correlation analysis of the sesquiterpene synthases, *Ovtps*3 and *Ovtps*6, was hampered by the limited variation of sesquiterpene content between the *O. vulgare* lines. Further support for the assertion of transcript level regulation at least for monoterpene formation comes from the combined *in vitro* assays of the heterologously

expressed terpene synthases. When approximately equal amounts of the expressed active terpene synthases were combined in the presence of GPP or FPP as substrates, the blends of monoterpenes or sesquiterpenes produced strongly resembled those of *O. vulgare* terpene blends with the absence of the monoterpene *trans*-sabinene hydrate and thymol. This suggests that terpene synthase expression levels directly control the composition of the essential oil, and provide no indication for operation of any further regulatory mechanisms like compartmentation or metabolite channeling at the site of *O. vulgare* terpene biosynthesis. However, control of total terpene yield may result from processes at other levels of organization. The *O. vulgare* cultivars used in chapter I, f02-04 and d01-06, showed major differences in the quantity of terpenes produced, with the quantity of cultivar f02-04 being approximately twice that of d01-06. This difference is likely caused by the fact that the leaf surface of f02-04 has a higher density of glandular trichomes, the sites of synthesis and storage of the monoterpenes and sesquiterpenes (Gershenzon et al., 1989; Turner et al., 1999).

Thymol and carvacrol formation depends on the presence of γ-terpinene in oregano and thyme

The six active terpene synthases characterized in chapter I produce the majority of terpenes found in *O. vulgare*. One major terpene product not formed by these enzymes is *trans*-sabinene hydrate. Despite the reports of a sabinene hydrate synthase activity in sweet marjoram (*Origanum majorana* L., previously: *Majorana hortensis* Moench*.)* (Hallahan and Croteau, 1988, 1989), no enzyme responsible for its formation from GPP could be identified in the present study.

Other compounds not directly formed by the characterized terpene synthases are the aromatic monoterpene alcohols, thymol and carvacrol. As mentioned above, they were predicted to be synthesized from γ-terpinene via *p*-cymene (Poulose and Croteau, 1978a; Poulose and Croteau, 1978b). Analysis of the terpene content of the oregano lines supported this hypothesis as no thymol or carvacrol are found in lines that lack γ-terpinene and *p*-cymene (Fig. 2). Therefore, the enzyme responsible for γ-terpinene formation *in vitro*, *Ov*TPS2, is likely to be a major terpene synthase activity in oregano *in vivo*. Between 25.2 and 48.4 % of the total terpene content of this species consists of compounds that are products of this enzyme *in vitro*. The role of *Ov*TPS2 in γ-terpinene formation *in vivo* is also supported by the fact that the *Ovtps2* gene was only present in lines that produced γ-terpinene (Fig. 2). In addition, the biochemical properties of this enzyme closely resemble those of a γ-terpinene synthase extracted from thyme leaves (Poulose and Croteau, 1978b).

The importance of γ-terpinene supply for the formation of thymol and carvacrol was confirmed in chapter II. Analysis of oregano, thyme and marjoram plant lines shows that an ortholog of *Ovtps*2 seems to be responsible for γ-terpinene formation in marjoram, another member of the genus *Origanum*. In thyme, a different monoterpene synthase gene, *Tvtps*1, was characterized as a γ-terpinene synthase (Julia Asbach, unpublished data), which is similar to another gene isolated from oregano (*Ovtps*8). *Tvtps*1 and *Ovtps*8 are 99.5 % identical at the nucleotide and the amino acid level, but both are only 90 % identical to *Ovtps*2. Transcript levels for *Tvtps*1 have been measured in different thyme chemotypes and found to strongly correlate with γ-terpinene content (Julia Asbach, unpublished data). Whether or not *Ovtps*8 has any importance in γ-terpinene formation in oregano needs to be tested in the future. Since no ESTs for this gene were found in a cDNA library made from oregano cultivar f02-04, *Ovtps*8 might not be actively involved in γ-terpinene biosynthesis in cultivars f02-04 and d06-01(chapter I).

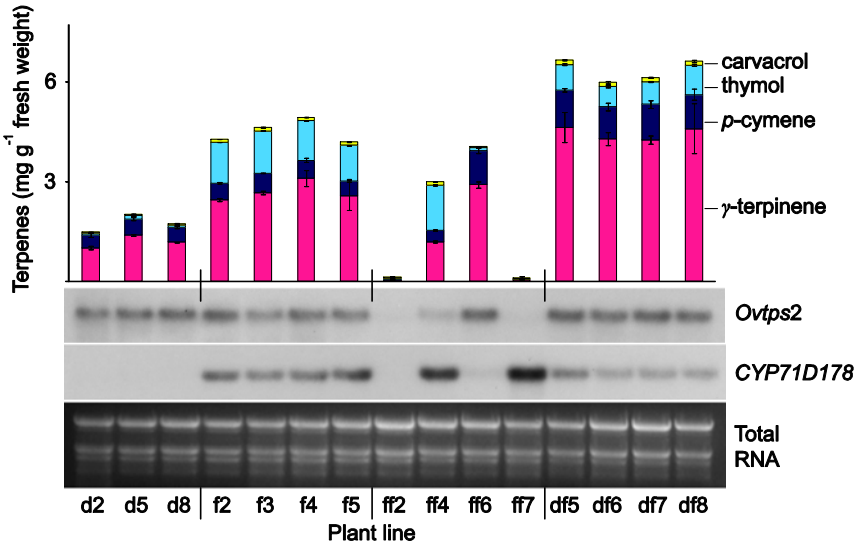


Figure 2 Expression analysis of *Ovtps*2 and *CYP71D178 in planta* compared to essential oil contents of various oregano plant lines. The amounts for γ-terpinene, *p*-cymene, thymol and carvacrol are shown for 15 oregano lines: three clonal lines of the cultivar d06-01 (d2, d5, d8), four clonal lines of cultivar f02-04 (f2, f3, f4, f5), four lines from a selfing of the cultivar f02-04 (ff2, ff4, ff6, ff7), and four lines from a cross of both cultivars (df5, df6, df7, df8). Transcript levels of the terpene synthase *Ovtps*2 and *CYP71D178* in leaves of *O. vulgare* were measured by RNA hybridization analyses. The bottom panel shows an ethidium-bromide-stained agarose gel with total RNA as control for equal RNA loading.

Cytochrome P450s are involved in thymol and carvacrol formation in oregano and thyme

From the beginning of the investigations, the biosynthesis of thymol and carvacrol from γ-terpinene was thought likely to be catalyzed by the action of one or more cytochrome P450

oxidases, carrying out a hydroxylation similar to that described for other monoterpenes like (-)-*S*-limonene in menthol biosynthesis in *Mentha* sp. (Lupien et al., 1999; Haudenschild et al., 2000). Several studies have reported on the roles of cytochrome P450s in monoterpene metabolism. Geraniol is hydroxylated by a geraniol-10-hydroxylase in *Catharanthus roseus* (Meijer et al., 1993), and sabinene hydroxylase catalyzes the formation of sabinol in sage (Karp et al., 1987). Several examples are known of limonene hydroxylases which catalyze hydroxylations at different carbon atoms in this cyclic monoterpene. In spearmint, a 6-hydroxylase forms (-)-*trans*-carveol from (-)-*S*-limonene, whereas in peppermint this substrate is hydroxylated by a 3-hydroxylase to form (-)-*trans*-isopiperitenol, an intermediate in the menthol biosynthesis pathway (Karp et al., 1990; Lupien et al., 1995). Recently, a cytochrome P450 candidate for the hydroxylation of (-)-*S*-limonene was reported for *Perilla* which is able to catalyze the hydroxylation at three different positions to either form (-)-*trans*-isopiperitenol, (-)-*trans*-carveol or (-)-perillyl alcohol (Mau et al., 2010). In caraway another limonene-6-hydroxylase catalyzes the reaction of the other enantiomer, (+)-*R*-limonene, into (+)-*trans*-carveol (Bouwmeester et al., 1998).

An oregano gene sequence (*CYP71D178*) with similarity to a known monoterpene hydroxylase from mint (*CYP71D18*) was first suggested to be involved in thymol and carvacrol biosynthesis in oregano based on RNA hybridization and relative qRT-PCR (chapter II). The gene expression of *CYP71D178* in oregano was well-correlated with thymol content in different sets of oregano lines (Fig. 2). Oregano line ff7, which showed a high transcript accumulation but only traces of thymol and carvacrol, lacked *γ*-terpinene underlying the importance of this compound as a precursor of phenolic monoterpene formation.

Nevertheless, it remained unclear if only one cytochrome P450 enzyme would be involved in thymol or carvacrol formation since conversion of *γ*-terpinene to thymol or carvacrol involves two formal oxidation steps. Hence, additional P450 sequences were sought from oregano, thyme and marjoram cultivars using primers from *CYP71D178*. Eleven sequences were found classified into five different cytochrome P450s that gave a clear differentiation into two large subgroups a) and b) (Fig. 3).

The two subgroups of P450s could conceivably represent two different catalytic activities, one for thymol and one for carvacrol. This hypothesis was first tested by absolute qRT-PCR on a group of oregano plant lines also used in chapter I and three oregano, thyme and marjoram cultivars with varying thymol and carvacrol content. Although qRT-PCR was partly hampered by the high nucleotide sequence identity between the different P450s, a clear differentiation of expression pattern was observed for three of the five cytochrome genes. High copy numbers of *CYP71D181* correlated with the very high carvacrol content in oregano cultivar ‘Ct’, and the expression levels of *CYP71D179* and *CYP71D182* correlated with thymol content. Plant lines

without transcript of any of the described P450s had a complete absence of thymol and carvacrol even if γ-terpinene was present. These correlations also suggested that expression of these CYP71D P450s is regulated at the transcript level in these plants (chapter II) similar to what was found for the terpene synthases of oregano (chapter I).

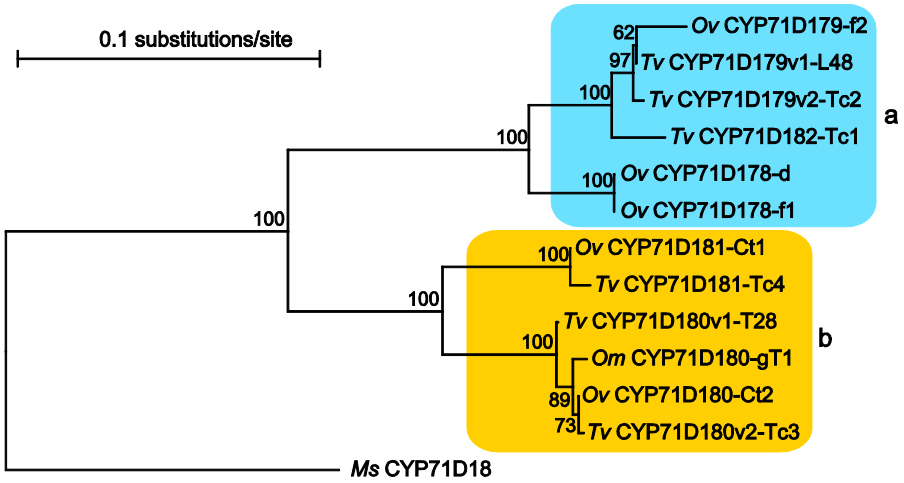


Figure 3 Dendrogram analysis of amino acid sequences of cytochrome P450s from oregano, thyme and marjoram compared to one member of the same cytochrome P450 subfamily CYP71D from mint, CYP71D18 which encodes a previously characterized monoterpene hydroxylase (Colby et al., 1993; Lupien et al., 1999; Haudenschild et al., 2000). Two subgroups are shown. Subgroup (**a**) contains CYP71D178, CYP71D179 and CYP71D182. Subgroup (**b**) consists of CYP71D180 and CYP71D181. The dendrogram was constructed using the neighbour-joining method. The CYP name of each cytochrome P450 is given after the abbreviation of the species of origin and followed by the internal numbering within one cultivar: *Ov* – *Origanum vulgare*, *Tv* – *Thymus vulgaris*, *Om* – *Origanum majorana*; d – oregano cv. d06-01, f –oregano cv. f02-04, Ct – oregano cv. 'Ct', Tc – thyme cv. 'Tc', gT – marjoram cv. 'gT'.

Although five P450s were expressed in yeast, microsomal preparations gave enzymatically active protein for only three, *CYP71D178*, *CYP71D180*v1 and *CYP71D181*. Despite the high substrate specificity reported for plant biosynthetic P450s (Schuler, 1996), all three enzymes accepted a variety of different monoterpenes as substrates, including γ-terpinene, α-terpinene, (-)-*R*-α-phellandrene, (+)-*R*-limonene and (-)-*S*-limonene, all cyclohexanoid monoterpenes with two double bonds. Given the correlation of three P450s with thymol and carvacrol accumulation and the proposed role of γ-terpinene in the pathway to thymol and carvacrol, γ-terpinene was expected to be the natural substrate. All three active P450 enzymes converted γ-terpinene to one of these phenolic monoterpenes in small amounts: CYP71D180 and CYP71D181 formed carvacrol and CYP71D178 formed both thymol and carvacrol. But, the main product in all cases was *p*-cymene. This aromatic monoterpene was suggested to be an intermediate of thymol

biosynthesis, between γ-terpinene and thymol (Poulose and Croteau, 1978a). However, none of the enzymes converted *p*-cymene to thymol, carvacrol or any other product *in vitro*.

Nevertheless, the three tested P450s were able to convert the limonene enantiomers, (+)-*R*- and (-)-*S*-limonene, into allylic alcohols with similar regiospecificity of the hydroxyl groups as in thymol and carvacrol. These substrates were tested because of the remarkably high amino acid identity of the oregano and thyme CYP71D P450s to the mint enzymes, CYP71D13 and CYP71D18. The oregano and thyme P450s share many structural and biochemical characteristics with the limonene hydroxylases of mint, which are the most closely-related P450s on the basis of amino acid similarity that have been characterized to date. The shared characteristics include the potential substrate recognition sites SRS1-SRS6. These sites also differ among CYP71D178 through CYP71D182 which might explain some of the biochemical properties observed *in vitro*. In SRS5, it was reported that a single amino acid substitution (F361I) converts the regiospecificity of CYP71D18 from a C6- to a C3-hydroxylase (Schalk and Croteau, 2000). At the corresponding positions, CYP71D178 through D182 all bear a phenylalanine (F) like the mint C6-hydroxylase so this position cannot be responsible for regiospecific differences in catalysis. However, only two amino acids downstream of this position, there is a marked difference among the enzymes with CYP71D178 containing an isoleucine residue, CYP71D179 and D182 containing a methionine residue, and CYP71D180 and D181 (as well as the mint limonene hydroxylases, CYP71D13 and CYP71D18) containing a leucine at this position.

At this point it remained unclear whether there are one or two separate P450-catalyzed steps on the pathway from γ-terpinene to thymol and carvacrol. The enzymes investigated *in vitro* could perform the first step to form *p*-cymene *in planta*, and one of the not yet characterized CYP71D P450s could be responsible for the second step from *p*-cymene to thymol or carvacrol. The co-expression of two or more CYP71D P450s in several of the investigated plant lines supported such a two-enzyme scenario (chapter II). However, based on the conversion of (+)-*R*-limonene to either a C3 oxidation product (CYP71D178) or a C6 oxidation product (CYP71D180 and CYP71D181) it was conceivable that thymol and carvacrol formation could also involve an allylic intermediate formed from γ-terpinene. In this case, aromatization would constitute the second step. This suggested that *p*-cymene would be an artifact, possibly due to not optimal *in vitro* assay conditions.

Thymol and carvacrol formation by transgenic *Arabidopsis thaliana* over-expressing *CYP71D178* or *CYP71D180*

To circumvent the difficulties inherent in carrying out *in vitro* assays with these P450s under non-natural conditions, a different approach was necessary. Therefore, two P450s, *CYP71D178* and *CYP71D180*v1, were transformed into *Arabidopsis thaliana* Col-0 (chapter III). First, the γ-terpinene synthase, *Ov*TPS2, was transformed into *Arabidopsis* to provide the potential substrate for the P450s. However, this approach was stopped since the emission of γ-terpinene was extremely low and only present in flowers. Co-expression of both the γ-terpinene synthase and *CYP71D178* resulted in the release of low levels of *p*-cymene only. The supply of substrate was enhanced by direct feeding of different monoterpenes in high concentrations via the surrounding air to transgenic *A. thaliana* over-expressing *CYP71D178* or *CYP71D180*v1 kept in closed glass vessels. By this experimental setup, the hypothesis of a direct conversion of γ-terpinene into thymol or carvacrol by oregano and thyme CYP71D P450s was tested.

Similar to the *in vitro* results from chapter II, *CYP71D178* over-expressers formed both thymol and carvacrol while *CYP71D180*v1 over-expressers formed only carvacrol from γ-terpinene (chapter III). *p*-Cymene was found as the major product and was also formed by wild type and vector control *Arabidopsis* plants. Given these results and the previously described spontaneous formation upon contact with oxygen (Granger et al., 1964), it was concluded that the formation of *p*-cymene *in vitro* and *in vivo* is very likely an artifact of enzyme catalysis in a heterologous system. The fact that CYP71D178 and 180 convert γ-terpinene to thymol and carvacrol, a process requiring two formal oxidations, is not unusual for a member of the P450 family. Catalysis of multi-step oxidations is well known for a number of cytochrome P450s (Halkier et al., 1995; Bak et al., 1998; Ro et al., 2005).

p-Cymene is not an intermediate in thymol and carvacrol formation by oregano and thyme CYP71D P450s

The aromatic monoterpene *p*-cymene was originally suggested as intermediate in the pathway of thymol and carvacrol biosynthesis proposed for thyme over thirty years ago (Poulose and Croteau, 1978a). The initial substrate, γ-terpinene, was predicted to be oxidized to *p*-cymene which in a second step is hydroxylated to form either thymol or carvacrol. This was a likely pathway since cytochrome P450s are often responsible for the hydroxylation of aromatic rings (Jerina and Daly, 1974), which are present in many pathways for plant secondary products such as furanocoumarins, anthocyanins, flavonoids and many more (Schuler, 1996).
In the work reported here, no proof was found for the intermediacy of *p*-cymene in the conversion of γ-terpinene to thymol and carvacrol. In the feeding experiments described in this

chapter with the *A. thaliana* lines over-expressing *CYP71D178* and *CYP71D180*v1 and the *in vitro* assays conducted with CYP71D178, 180v1 and 181 protein expressed in yeast (chapter II), it was found that *p*-cymene itself was not converted to form thymol and carvacrol in any amounts above background levels. Instead, *p*-cymene formed two other products at much higher rates, cuminol and *p*-cymene-8-ol, which carry hydroxyl groups at carbon positions C7 and C8 outside the aromatic ring. These two derivatives were most probably formed by enzymes from *Arabidopsis thaliana* itself, possibly by one of the 246 P450s found in the *Arabidopsis* genome (Paquette et al., 2000; Werck-Reichhart et al., 2002; Schuler and Werck-Reichhart, 2003). Some of these P450s might be capable of catalyzing the hydroxylation of the aromatic *p*-cymene to thymol possibly by an NIH shift mechanism (Guroff et al., 1967).

Thyme and oregano CYP71D P450s have narrow product specificity but broad substrate specificity

While *CYP71D178* over-expressing *A. thaliana* plants formed both thymol and carvacrol, *CYP71D180*v1 over-expressers produced only carvacrol from *γ*-terpinene. Such differences in the regiospecificity of hydroxylation reactions are well known for cytochrome P450s (Schuler, 1996) and have been especially well described for several P450s hydroxylating limonene in mint, caraway and *Perilla* (Karp et al., 1990; Bouwmeester et al., 1998; Mau et al., 2010).

The oregano and thyme P450s, CYP71D178, CYP71D180v1 and CYP71D181, also showed such a difference in the regiospecificity of the hydroxylation from (-)-*S*- and (+)-*R*-limonene *in vitro* (chapter II). In chapter III, with *A. thaliana* over-expression lines, we found the same pattern. CYP71D178 over-expressing plant lines form mainly (-)-*trans*-carveol from (-)-*S*-limonene and (+)-*trans*-isopiperitenol from (+)-*R*-limonene whereas CYP71D180v1lines form only carveols from either substrate. The position of the hydroxyl-group in carveol at carbon C6 is identical to C2 in carvacrol whereas the C3-hydroxylation in isopiperitenol is identical in its position to the hydroxyl group in thymol (Fig. 4).

As indicated above, the reason for these differences in the hydroxylation position might be related to differences found in the substrate recognition sites described in chapter II. Whether these differences, especially in SRS5 are responsible for the formation of either thymol or carvacrol and isopiperitenol or carveol needs further investigations, e.g. site directed mutagenesis combined with structural modeling of these P450s. *In vitro* CYP71D180 and CYP71D181 had shown identical hydroxylation patterns with both limonene substrates. Whether CYP71D179 and CYP71D182 show similar or different hydroxylation abilities compared to CYP71D178 is currently under investigation.

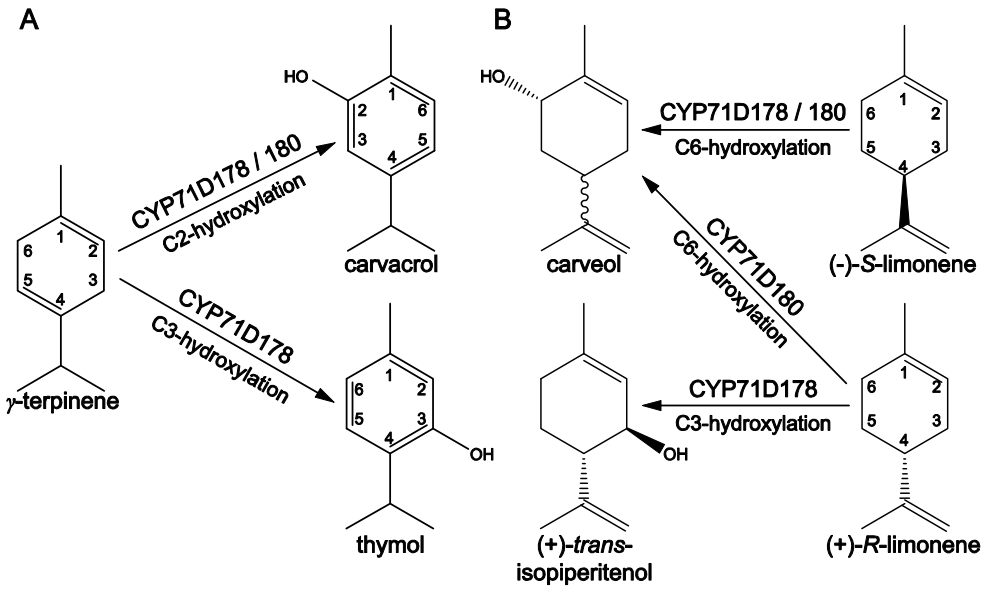


Figure 4 Summary of hydroxylated products formed from *γ*-terpinene and limonene fed to *A. thaliana* lines transformed with CYP71D genes. (**A**) Products formed from *γ*-terpinene: CYP71D178 catalyzes C2- and C3-hydroxylations to carvacrol or thymol while CYP71D180v1 forms only carvacrol. (**B**) Products formed from (+)-*R*- and (-)-*S*-limonene: CYP71D178 catalyzes C3- or C6-hydroxylations depending on the limonene enantiomer while CYP71D180v1 catalyzes only C6-hydroxylations from both enantiomers.

The narrow product specificity of hydroxylation at only two positions in the cyclohexanoid ring is paired with broad substrate specificity. The three CYP71D P450s studied with yeast-expressed proteins and in *A. thaliana* accepted five different cyclohexanoid monoterpenes as substrates. All accepted substrates contain two double bonds, at least one of which is within the cyclohexanoid ring. The position of the double bonds within the ring had an effect on the product. While *γ*-terpinene was hydroxylated at either C2 or C3 by CYP71D178, *α*-terpinene was exclusively hydroxylated at C2 by both enzymes. Moreover, the rate of formation of carvacrol by CYP71D180 was higher with *α*-terpinene as substrate. Thus, although *γ*-terpinene is the native substrate for thymol formation, carvacrol can be formed from *α*-terpinene as well.

Arabidopsis readily glycosylates monoterpene alcohols as detoxification reactions

A. thaliana was found to convert hydroxylated monoterpenes very efficiently to glycosides. Only low amounts of free monoterpene alcohols were detected as volatiles in the headspace. A similar phenomenon was previously described for petunia over-expressing a linalool synthase from *Clarkia breweri* where the linalool (a monoterpene alcohol) formed was completely bound as glycosides (Lücker et al., 2001). The formation of glycosides from thymol, carvacrol and other monoterpene alcohols might be detoxification reactions to prevent cell damage by these compounds which are known to have strong anti-herbivore and anti-microbial activities (Isman,

2000; Hummelbrunner and Isman, 2001; Ultee et al., 2002; Sedy and Koschier, 2003; Floris et al., 2004; Braga et al., 2008).

Interestingly, thymol- and carvacrol-glycosides have also been reported to occur in oregano at levels of 80-300 µg g^{-1} fresh weight (Skoula and Harborne, 2002; Stahl-Biskup, 2002). The amounts of free thymol and carvacrol, however, which are stored in the glandular trichomes, are 30 to 400 times higher than these glycosidically bound forms (Stahl-Biskup, 1993, 2002). Observations of the glycoside content in relation to the filling of the glandular trichomes indicated that glycosides are formed when the storage capacity of the subcuticular space is reached and is thought to be a protection mechanism to prevent cell damage, especially membrane destruction, by excess lipophilic volatiles such as phenols or alcohols from destroying membranes (Stahl-Biskup, 1993, 2002).

A. thaliana lacks specialized storage compartments for lipophilic secondary metabolites, like glandular trichomes, resin ducts and secretory cavities, and therefore needs to employ a different strategy to prevent autotoxicity. Glycosylation is such a mechanism which is involved in inactivation or detoxification of xenobiotics and other harmful components (Vogt and Jones, 2000; Meßner et al., 2003; Gachon et al., 2005). The conjugation of plant metabolites to sugar moieties is performed by family 1 glycosyltransferases. They are known to occur in conjunction with cytochrome P450s as part of a detoxification sequence (Pedras et al., 2001). *Arabidopsis* contains more than 100 glycosyltransferases but the *in planta* functions are established for only about 10 % (Yonekura-Sakakibara, 2009). One or more of these enzymes is probably responsible for the glycosylation of the hydroxylated products formed by CYP71D178 and CYP71D180.

The reaction of oregano and thyme CYP71D P450s probably involves an allylic alcohol intermediate

The data presented in chapters II and III clearly show that instead of *p*-cymene serving as an intermediate for thymol or carvacrol formation by the oregano and thyme CYP71D P450s, the mechanism of these reactions might involve an allylic alcohol intermediate formed from *γ*-terpinene which is then followed by a second oxidation resulting in aromatization.

An initial allylic hydroxylation is supported by the high sequence similarity (> 73 % at the amino acid level) of the oregano and thyme CYP71D P450s with CYP71D13 and CYP71D18 from mint, which both carry out allylic hydroxylation of the cyclohexanoid monoterpene, limonene. Sequence similarity is especially high in the potential substrate recognition sites (chapter II). Moreover, CYP71D13 and 18 catalyze the hydroxylation of both limonene substrates into products with identical regiospecificity to that seen in thymol and carvacrol. Following the first allylic hydroxylation, a second oxidation may take place to forma ketone.

Multi-step oxidations are a characteristic feature of P450s as stated above (Halkier et al., 1995; Bak et al., 1998; Ro et al., 2005). The proposed α,β-unsaturated ketone is inherently unstable and should aromatize to thymol or carvacrol via a keto-enol tautomerism.

The artifactual formation of *p*-cymene might also be explained by details of the reaction mechanism. P450 oxidation of γ-terpinene would be expected to be initiated by abstraction of a hydrogen radical (Meunier et al., 2004; Shaik et al., 2005). If the resulting γ-terpinene radical species were released from the active site, it might spontaneously oxidize to *p*-cymene even more readily than *p*-cymene itself upon contact with oxygen (Granger et al., 1964).

A pathway for thymol and carvacrol formation in oregano and thyme

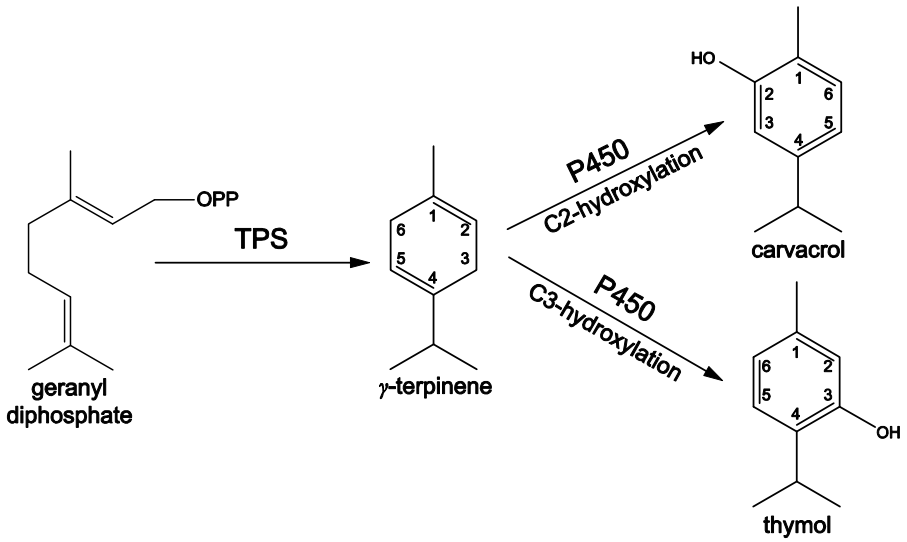


Figure 5 Proposed pathway for thymol and carvacrol formation in oregano and thyme.

Based on the data presented in the three chapters, only two enzymes seem to be necessary to catalyze the reactions from geranyl diphosphate (GPP) to thymol and carvacrol in oregano and thyme. The first step from GPP to γ-terpinene is performed by monoterpenes synthases (TPS). γ-Terpinene is then transformed into thymol and carvacrol by the action of single cytochrome P450s. The two subgroups found in the amino acid alignment of the five CYP71D P450s from oregano and thyme seem to represent the functional difference. Therefore, CYP71D178, CYP71D179 and CYP71D182 are proposed to be thymol synthases while CYP71D180 and CYP71D182 are proposed to be carvacrol synthases.

Outlook

Thymol and / or carvacrol are found in several plant species from at least three different plant families often together with *γ*-terpinene and *p*-cymene. This indicates a similar mechanism of thymol and carvacrol formation, possibly via cytochrome P450s of the CYP71D subfamily. Nevertheless, it is conceivable that different pathways have evolved, especially in plant species outside the Lamiaceae. Elsewhere in plant metabolism, different pathways or different enzymes are sometimes involved in the biosynthesis of the same product.

The oregano and thyme CYP71D P450s characterized seem to have clear differences in product spectra between thymol and carvacrol. However, this point is still equivocal, since CYP71D178 seems to produce both compounds. The difference in product spectra might be related to amino acid difference found in the substrate recognition sites. This might be tested either *in vitro* by site directed mutagenesis or by structural modeling and substrate fitting.

Another aspect not discussed in this thesis is the role of the native cytochrome P450 reductase for the reaction mechanism of thymol and carvacrol formation. Almost all P450s require such a reductase for electron transfer to the active site. Though most plant P450s work very well with other plant P450 reductases the efficiency of the electron transfer may still depend on the CPR homolog present, and thus different CPRs may differentially influence cytochrome P450 performance (Hasemann et al., 1995; Jensen and Møller, 2010). This possibility could be investigated by isolation of the CPR from thyme or oregano and its utilization in functional expression.

6 Summary

Plant secondary compounds are of great importance not only to the plant as defense but also for pharmaceutical and medicinal purposes. Understanding the mechanisms underlying the formation and regulation of plant secondary compounds is essential to further investigate their roles in plant defense and develop new strategies to make these compounds more available for pharmaceutical and nutritional usage. Thymol and carvacrol are two aromatic monoterpenes often found in the essential oil of two culinary herbs, oregano (*Origanum vulgare* L.) and thyme (*Thymus vulgaris* L.) but also in a great diversity of other plant species. These compounds have a broad range of biological activities acting as antimicrobial compounds, insecticides, anti-oxidants and pharmaceutical agents. A pathway for the biosynthesis of thymol from the monoterpene *γ*-terpinene via an intermediate *p*-cymene was proposed in the late 1970s which has never been validated.

The research conducted for this thesis led to the elucidation of a new pathway to thymol and carvacrol and generated knowledge about the properties of the enzymes involved.
Terpene synthases catalyze the formation of basic terpene skeletons from acyclic precursors. The genes coding for terpene synthases were investigated in two different oregano cultivars. Seven terpene synthase genes, *Ovtps*1 through *Ovtps*7, were isolated. Heterologous expression of these genes in *E. coli* resulted in six active terpene synthases which were found to form multiple mono- or sesquiterpenes. Together these terpene synthases are responsible for the direct production of the majority of terpenes found in *O. vulgare* essential oil. The isolated monoterpene synthase genes of *O. vulgare* appear to play a major role in controlling terpene composition in this species since the transcript levels of individual genes correlate closely with the amounts of the encoded enzyme products found in the essential oil.
The enzymes responsible for *γ*-terpinene formation *in vivo* are likely to be the major terpene synthase activities in oregano and thyme. These enzymes provide the first intermediate in thymol and carvacrol biosynthesis. Only plant lines expressing the encoding genes contain *γ*-terpinene and related compounds such as *p*-cymene, thymol and carvacrol.

Cytochrome P450 enzymes (P450s) are known to catalyze a number of oxidations of terpene metabolism and were likely to be involved in the reactions from *γ*-terpinene to thymol. Eleven cytochrome P450 gene sequences were isolated from oregano, thyme and marjoram that were assigned to five gene names, *CYP71D178* through *CYP71D182*. The transcript levels of most of these genes are well-correlated with the occurrence of thymol and carvacrol. Heterologous expression of *CYP71D178*, *CYP71D180* and *CYP71D181* in yeast resulted in active proteins catalyzing the formation of *p*-cymene, thymol and carvacrol from *γ*-terpinene. *p*-Cymene was

not accepted as a substrate *in vitro*. This suggested that γ-terpinene is directly converted to thymol and carvacrol *in vivo* with *p*-cymene as a side product. The properties and sequence motifs of these P450s are similar to those of well-characterized limonene hydroxylases isolated from mint, CYP71D13 and CYP71D18. Moreover, the oregano and thyme CYP71D P450s hydroxylated limonene with similar regiospecificity as found in thymol and carvacrol.

In order to circumvent the difficulties inherent in carrying out *in vitro* assays with these P450s under non-natural conditions, a different approach was necessary. Therefore, two of the oregano and thyme P450s were transformed into *Arabidopsis thaliana*. Transgenic *A. thaliana* plants over-expressing *CYP71D178* or *CYP71D180*v1 were fed with different monoterpenes as substrates. Thymol and carvacrol were formed by these transgenic plants from γ-terpinene while *p*-cymene was not accepted as a substrate by the introduced CYP71D P450s. The majority of the hydroxylated products formed by transgenic *Arabidopsis* plants were not released as free volatiles but bound as glycosides. This might be a detoxification mechanism to prevent cell damage.
Further experiments with structurally similar monoterpenes such as α-terpinene, (-)-*R*-α-phellandrene, (-)-*S*-limonene and (+)-*R*-limonene revealed that these P450s have broad substrate specificities paired with narrow product specificity. Hydroxylations catalyzed by these oregano and thyme P450s occur only at two distinct carbon positions within the cyclohexanoid ring, either at C3 or at C2 (C6).

In conclusion, it is proposed that the formation of thymol and carvacrol is catalyzed by single P450s directly from γ-terpinene via a two-step oxidation, whereas *p*-cymene is a side product resulting from premature release of the substrate from the active site. The mechanism of these reactions might involve an allylic alcohol intermediate formed from γ-terpinene which is then followed by a second oxidation resulting in aromatization.
A pathway for thymol and carvacrol is proposed to start with the formation of γ-terpinene by a monoterpene synthase. The second step is catalyzed by cytochrome P450s in a two-step oxidation. CYP71D178, CYP71D179 and CYP71D82 are proposed to be thymol synthases while CYP71D180 and CYP71D181 are proposed to be carvacrol synthases.

7 Zusammenfassung

Sekundäre Pflanzeninhaltsstoffe (auch pflanzliche Naturstoffe genannt) sind einerseits wichtig für Pflanzen zur Abwehr von Fraßschäden, andererseits werden sie häufig auch im pharmazeutischen oder medizinischen Bereich eingesetzt. Einblicke in die Mechanismen der Biosynthese und Regulation dieser Naturstoffe sind unverzichtbar, um neue Strategien zur Nutzbarmachung dieser Naturstoffe zu entwickeln und ihre eigentliche Rolle in der Pflanzenabwehr zu verstehen. Thymol und Carvacrol sind zwei Naturstoffe, die eine hohe Bioaktivität besitzen, z.B. als antimikrobielle Agentien, als Insektizid, Antioxidans oder als pharmazeutisches Mittel. Diese beiden phenolischen Monoterpene sind typische Inhaltsstoffe der Ätherischen Öle in Oregano (*Origanum vulgare* L.) und Thymian (*Thymus vulgaris* L.). Bereits in den späten 1970er Jahren wurde ein möglicher Weg für die Biosynthese von Thymol beschrieben. Es wurde postuliert, dass aus γ-Terpinen, einem Monoterpen, über ein aromatisches Intermediat (*p*-Cymen) Thymol gebildet wird.

Die Zielstellung der vorliegenden Arbeit lag darin, den genauen Biosyntheseweg von Thymol und Carvacrol in den beiden Pflanzen, Oregano und Thymian, zu entschlüsseln.

Terpensynthasen sind die Enzyme, die für die Biosynthese von Mono- und Sesquiterpenen verantwortlich sind. Daher wurden zuerst die Gene verschiedener Terpensynthasen in zwei Kultursorten von Oregano untersucht. Dabei wurden sieben Terpensynthasegene (*Ovtps*1 bis *Ovtps*7) isoliert, deren heterologe Expression in *E. coli* resultierte in sechs aktiven Terpensynthasen, die jeweils mehrere verschiedene Mono- oder Sesquiterpene bilden. Zusammen sind diese Enzyme für die Biosynthese eines Großteils der Terpene im ätherischen Öl von Oregano verantwortlich. Die Regulation scheint sich dabei vorwiegend auf der Ebene der Gene abzuspielen. Die Transkriptmengen der einzelnen Monoterpensynthasegene korrelieren sehr genau mit den Produkten der einzelnen Terpensynthasen im ätherischen Öl. Besonders wichtig sind die Monoterpensynthasen aus Oregano und Thymian, die γ-Terpinen bilden, da dieses Monoterpen die Vorstufe für die Biosynthese von Thymol und Carvacrol darstellt.

Die nächsten Schritte im ursprünglich postulierten Biosyntheseweg für Thymol beinhalten Reaktionen, die häufig durch Enzyme aus der Familie der sogenannten Cytochrom P450 (P450) katalysiert werden. Aus verschiedenen Oregano, Thymian und Majoran Kultursorten konnten insgesamt elf Gene von Cytochrom P450 Enzymen isoliert werden, die sich auf fünf P450 verteilen: *CYP71D178* bis *CYP71D182*. Die Expression der meisten dieser Gene korreliert deutlich mit dem Vorkommen von Thymol und Carvacrol in den entsprechenden Pflanzen. Drei der isolierten Gene (*CYP71D178*, *CYP71D180* und *CYP71D181*) wurden in *S. cerevisiae* heterolog exprimiert. Die resultierenden Enzyme setzen γ-Terpinen in *p*-Cymen, Thymol und

Carvacrol um. Da *p*-Cymen nicht umgewandelt werden konnte wurde angenommen, dass *p*-Cymen ein Nebenprodukt sein könnte.

Die Proteinsequenzen der untersuchten P450 Enzyme weisen große Ähnlichkeit mit anderen P450 Enzymen auf, besonders mit zwei Limonene-Hydroxylasen aus Minze (CYP71D13 und CYP71D18). Auch die enzymatischen Eigenschaften ähneln denen aus der Minze isolierten Enzyme.

Um Probleme mit der Bildung von Artefakten *in vitro* zu umgehen, wurden zwei der P450 Gene in *Arabidopsis thaliana* transformiert. Die daraus resultierenden transgenen *A. thaliana* Pflanzen, die entweder *CYP71D178* oder *CYP71D180* überexprimieren, wurden mit verschiedenen Monoterpenen gefüttert. Nach Zugabe von *γ*-Terpinen bildeten die transgenen Pflanzen Thymol und Carvacrol. *p*-Cymen konnte auch von den in *A. thaliana* überexprimierten CYP71D P450 Enzymen nicht als Substrat verwendet werden, um Thymol oder Carvacrol zu synthetisieren. Der überwiegende Teil der in den transgenen Pflanzen gebildeten hydroxylierten Produkte wurde nicht als flüchtige Terpene abgegeben, sondern in Glykosiden gebunden. Der zugrunde liegende Mechanismus könnte eine Entgiftungsreaktion der Pflanze sein, um Schäden an den Zellen zu verhindern.

In weiteren Experimenten wurden strukturell ähnliche Monoterpene getestet, wie *α*-Terpinen, (-)-*R*-*α*-Phellandren, (+)-*R*-Limonen und (-)-*S*-Limonen. Diese Versuche zeigten, dass die getesteten Cytochrom P450 Enzyme eine relativ große Menge an verschiedenen Monoterpenen als Substrat verwenden können. Die entstandenen Produkte sind allerdings nur an zwei verschiedenen Positionen des Cyclohexan-Rings hydroxyliert, entweder am Kohlenstoffatom C3 oder C2 (C6). Daraus lässt sich ableiten, dass Thymol und Carvacrol vermutlich durch zwei nacheinander ablaufende Oxidationen direkt aus *γ*-Terpinen gebildet werden. *p*-Cymen scheint ein Nebenprodukt durch zu frühes Entweichen aus dem aktiven Zentrum des Enzyms zu sein.

Als Schlussfolgerung ergibt sich aus den vorliegenden Ergebnissen, dass *p*-Cymen kein Zwischenprodukt in der Biosynthese von Thymol und Carvacrol aus *γ*-Terpinen durch die P450 Enzyme aus Oregano und Thymian darstellt. Eine mögliche Zwischenstufe könnte ein Allylalkohol sein, aus dem durch eine weitere Oxidation die aromatischen Endprodukte gebildet werden. Es wird postuliert, dass der Biosyntheseweg von Thymol und Carvacrol in Oregano und Thymian aus zwei Teilen besteht. Zuerst wird *γ*-Terpinen durch Monoterpensynthasen gebildet, welches im zweiten Schritt durch Cytochrom P450 Enzyme in einer zweistufigen Oxidation in Thymol oder Carvacrol umgewandelt wird. CYP71D178, CYP71D179 und CYP71D182 sind vermutlich Thymol-Synthasen und CYP71D180 und CYP71D182 sind vermutlich Carvacrol-Synthasen.

8 References

Aharoni A, Giri AP, Verstappen FWA, B ertea CM, Sevenier R, Sun Z K, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell **16:** 3110-3131

Arnold K, Bordoli L, Ko pp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:** 195-201

Bak S, Kah n RA, Nielsen HL, Møller B L, Hal kier B A (1998) Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Molecular Biology **36:** 393-405

Bak S, Olsen CE, Halkier BA, Møller BL (2000) Transgenic Tobacco and Arabidopsis Plants Expressing the Two Multifunctional Sorghum Cytochrome P450 Enzymes, CYP79A1 and CYP71E1, Are Cyanogenic and Accumulate Metabolites Derived from Intermediates in Dhurrin Biosynthesis. Plant Physiology **123:** 1437-1448

Bechtold N, Ellis J, Pelletier G (1993) *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences **316:** 1194-1199

Bertea CM, Schalk M, Karp F, Maffei M, Croteau R (2001) Demonstration That Menthofuran Synthase of Mint (Mentha) Is a Cytochrome P450 Monooxygenase: Cloning, Functional Expression, and Characterization of the Responsible Gene. Archives of Biochemistry and Biophysics **390:** 279-286

Bohlmann J, Meyer-Ga uen G, Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proceedings of the National Academy of Sciences of the United States of America **95:** 4126-4133

Bolwell GP, Boz ak K, Zimmerlin A (1994) Plant Cytochrome-P450. Phytochemistry **37:** 1491-1506

Bouwmeester HJ, Gershenzon J, Ko nings MCJM, Croteau R (1998) Biosynthesis of the monoterpenes limonene and carvone in the fruit of caraway - I. Demonstration of enzyme activities and their changes with development. Plant Physiology **117:** 901-912

Bradford MM (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry **72:** 248-254

Braga PC, Culici M, Alfieri M, Dal Sasso M (2008) Thymol inhibits *Candida albicans* biofilm formation and mature biofilm. International Journal of Antimicrobial Agents **31:** 472-477

Buckingham J, ed (1998) Dictionary of natural products. Chapman and Hall, London

Catalan C, A.N. and Lampasona, Marina E.P. (2002) The chemistry of the genus *Lippia* (Verbenaceae). *In* SE Kintzios, ed, Oregano: The genera *Origanum* and *Lippia*, Ed 1 Vol 25. Taylor and Francis, London, pp 127-149

Chen S, Glawischnig E, Jørgensen K, Naur P, Jørgensen B, Olsen C-E , Hansen CH, Rasmussen H, Pickett JA, Halkier BA (2003) CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis. Plant Journal **33:** 923-937

Colby SM, Alonso WR, Katahira EJ, Mcgarv ey DJ, Croteau R (1993) 4*S*-Limonene synthase from the oil glands of spearmint (*Mentha spicata*) - cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. Journal of Biological Chemistry **268:** 23016-23024

Craveiro AA, Alencar JW, Matos FJA, Andrade CHS, Machado MIL (1981) Essential Oils from Brazilian Verbenaceae - Genus *Lippia*. Journal of Natural Products **44:** 598-601

Crocoll C, Asbach J, Novak J, Gershen zon J, Degenhardt J (2010) Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Molecular Biology **73:** 587-603

Croteau R, Davis E, Ringer K, Wildung M (2005) (−)-Menthol biosynthesis and molecular genetics. Naturwissenschaften **92:** 562-577

Davis EM, Croteau R (2000) Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. *In* Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids, Vol 209. Springer, Berlin Heidelberg, pp 53-95

de Kraker J W, Schurink M, Franssen MCR, Ko nig WA, de Groot A, Bo uwmeester HJ (2003) Hydroxylation of sesquiterpenes by enzymes from chicory (Cichorium intybus L.) roots. Tetrahedron **59:** 409-418

Degenhardt J, Köllner T G, Ge rshenzon J (2010) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry **70:** 1621-1637

Du LC, Lyk kesfeldt J, Olsens C E, Halkier BA (1995) Involvement of cytochrome P450 in oxime production in glucosinolate biosynthesis as demonstrated by an in vitro microsomal enzyme system isolated from jasmonic acid-induced seedlings of Sinapis alba L. Proceedings of the National Academy of Sciences of the United States of America **92:** 12505-12509

Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Faldt J, Miller B, Bohlmann J (2003) (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: Function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell **15:** 1227-1241

Emanuelsson O, Ni elsen H, Brun ak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology **300:** 1005-1016

Escobar P, Leal SM, He rrera LV, Martinez JR, Stashenko E (2010) Chemical composition and antiprotozoal activities of Colombian Lippia spp essential oils and their major components. Memorias Do Instituto Oswaldo Cruz **105:** 184-190

Fahn A (1988) Secretory-Tissues in Vascular Plants. New Phytologist **108:** 229-257

Feyereisen R (2005) Insect cytochrome P450. Elsevier

Floris I, Satta A, Cabras P, Garau VL, Angioni A (2004) Comparison between two thymol formulations in the control of Varroa destructor: Effectiveness, persistence, and residues. Journal of Economic Entomology **97:** 187-191

Gachon CM M, Langlois-Meurinne M, Saindrenan P (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends in Plant Science **10:** 542-549

Gang DR (2005) Evolution of flavors and scents. Annual Review of Plant Biology **56:** 301-325

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nature Chemical Biology **3:** 408-414

Gershenzon J, Krei s, W. (1999) Biochemistry of terpenoids: monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins. CRC Press, Boca Raton, FL

Gershenzon J, Maffei M, Croteau R (1989) Biochemical and histochemical-localization of monoterpene biosynthesis in the glandular trichomes of spearmint (*Mentha spicata*). Plant Physiology **89:** 1351-1357

Gershenzon J, Mccaskill D, Rajaonarivony JIM, Mihaliak C, Karp F, Croteau R (1992) Isolation of secretory-cells from plant glandular trichomes and their use in biosynthetic-studies of monoterpenes and other gland products. Analytical Biochemistry **200:** 130-138

Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *In* G Christine, RF Gerald, eds, Methods in Enzymology, Vol Volume 350. Academic Press, pp 87-96

Gotoh O (1992) Substrate Recognition Sites in Cytochrome-P450 Family-2 (Cyp2) Proteins Inferred from Comparative Analyses of Amino-Acid and Coding Nucleotide-Sequences. Journal of Biological Chemistry **267:** 83-90

Granger R, Passet J, Ve rdier R (1964) Le gamma-terpinene precurseur du *p*-cymene dans *Thymus vulgaris* L. Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences **258:** 5539-5541

Groves J (2005) Models and Mechanisms of Cytochrome P450 Action. *In* PR Ortiz de Montellano, ed, Cytochrome P450. Springer US, pp 1-43

Guroff G, Daly JW, Jerina DM, Renson J, Witk op B, Udenfrie.S (1967) Hydroxylation-Induced Migration: The NIH Shift. Science **157:** 1524-&

Gwinn KD, Ownley BH, Greene SE, Clark MM, Taylor CL, Springfield TN, Trently DJ, Green JF, Reed A, Hamilton SL (2010) Role of Essential Oils in Control of Rhizoctonia Damping-Off in Tomato with Bioactive *Monarda* Herbage. Phytopathology **100:** 493-501

Halkier BA (1996) Catalytic reactivities and structure/function relationships of cytochrome P450 enzymes. Phytochemistry **43:** 1-21

Halkier BA, Møller BL (1991) Involvement of Cytochrome P-450 in the Biosynthesis of Dhurrin in Sorghum bicolor (L.) Moench. Plant Physiology **96:** 10-17

Halkier B A, Nielsen HL, Koch B, Moller BL (1995) Purification and Characterization of Recombinant Cytochrome P450(Tyr) Expressed at High-Levels in Escherichia-Coli. Archives of Biochemistry and Biophysics **322:** 369-377

Hallahan TW, Croteau R (1988) Monoterpene biosynthesis - demonstration of a geranyl pyrophosphate - sabinene hydrate cyclase in soluble enzyme preparations from sweet marjoram (*Majorana hortensis*). Archives of Biochemistry and Biophysics **264:** 618-631

Hallahan TW, Croteau R (1989) Monoterpene biosynthesis - mechanism and stereochemistry of the enzymatic cyclization of geranyl pyrophosphate to (+)-*cis*-sabinene and (+)-*trans*-sabinene hydrate. Archives of Biochemistry and Biophysics **269:** 313-326

Hannemann F, Bichet A, Ewen KM, Bernhardt R (2007) Cytochrome P450 systems - biological variations of electron transport chains. Biochimica Et Biophysica Acta-General Subjects **1770:** 330-344

Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry **68:** 2831-2846

Hasemann CA, Kuru mbail RG, Boddupalli SS, Peterson JA, Deisen hofer J (1995) Structure and function of cytochromes P450:a comparative analysis of three crystal structures. Structure **3:** 41-62

Haudenschild C, Schalk M, Karp F, Croteau R (2000) Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha* spp.) in *Escherichia coli* and *Saccharomyces cerevisiae*. Archives of Biochemistry and Biophysics **379:** 127-136

He X, Orti z de Montellano PR (2004) Radical Rebound Mechanism in Cytochrome P-450-catalyzed Hydroxylation of the Multifaceted Radical Clocks α- and β-Thujone. Journal of Biological Chemistry **279:** 39479-39484

Huang M, Abel C, Sohrabi R, Petri J, Haupt I, Co simano J, Gershenzon J, Tholl D (2010) Variation of Herbivore-Induced Volatile Terpenes among *Arabidopsis* Ecotypes Depends on Allelic Differences and Subcellular Targeting of Two Terpene Synthases, TPS02 and TPS03. Plant Physiol. **153:** 1293-1310

Hummelbrunner LA, I sman MB (2001) Acute, sublethal, antifeedant, and synergistic effects of monoterpenoid essential oil compounds on the tobacco cutworm, *Spodoptera litura* (Lep., Noctuidae). Journal of Agricultural and Food Chemistry **49:** 715-720

Iijima Y, Davidovich-Rikanati R, Fridman E, Gang DR, Bar E, Lewinsohn E, Pichersky E (2004b) The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. Plant Physiology **136:** 3724-3736

Iijima Y, Gang DR, Fri dman E, L ewinsohn E, Pichersky E (2004a) Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil. Plant Physiology **134:** 370-379

Isman M B (2000) Plant essential oils for pest and disease management. Crop Protection **19:** 603-608

Jensen K, Møller BL (2010) Plant NADPH-cytochrome P450 oxidoreductases. Phytochemistry **71:** 132-141

Jerina DM, Daly JW (1974) Arene Oxides: A New Aspect of Drug Metabolism. Science **185:** 573-582

Jez JM, Ferrer JL, Bowman ME, Dixon RA, Noel JP (2000) Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry **39:** 890-902

Jiang Y, Ortiz de Montellano PR (2009) Cooperative Effects on Radical Recombination in CYP3A4-Catalyzed Oxidation of the Radical Clock β-Thujone. ChemBioChem **10:** 650-653

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology **292:** 195-202

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris AM, Goodenough PW, Johnson CB (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. Plant Cell **19:** 1994-2005

Karimi M, Inze D, Depicker A (2002) GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. Trends in Plant Science **7:** 193-195

Karp F, Harris JL, Croteau R (1987) Metabolism of Monoterpenes: Demonstration of the Hydroxylation of (+)-Sabinene to (+)-*cis*-Sabinol by an Enzyme Preparation from Sage (*Salvia officinalis*) Leaves Archives of Biochemistry and Biophysics **256:** 179-193

Karp F, Mihaliak CA, Harris JL, Croteau R (1990) Monoterpene Biosynthesis - Specificity of the Hydroxylations of (-)-Limonene by Enzyme Preparations from Peppermint (*Mentha-Piperita*), Spearmint (*Mentha-Spicata*), and Perilla (*Perilla-Frutescens*) Leaves. Archives of Biochemistry and Biophysics **276:** 219-226

Karuzina II, Archakov AI (1994) The oxidative inactivation of cytochrome P450 in monooxygenase reactions. Free Radical Biology and Medicine **16:** 73-97

Kintzios SE (2002) Profile of the multifaced prince of the herbs. *In* SE Kintzios, ed, Oregano: the genera *Origanum* and *Lippia*, Ed 1 Vol 25. Taylor&Francis, London, pp 3-8

Kokkini S, Vokou D (1989) Carvacrol-rich plants in Greece. Flavour and Fragrance Journal **4:** 1-7

Köllner T G, Held M, Lenk C, Hiltpold I, Turlings TCJ, Gershenzon J, Degenhardt J (2008) A maize (*E*)-*β*-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. Plant Cell **20:** 482-494

Köllner TG, Schnee C, Gershenzon J, Degenhardt J (2004) The variability of sesquiterpenes cultivars is controlled by allelic emitted from two *Zea mays* variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell **16:** 1115-1131

Landmann C, Fink B, Festner M, Dregus M, Engel K-H, Schwab W (2007) Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). Archives of Biochemistry and Biophysics **465:** 417-429

Lücker J, Bouwmeester HJ, Schwab W, Blaas J, van der Plas LHW, Verhoeven HA (2001) Expression of Clarkia S-linalool synthase in transgenic petunia plants results in the accumulation of S-linalyl-beta-D-glucopyranoside. Plant Journal **27:** 315-324

Lupien S, Karp F, Ponnamperuma K, Wildung M, Croteau R (1995) Cytochrome P450 limonene hydroxylases of Mentha species. Drug Metabol Drug Interact **12:** 245-260

Lupien S, Karp F, Wildung M, Croteau R (1999) Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. Archives of Biochemistry and Biophysics **368:** 181-192

Mahmoud SS, Croteau RB (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. Proceedings of the National Academy of Sciences of the United States of America **98:** 8915-8920

Makris T, Denisov I, Schlichting I, Sligar S (2005) Activation of Molecular Oxygen by Cytochrome P450. *In* PR Ortiz de Montellano, ed, Cytochrome P450. Springer US, pp 149-182

Mansuy D, Renaud J-P (1995) Heme-thiolate proteins different from cytochromes P450 catalyzing monooxygenations. Plenum Publishing Corp.; Plenum Press

Matos FJD, Machado MIL, Silva MGD, Craveiro AA, Alencar JW (2000) Essential oils of *Lippia alnifolia* Schau. (Verbenaceae) and Lippia aff. gracillis HBK, two aromatic medicinal shrubs from northeast Brazil. Journal of Essential Oil Research **12:** 295-297

Mau CJD, Karp F, Ito M, Honda G, Croteau RB (2010) A candidate cDNA clone for (-)-limonene-7-hydroxylase from *Perilla frutescens*. Phytochemistry **71:** 373-379

Meesters R JW, Duisken M, Hollender J (2009) Cytochrome P450-catalysed arene-epoxidation of the bioactive tea tree oil ingredient *p*-cymene: indication for the formation of a reactive allergenic intermediate? Xenobiotica **39:** 663-671

Meijer AH, Dewaal A, Verpoorte R (1993) Purification of the cytochrome P-450 enzyme geraniol 10-hydroxylase from cell cultures of *Catharanthus roseus*. Journal of Chromatography **635:** 237-249

Mesa-Arango AC, Montiel-Ramos J, Zapata B, Duran C, Betancur-Galvis L, Stashenko E (2009) Citral and carvone chemotypes from the essential oils of Colombian *Lippia alba* (Mill.) NE Brown: composition, cytotoxicity and antifungal activity. Memorias Do Instituto Oswaldo Cruz **104:** 878-884

Meßner B, Thulke O, Schaffner AR (2003) *Arabidopsis* glucosyltransferases with activities toward both endogenous and xenobiotic substrates. Planta **217:** 138-146

Meunier B, de Visser SP, Shaik S (2004) Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. Chemical Reviews **104:** 3947-3980

Mihaliak CA, Karp F, and Croteau R (1993) Cytochrome P-450 Terpene Hydroxylases *In* PJ Lea, ed, Methods in Plant Biochemistry (Enzymes of Secondary Metabolism), Vol 9. Academic Press, London pp 261-279

Nebert D W, Nelson DR, Coon MJ, Estabrook RW, Feyereisen R, Fujii-Kuriyama Y, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Loper JC, Sato R, Waterman MR, Waxman DJ (1991) CORRIGENDUM. DNA and Cell Biology **10:** 397-398

Nebert D W, Nelson DR, Coon MJ, Estabrook RW, Feyereisen R, Fujii-Kuriyama Y, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Loper JC, Sato R, Waterman MR, Waxman DJ (1991) The P450 Superfamily: Update on New Sequences, Gene Mapping, and Recommended Nomenclature. DNA and Cell Biology **10:** 1-14

Nelson DR (2009) The Cytochrome P450 Homepage. Human Genomics **4:** 59-65

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW (1993) The P450 Superfamily: Update on New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of Enzymes, and Nomenclature. DNA and Cell Biology **12:** 1-51

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW (1996) P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics **6:** 1-42

Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative Genomics of Rice and *Arabidopsis*. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot. Plant Physiology **135:** 756-772

Novak J, Lukas B, Bolzer K, Grausgruber-Groger S, Degenhardt J (2008) Identification and characterization of simple sequence repeat markers from a glandular *Origanum vulgare* expressed sequence tag. Molecular Ecology Resources **8:** 599-601

Ohkawa H, Tsujii H, Shimoji M, Imajuku Y, Imaishi H (1999) Cytochrome P450 biodiversity and plant protection. Journal of Pesticide Science **24:** 197-203

Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem **239:** 2370 - 2378

Ortiz de Mo ntellano PR (2010) Hydrocarbon hydroxylation by cytochrome P450 enzymes. Chemical Reviews **110:** 932-948

Palovaara J , Hakm an I (2008) Conifer WOX-related homeodomain transcription factors, developmental consideration and expression dynamic of WOX2 during *Picea abies* somatic embryogenesis. Plant Molecular Biology **66:** 533-549

Paquette SM, Bak S, Feyereis en R (2000) Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol **19:** 307 - 317

Pechous SW, Whitaker BD (2004) Cloning and functional expression of an (*E,E*)-α-farnesene synthase cDNA from peel tissue of apple fruit. Planta **219:** 84-94

Pedras MS C, Zaharia IL, Gai Y, Zhou Y, Ward DE (2001) In planta sequential hydroxylation and glycosylation of a fungal phytotoxin: Avoiding cell death and overcoming the fungal invader. Proceedings of the National Academy of Sciences of the United States of America **98:** 747-752

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research **29**

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. Cytochrome P450, Pt B **272:** 51-64

Poulos TL, Finzel BC, Gunsalus IC, Wagner GC, Kraut J (1985) The Scripps Institute P450 protein site The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. J Biol Chem **260:** 16122 - 16130

Poulose AJ, Croteau R (1978a) Biosynthesis of aromatic monoterpenes : Conversion of γ-terpinene to *p*-cymene and thymol in *Thymus vulgaris* L. Archives of Biochemistry and Biophysics **187:** 307-314

Poulose AJ, Croteau R (1978b) γ-Terpinene synthetase - key enzyme in biosynthesis of aromatic monoterpenes. Archives of Biochemistry and Biophysics **191:** 400-411

Ravichandran KG, Boddupalli SS, Hasema nn CA, Pet erson JA, Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science **261:** 731-736

Ren MZ , Chen QJ, Li L, Zhang R, Guo SD (2005) Successive chromosome walking by compatible ends ligation inverse PCR. Molecular Biotechnology **30:** 95-101

Ro DK, Arimura GL, Lau SYW, Piers E, Bohlma nn J (2005) Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. Proceedings of the National Academy of Sciences of the United States of America **102:** 8060-8065

Ro DK, Ou ellet M, Paradise EM, Bu rd H, Eng D, Paddon CJ , Newm an JD, Keasling JD (2008) Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. Bmc Biotechnology **8**

Rupasinghe S, Baudry J, Schuler MA (2003) Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. Protein Engineering **16:** 721-731

Sallaud C, Rontein D, Onillon S, Jab es F, Duffe P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M, Tissier A (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*. Plant Cell **21:** 301-317

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual, Ed 2nd Edition. Cold Spring Harbor Laboratory Press, New York

Sapir-Mir M, Mett A, Bel ausov E, Tal-Meshulam S, Frydman A, Gidoni D, Eyal Y (2008) Peroxisomal Localization of *Arabidopsis* Isopentenyl Diphosphate Isomerases Suggests That Part of the Plant Isoprenoid Mevalonic Acid Pathway Is Compartmentalized to Peroxisomes. Plant Physiology **148:** 1219-1228

Schalk M, Croteau R (2000) A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6-to a C3-hydroxylase. Proceedings of the National Academy of Sciences of the United States of America **97:** 11948-11953

Schuler M A (1996) Plant cytochrome P450 monooxygenases. Critical Reviews in Plant Sciences **15:** 235-284

Schuler M A, Werck-Reichhart D (2003) Functional genomics of P450s. Annual Review of Plant Biology **54:** 629-667

Sedy KA, Koschier E H (2003) Bioactivity of carvacrol and thymol against *Frankliniella occidentalis* and *Thrips tabaci*. Journal of Applied Entomology-Zeitschrift Für Angewandte Entomologie **127:** 313-316

Shaik S, Kumar D, de V isser SP, Alt un A, Thiel W (2005) Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes†. Chemical Reviews **105:** 2279-2328

Shephard EA, Phillips IR, Bayney RM, Pike SF, Rabin BR (1983) Quantification of NADPH: cytochrome P-450 reductase in liver microsomes by a specific radioimmunoassay technique. Biochemical Journal **211:** 333-340

Skoula M, Gotsiou P, Naxakis G, Johnson CB (1999) A chemosystematic investigation on the mono- and sesquiterpenoids in the genus *Origanum* (Labiatae). Phytochemistry **52:** 649-657

Skoula M, Harborne J B (2002) The taxonomy and chemistry of *Origanum*. *In* SE Kintzios, ed, Oregano: The genera *Origanum* and *Lippia*, Ed 1 Vol 25. Taylor and Francis, London, pp 67-108

Stahl-Biskup E (2002) Essential oil chemistry of the genus *Thymus* - a global view. *In* E Stahl-Biskup, Saez, F., ed, Thyme: The genus Thymus, Ed 1 Vol 24. Taylor & Francis, London and New York, pp 75-124

Stahl-Biskup E, Intert, F ., Holthuijzen, J., Stengele, M., Schulz , G. (1993) Glycosidically bound volatiles - a review 1986-1991. Flavour and Fragrance Journal **8:** 61-80

Starks CM, Back KW, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science **277:** 1815-1820

Stashenko EE, Martinez JR, Ruiz CA, Arias G, Duran C, Salgar W, Cala M (2010) *Lippia origanoides* chemotype differentiation based on essential oil GC-MS and principal component analysis. Journal of Separation Science **33:** 93-103

Tan S, Hu, Y and Lin,Z (2007) Cloning and functional expression of bornyl synthase from *Rosmarinus officinalis*.

Tattersall D B, Bak S, Jones PR, Olsen CE, Ni elsen JK, Hans en ML, Hoj PB, Moller BL (2001) Resistance to an Herbivore Through Engineered Cyanogenic Glucoside Synthesis. Science **293:** 1826-1828

Theis N, L erdau M (2003) The evolution of function in plant secondary metabolites. International Journal of Plant Sciences **164:** S93-S102

Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Current Opinion in Plant Biology **9:** 297-304

Thompson JD (2002) Population structure and the spatial dynamics of genetic polymorphism in thyme. *In* E Stahl-Biskup, Saez, F., ed, Thyme: The genus Thymus, Ed 1 Vol 24. Taylor & Francis, London, pp 44-74

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research **25:** 4876-4882

Thompson JD, Manicacci D, Tarayre M (1998) Thirty-five years of thyme: A tale of two polymorphisms. Bioscience **48:** 805-815

Turner G, Gershenzon J, Nielson EE, Froehlich JE, Croteau R (1999) Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. Plant Physiology **120:** 879-886

Ultee A, Bennik MHJ, Moezelaar R (2002) The phenolic hydroxyl group of carvacrol is essential for action against the food-borne pathogen Bacillus cereus. Applied and Environmental Microbiology **68:** 1561-1568

Urban P, Werck- Reichhart D, Teutsch HG, Du rst F, Regnier S, Kazmaier M, Pompon D (1994) Characterization of Recombinant Plant Cinnamate 4-Hydroxylase Produced in Yeast - Kinetic and Spectral Properties of the Major Plant P450 of the Phenylpropanoid Pathway. European Journal of Biochemistry **222:** 843-850

Van de Peer Y, De Wac hter R (1994) Treecon for Windows - a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Computer Applications in the Biosciences **10:** 569-570

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends in Plant Science **5:** 380-386

Vokou D, Kokkini S, Bessier e JM (1993) Geographic-variation of Greek oregano (*Origanum vulgare* ssp *hirtum*) essential oils. Biochemical Systematics and Ecology **21:** 287-295

Werck-Reichhart D, Bak Sr, Paquette S (2002) Cytochromes P450. *In* The Arabidopsis Book, Vol null. The American Society of Plant Biologists, pp 1-28

Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. Biochemistry **37:** 12213-12220

Williams PA, Cosme J, S ridhar V, Johnson E, McRee DE (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol Cell **5:** 121 - 131

Wise ML, Croteau R (1999) Monoterpene biosynthesis. *In* DE Cane, ed, Comprehensive Natural Products Chemistry. Elsevier, Amsterdam, pp 97-153

Wise ML, Savage T J, Katahira E, Croteau R (1998) Monoterpene synthases from common sage (*Salvia officinalis*) - cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. Journal of Biological Chemistry **273:** 14891-14899

Wittstock U, Halkier BA (2000) Cytochrome P450CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. Journal of Biological Chemistry **275:** 14659-14666

Wüst M, Croteau RB (2002) Hydroxylation of specifically deuterated limonene enantiomers by cytochrome P450 limonene-6-hydroxylase reveals the mechanism of multiple product formation. Biochemistry **41:** 1820-1827

Wüst M, Little DB, Scha lk M, Croteau R (2001) Hydroxylation of Limonene Enantiomers and Analogs by Recombinant (-)-Limonene 3- and 6-Hydroxylases from Mint (*Mentha*) Species: Evidence for Catalysis within Sterically Constrained Active Sites. Archives of Biochemistry and Biophysics **387:** 125-136

Yamazaki M, Shibata S, Usui T (1963) The Biogenesis of Plant Products. 2. Biogenesis of Thymol. Chemical & Pharmaceutical Bulletin **11:** 363-&

Yonekura-Sakakibara K (2009) Functional genomics of family 1 glycosyltransferases in Arabidopsis. Plant Biotechnology **26:** 267-274

Yu X H, Liu CJ (2006) Development of an analytical method for genome-wide functional identification of plant acyl-coenzyme A-dependent acyltransferases. Analytical Biochemistry **358:** 146-148

Zavala-Paramo G, Chavez-Moctezuma MP, Garcia-Pineda E, Yin S, Cha ppell J, Lo zoya-Gloria E (2000) Isolation of an elicitor-stimulated 5-epi-aristolochene synthase gene (gPEAS1) from chili pepper (*Capsicum annuum*). Physiologia Plantarum **110:** 410-418

9 Supplementary Material

9.1 Supplementary Material for Chapter I

Table S1 Oligonucleotide sequences used for gene expression and synthesis of RNA hybridization probes

Oligo sequences for gene expression. Start and stop codons are underlined.

*Ov*TPS1-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCTACCATTAGCATACATCATGT-3'
*Ov*TPS1truSig-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCGTCATCTGCCTCTCGCCTC-3'
*Ov*TPS1truRR-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGACCCGGCGTTCCGCAAACTACGAGC-3'
*Ov*TPS1-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCATACGTATGGGTGGAAGAAC-3'
*Ov*TPS2-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTACCCTTAGCATGCAAGTGTC -3'
*Ov*TPS2truSig-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCATCTAAACCAATGGTGGC-3'
*Ov*TPS2-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACACATATGGCTCGAAAATAAGGC-3'
*Ov*TPS3-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCAGAAATCTGTGCATCGGCT-3'
*Ov*TPS3-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAATATGATACTCGATCGAGTGTAT-3'
*Ov*TPS4-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAAATATATTCACCGGTGGTTCC-3'
*Ov*TPS4-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAATGGGATCAACAAACACCATC-3'
*Ov*TPS5-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAAGGTTCAAAACCAGCT-3'
*Ov*TPS5truRR-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCGACGTTCTGGAAACTAC-3'
*Ov*TPS5-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCATGCATATGTCTCGAACAAC-3'
*Ov*TPS6-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAATTTCCGGCATCGGTTGCT-3'
*Ov*TPS6-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTATACGGGATCAACGAGTATGGATTT-3'
*Ov*TPS7-gt-fwd	5'-GGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTCTACCATTAGCATAAATCTTA-3'
*Ov*TPS7truRR-gt-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGACCCGGCGTTCCGCAAACTACGAGC-3'
*Ov*TPS7-gt-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCATACGTATGGGTGGAAGAAC-3'

Oligos for RNA hybridization probe synthesis:

*Ovtps*1-fwd	5'-TTCTCAAAGACCGCGGCATCAACACTATCCT-3'
*Ovtps*1-rev	5'-CTGCAGCCCCCGCCCCTTGTTC-3'
*Ovtps*2-fwd	5'-TCACCGACGCGATTCGAAAATGGGACTT-3'
*Ovtps*2-rev	5'-GGGGTCCGCCATCGCCGTGTTC-3'
*Ovtps*3-fwd	5'-GTGATGGCATACATGGTAGAGGCAGAATGGTGT-3'
*Ovtps*3-rev	5'-ACGTACAATAAATTGATGACGCGAGCAAGATTGAGAAC-3'
*Ovtps*4-fwd	5'-GAGGGGGAGCGTTGTACCGCGTTGAAT-3'
*Ovtps*4-rev	5'- GGCTCGATGCATTCTTCGTTCATGTCCTTCC-3'
*Ovtps*5-fwd	5'-TCGATATCGTCTCCGACGATCATTTCCCAGT-3'
*Ovtps*5-rev	5'-TGTCTCGAACAACAGCCCTCCCATCTGTTTATGTATT-3'
*Ovtps*6-fwd	5'-CACACTATTCACTAGCGTTGTTCGAAGGTGGGACAT-3'
*Ovtps*6-rev	5'-GGTCGTAAGGATTGGAACAGAGGCTGGTCGT-3'

Oligos for qRT-PCR

*Ovtps*2-fwd	5'-GTGGCTGAGTTTGGTGGAAGG-3'
*Ovtps*2-rev	5'-TTGGCGTTCTCTAGGTATTCTGC-3'
*Ovtps*3-fwd	5'-AGGCAGAATGGTGTTTTAGCAAG-3'

*Ovtps*3-rev	5'-GATCCTCCATCCCAACTAAAGAAG-3'
*Ovtps*5-fwd	5'-ACCTCAACAATGCCAAAGTTTCG-3'
*Ovtps*5-rev	5'-TGGTATCCGTACACGCTCTCG-3'
*Ovtps*6-fwd	5'-AGGGCGTATCGGAGGAAGAAG-3'
*Ovtps*6-rev	5'-GGTCGTAAGGATTGGAACAGAGG-3'
OvEF1alpha-fwd	5'-CTCCAGTTCTTGATTGCCACAC-3'
OvEF1alpha-rev	5'-GCTCCTTTCCAGACCTCCTATC-3'

Table S2 Essential oil composition of the *Origanum vulgare* plant lines (June 2006). Mean values (± SE) in mg g^{-1} fresh weight (n=3).

Plant line: Terpene	d2	d5	d8	f2	f3	f4	f5	ff2	ff4	ff6	ff7	df5	df6	df7	df8
α-thujene	0.07 (0.02)	0.08 (0.00)	0.07 (0.01)	0.13 (0.01)	0.14 (0.01)	0.14 (0.03)	0.14 (0.04)	0.02 (0.00)	0.07 (0.00)	0.12 (0.01)	0.03 (0.00)	0.21 (0.04)	0.25 (0.07)	0.24 (0.04)	0.20 (0.03)
α-pinene	0.07 (0.02)	0.08 (0.00)	0.07 (0.01)	0.14 (0.02)	0.16 (0.00)	0.15 (0.03)	0.14 (0.04)	0.05 (0.01)	0.06 (0.00)	0.14 (0.02)	0.12 (0.00)	0.11 (0.02)	0.12 (0.03)	0.12 (0.02)	0.10 (0.02)
camphene	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)	0.04 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)	0.03 (0.00)	0.07 (0.01)	0.00 (0.00)	0.00 (0.00)	0.05 (0.02)	0.00 (0.00)	0.02 (0.02)
sabinene	0.37 (0.05)	0.51 (0.02)	0.44 (0.01)	1.03 (0.05)	1.12 (0.00)	1.21 (0.11)	1.08 (0.24)	0.48 (0.03)	0.47 (0.01)	0.84 (0.06)	0.72 (0.01)	0.70 (0.07)	0.52 (0.05)	0.56 (0.03)	0.56 (0.11)
β-pinene	0.04 (0.01)	0.06 (0.01)	0.04 (0.01)	0.09 (0.01)	0.11 (0.01)	0.09 (0.02)	0.09 (0.02)	0.03 (0.00)	0.04 (0.00)	0.06 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.02)	0.05 (0.00)	0.05 (0.02)
myrcene	0.12 (0.02)	0.18 (0.04)	0.14 (0.01)	0.27 (0.01)	0.31 (0.00)	0.32 (0.03)	0.28 (0.08)	0.12 (0.01)	0.14 (0.00)	0.28 (0.02)	0.21 (0.01)	0.29 (0.04)	0.31 (0.04)	0.28 (0.02)	0.28 (0.06)
α-terpinene	0.06 (0.01)	0.06 (0.01)	0.08 (0.02)	0.10 (0.01)	0.11 (0.03)	0.10 (0.03)	0.11 (0.05)	0.01 (0.01)	0.08 (0.00)	0.10 (0.01)	0.03 (0.02)	0.24 (0.05)	0.24 (0.02)	0.24 (0.01)	0.24 (0.05)
p-cymene	0.38 (0.07)	0.49 (0.04)	0.45 (0.04)	0.50 (0.03)	0.59 (0.01)	0.54 (0.06)	0.44 (0.05)	0.01 (0.01)	0.35 (0.01)	1.03 (0.08)	0.00 (0.00)	1.12 (0.05)	0.96 (0.12)	1.07 (0.12)	1.03 (0.17)
limonene	0.05 (0.01)	0.07 (0.01)	0.06 (0.00)	0.13 (0.01)	0.22 (0.07)	0.15 (0.02)	0.13 (0.03)	0.04 (0.02)	0.03 (0.01)	0.13 (0.01)	0.14 (0.00)	0.08 (0.02)	0.07 (0.01)	0.06 (0.00)	0.06 (0.01)
β-phellandrene	0.05 (0.01)	0.06 (0.00)	0.05 (0.00)	0.12 (0.01)	0.07 (0.07)	0.14 (0.01)	0.13 (0.03)	0.06 (0.00)	0.05 (0.01)	0.10 (0.01)	0.12 (0.00)	0.07 (0.01)	0.06 (0.01)	0.04 (0.02)	0.06 (0.01)
cis-ocimene	0.30 (0.02)	0.40 (0.01)	0.36 (0.01)	0.16 (0.01)	0.19 (0.00)	0.21 (0.03)	0.17 (0.03)	0.31 (0.02)	0.08 (0.01)	0.04 (0.01)	0.33 (0.01)	0.64 (0.09)	0.55 (0.05)	0.54 (0.01)	0.56 (0.10)
trans-*β*-ocimene	0.16 (0.03)	0.16 (0.01)	0.14 (0.02)	0.25 (0.01)	0.30 (0.00)	0.30 (0.03)	0.29 (0.07)	0.19 (0.01)	0.25 (0.01)	0.20 (0.01)	0.07 (0.01)	0.36 (0.04)	0.28 (0.02)	0.32 (0.02)	0.33 (0.07)
γ-terpinene	1.01 (0.06)	1.39 (0.02)	1.18 (0.01)	2.46 (0.06)	2.67 (0.06)	3.11 (0.24)	2.59 (0.45)	0.03 (0.01)	1.19 (0.03)	2.91 (0.09)	0.05 (0.01)	4.64 (0.45)	4.30 (0.19)	4.26 (0.13)	4.60 (0.75)
cis-sabinene-hydrate	0.09 (0.02)	0.07 (0.01)	0.06 (0.01)	0.17 (0.00)	0.25 (0.02)	0.21 (0.01)	0.20 (0.02)	0.10 (0.00)	0.05 (0.01)	0.14 (0.01)	0.16 (0.02)	0.14 (0.05)	0.09 (0.01)	0.09 (0.01)	0.08 (0.00)
terpinolene	0.02 (0.02)	0.03 (0.02)	0.02 (0.01)	0.02 (0.02)	0.05 (0.01)	0.02 (0.01)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.03 (0.01)	0.03 (0.03)	0.02 (0.01)	0.07 (0.05)	0.00 (0.00)	0.02 (0.02)
trans-sabinene-hydrate	0.83 (0.04)	1.28 (0.01)	1.16 (0.06)	3.48 (0.03)	3.69 (0.07)	3.75 (0.11)	3.38 (0.25)	2.01 (0.06)	0.03 (0.00)	3.57 (0.05)	4.96 (0.17)	0.09 (0.05)	0.05 (0.01)	0.02 (0.02)	0.04 (0.00)
α-terpineol	0.06 (0.04)	0.09 (0.01)	0.09 (0.02)	0.17 (0.02)	0.22 (0.02)	0.23 (0.05)	0.17 (0.01)	0.08 (0.00)	0.05 (0.01)	0.19 (0.00)	0.22 (0.01)	0.14 (0.05)	0.08 (0.02)	0.09 (0.02)	0.09 (0.00)
thymol	0.07 (0.01)	0.10 (0.01)	0.06 (0.01)	1.23 (0.04)	1.27 (0.06)	1.19 (0.00)	1.07 (0.06)	0.03 (0.01)	1.36 (0.06)	0.10 (0.00)	0.01 (0.01)	0.77 (0.01)	0.61 (0.02)	0.68 (0.01)	0.88 (0.11)
carvacrol	0.02 (0.01)	0.04 (0.01)	0.05 (0.01)	0.09 (0.00)	0.11 (0.01)	0.10 (0.01)	0.10 (0.01)	0.03 (0.01)	0.10 (0.01)	0.02 (0.01)	0.03 (0.03)	0.14 (0.02)	0.12 (0.03)	0.13 (0.02)	0.12 (0.03)
trans-*β*-caryophyllene	0.26 (0.03)	0.36 (0.00)	0.30 (0.02)	0.23 (0.01)	0.26 (0.04)	0.26 (0.01)	0.26 (0.02)	0.28 (0.02)	0.23 (0.01)	0.16 (0.01)	0.26 (0.03)	0.45 (0.01)	0.37 (0.01)	0.43 (0.01)	0.46 (0.05)
α-humulene	0.07 (0.01)	0.08 (0.00)	0.09 (0.00)	0.06 (0.01)	0.06 (0.01)	0.07 (0.01)	0.04 (0.02)	0.07 (0.02)	0.06 (0.01)	0.03 (0.00)	0.06 (0.01)	0.07 (0.00)	0.10 (0.03)	0.10 (0.05)	0.06 (0.00)
alloaromadendrene	0.02 (0.01)	0.04 (0.00)	0.05 (0.01)	0.02 (0.00)	0.03 (0.00)	0.01 (0.01)	0.03 (0.02)	0.03 (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.04 (0.02)	0.02 (0.01)	0.03 (0.01)	0.06 (0.01)
germacrene D	0.32 (0.01)	0.46 (0.01)	0.39 (0.02)	0.70 (0.01)	0.77 (0.01)	0.82 (0.03)	0.77 (0.04)	1.22 (0.12)	0.67 (0.03)	0.33 (0.01)	0.47 (0.02)	0.89 (0.04)	0.81 (0.07)	0.87 (0.03)	1.09 (0.19)
bicyclo-germacrene	0.07 (0.03)	0.06 (0.01)	0.05 (0.01)	0.09 (0.01)	0.11 (0.00)	0.16 (0.05)	0.27 (0.12)	0.09 (0.01)	0.03 (0.00)	0.03 (0.00)	0.24 (0.01)	0.11 (0.01)	0.09 (0.03)	0.11 (0.00)	0.10 (0.01)
1,6-germa-cradiene-5-ol	0.19 (0.01)	0.33 (0.02)	0.27 (0.01)	0.07 (0.00)	0.07 (0.01)	0.08 (0.01)	0.07 (0.01)	0.05 (0.01)	0.10 (0.00)	0.08 (0.00)	0.05 (0.02)	0.26 (0.03)	0.12 (0.01)	0.17 (0.01)	0.21 (0.05)
total	4.68 (0.19)	6.51 (0.17)	5.64 (0.25)	11.74 (0.22)	12.94 (0.38)	13.37 (0.76)	11.96 (1.35)	5.37 (0.31)	5.54 (0.11)	10.71 (0.38)	8.40 (0.09)	11.63 (1.09)	10.29 (0.68)	10.52 (0.34)	11.30 (1.07)
monoterpenes	3.75 (0.20)	5.19 (0.28)	4.50 (0.36)	10.56 (0.12)	11.63 (0.32)	11.98 (1.12)	10.52 (1.85)	3.63 (0.22)	4.44 (0.06)	10.08 (0.15)	7.30 (0.09)	9.81 (1.71)	8.78 (0.59)	8.80 (0.31)	9.32 (1.37)
sesquiterpenes	0.93 (0.02)	1.32 (0.01)	1.15 (0.04)	1.18 (0.03)	1.31 (0.68)	1.40 (0.06)	1.44 (0.16)	1.74 (0.12)	1.10 (0.11)	0.63 (0.03)	1.10 (0.04)	1.82 (0.06)	1.51 (0.07)	1.72 (0.05)	1.98 (0.30)

Figure S1 Amino acid sequence alignment of the *Origanum vulgare* terpene synthases and closely related enzymes from other plants. The DDxxD and the RR(x8)W motifs (solid lines) and the ChloroP predicted transit peptides (solid line under last amino acid) are shown.

Figure on next page as fold-out.

Figure S1

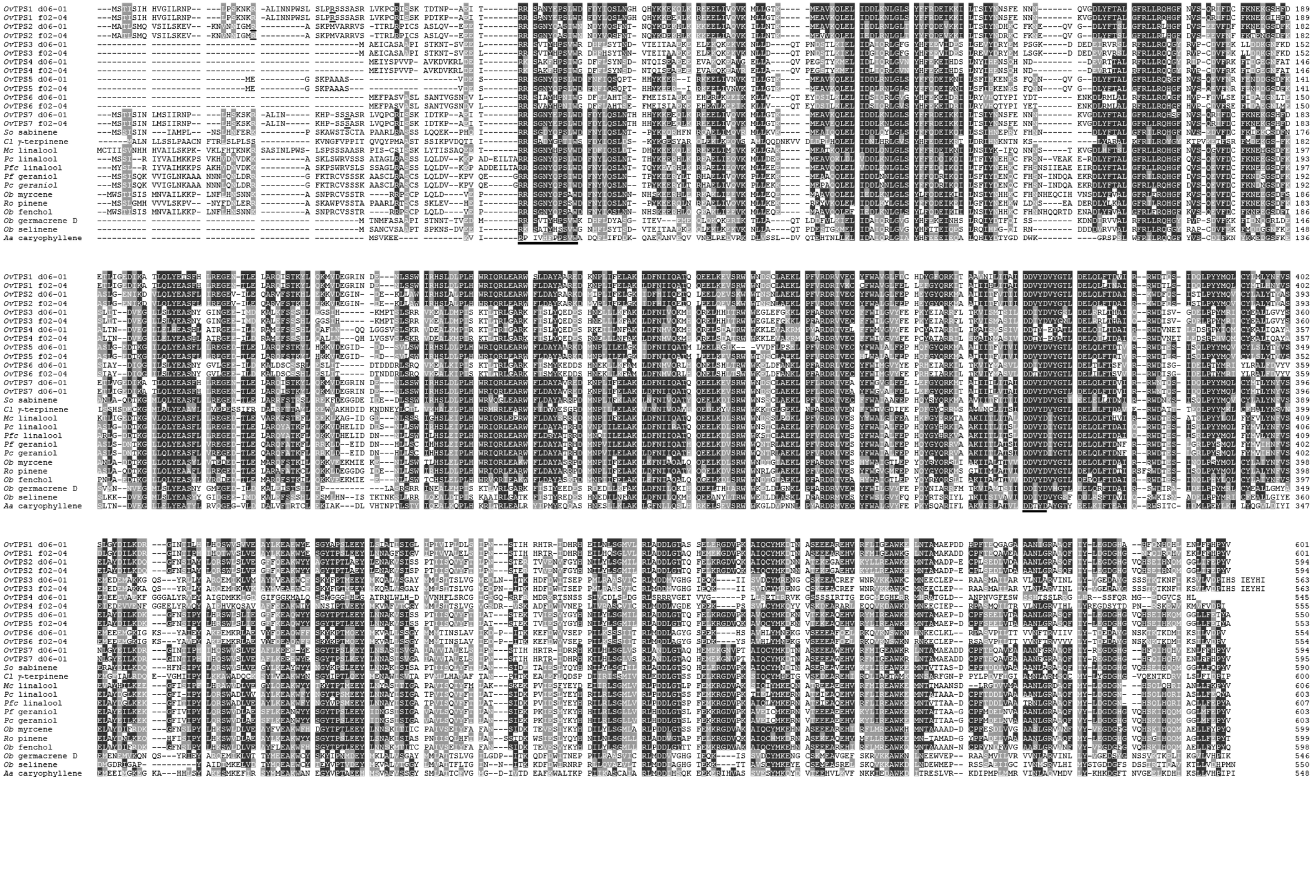

OvTPS1 d06-01
OvTPS1 f02-04
OvTPS2 d06-01
OvTPS2 f02-04
OvTPS3 d06-01
OvTPS3 f02-04
OvTPS4 d06-01
OvTPS4 f02-04
OvTPS5 d06-01
OvTPS5 f02-04
OvTPS6 d06-01
OvTPS6 f02-04
OvTPS7 d06-01
OvTPS7 f02-04
So sabinene
Cl γ-terpinene
Mc linalool
Pc linalool
Pfc linalool
Pf geraniol
Pc geraniol
Ob myrcene
Ro pinene
Ob fenchol
Ob germacrene D
Ob selinene
Aa caryophyllene

Table S3 Putative terpene synthase genes represented by ESTs in a glandular trichome cDNA library of *Origanum vulgare* cultivar f0204

Gene	# ESTs
*Ovtps*1	16
*Ovtps*2	18
*Ovtps*3	19
*Ovtps*4	8
*Ovtps*5	5
*Ovtps*6	3
*Ovtps*7	-

Table S4 Copy numbers per µg total RNA (normalized for mg plant material) from absolute qRT-PCR. Mean values (±SE), (n=9 except for NoRT and ff2 single plants n=3)

	*Ovtps*2	*Ovtps*5
d2	7254.1	1575.9
	(920.6)	(193.2)
d5	6150.0	1326.0
	(421.2)	(108.4)
f4	15795.8	6590.7
	(1347.5)	(716.9)
f5	10254.8	4190.5
	(448.8)	(136.9)
ff1	345.8	2609.8
	(61.8)	(205.5)
ff1 NoRT	238.5	59.6
	(52.8)	(12.9)
ff2-1	8835.1	2248.0
	(223.9)	(30.3)
ff2-2	227.1	1820.1
	(21.8)	(118.9)
ff2-2 NoRT	-	23.8
		(3.9)
ff2-3	10715.7	3820.6
	(166.1)	(244.7)
ff4	9862.2	2646.1
	(1508.0)	(364.5)
ff5	8656.4	4832.9
	(407.5)	(189.0)
ff6	11897.3	4384.0
	(846.1)	(280.6)
ff7	219.0	4510.6
	(52.2)	(325.9)
ff7 NoRT	129.8	36.4
	(53.1)	(12.0)
ff8	431.6	232.1
	(63.5)	(28.4)
ff8 NoRT	142.0	29.7
	(30.4)	(10.8)
df6	51240.6	4265.6
	(8927.3)	(705.9)
df8	30449.3	2653.9
	(2957.2)	(282.4)

9.2 Supplementary Material for Chapter II

Table S5 Essential oil composition of the oregano (*Ov*), thyme (*Tv*) and marjoram (*Om*) plant lines. Mean values (± SE) in mg g^{-1} fresh weight (n=9, except for *Tv*-Tc, *Ov*-Ct and *Om*-gT n=3).

Plant line: Terpene	*Tv*-Tc	*Ov*-Ct	*Om*-gT	*Ov*-d2	*Ov*-d5	*Ov*-f4	*Ov*-f5	*Ov*-df6	*Ov*-df8	*Ov*-ff1	*Ov*-ff2	*Ov*-ff4	*Ov*-ff5	*Ov*-ff6	*Ov*-ff7	*Ov*-ff8
α-thujene	33.5 (0.8)	50.4 (5.2)	cov.	cov.	cov.	36.9 (1.9)	34.4 (4.7)	cov.	cov.	2.4 (0.3)	7.4 (1.4)	15.5 (1.2)	19.3 (0.8)	cov.	cov.	2.5 (0.1)
α-pinene	22.2 (0.5)	26.4 (2.8)	10.8 (0.3)	1.8 (0.5)	2.3 (0.3)	32.2 (2.0)	33.4 (1.3)	19.7 (0.9)	21.4 (1.5)	5.5 (0.3)	11.1 (1.4)	9.9 (0.8)	22.6 (1.0)	17.9 (0.4)	15.2 (1.2)	5.2 (0.4)
camphene	22.6 (0.5)	6.8 (0.5)	n.d.	n.d.	n.d.	0.9 (0.3)	0.8 (0.3)	5.1 (0.3)	5.9 (0.6)	n.d.	6.9 (1.8)	7.8 (0.6)	n.d.	13.3 (0.4)	n.d.	n.d.
sabinene	4.8 (0.2)	20.0 (1.5)	96.9 (1.0)	37.4 (2.5)	36.1 (1.2)	356.2 (25.1)	382.7 (14.4)	188.4 (12.8)	176.2 (8.9)	94.3 (5.3)	85.6 (4.0)	45.7 (2.8)	577.1 (21.2)	124.9 (2.1)	158.7 (10.1)	88.9 (6.9)
β-myrcene	48.5 (1.5)	79.7 (7.9)	25.3 (0.5)	10.7 (0.5)	9.6 (0.3)	83.3 (3.7)	80.1 (8.4)	110.4 (6.3)	118.5 (4.0)	16.2 (0.7)	22.2 (1.8)	25.2 (1.8)	45.5 (1.5)	36.2. (0.7)	40.6 (2.9)	14.9 (1.1)
α-phellandrene	3.5 (0.4)	2.9 (0.4)	5.7 (0.1)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
α-terpinene	13.5 (1.9)	n.d.	32.7 (1.2)	9.9 (0.6)	8.1 (0.4)	80.4 (5.9)	65.6 (6.2)	157.3 (12.8)	99.7 (5.1)	8.3 (0.4)	21.9 (4.4)	28.5 (2.2)	46.4 (2.3)	45.8 (1.5)	45.4 (3.2)	6.9 (0.5)
p-cymene	400.4 (1.6)	216.1 (16.5)	45.9 (3.4)	22.3 (1.6)	35.2 (3.5)	223.6 (17.9)	154.1 (8.7)	265.5 (18.3)	219.8 (6.3)	n.d.	48.1 (16.1)	107.4 (6.6)	138.0 (10.3)	108.3 (6.8)	n.d.	n.d.
limonene	n.d.	n.d.	n.d.	6.4 (1.3)	6.3 (1.2)	60.6 (9.5)	57.2 (6.1)	19.1 (1.2)	18.5 (0.7)	11.8 (0.8)	14.6 (1.1)	8.3 (0.6)	22.1 (2.4)	26.7 (2.3)	39.8 (3.8)	10.5 (1.0)
β-phellandrene	n.d.	n.d.	n.d.	3.3 (1.1)	4.0 (1.4)	55.0 (8.3)	53.6 (11.8)	6.6 (2.2)	12.2 (0.4)	10.3 (2.0)	7.6 (3.8)	n.d.	51.8 (2.0)	21.4 (3.2)	50.5 (10.1)	9.0 (1.9)
cis-ocimene	1.5 (0.5)	16.5 (1.6)	38.1 (1.3)	44.9 (1.8)	41.1 (1.7)	21.8 (2.0)	23.2 (3.0)	237.5 (9.0)	271.8 (9.5)	69.9 (3.1)	28.0 (14.1)	6.7 (0.4)	4.0 (1.3)	n.d.	94.5 (5.0)	70.2 (5.2)
(E)-β-ocimene	2.5 (0.5)	18.2 (0.5)	4.7 (0.4)	13.6 (0.4)	11.6 (0.4)	59.1 (4.5)	71.3 (2.7)	78.9 (6.7)	86.2 (4.6)	20.1 (0.5)	15.3 (1.5)	25.2 (2.0)	317.6 (11.4)	15.4 (0.5)	12.8 (0.6)	16.8 (1.2)
γ-terpinene	858.4 (21.0)	497.4 (45.6)	380.1 (5.4)	164.5 (11.4)	133.8 (4.8)	877.8 (38.0)	944.6 (42.1)	2810.2 (170.9)	2868.8 (97.9)	10.3 (0.5)	211.1 (52.3)	301.1 (24.4)	553.5 (21.8)	460.2 (15.0)	48.2 (3.2)	8.7 (0.5)
cis-sabinene hydrate	36.8 (0.8)	17.2 (0.4)	32.3 (1.0)	n.d.	n.d.	28.8 (1.5)	26.4 (0.6)	8.5 (0.3)	9.5 (0.3)	9.5 (0.3)	10.9 (0.6)	n.d.	13.1 (0.3)	13.8 (0.2)	14.8 (0.8)	8.2 (0.9)
Linalool	48.9 (1.6)	9.8 (0.1)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
trans-sabinene hydrate	n.d.	n.d.	859.2 (3.3)	91.9 (4.1)	87.0 (2.2)	932.3 (29.5)	918.7 (29.1)	2.3 (0.9)	5.0 (0.2)	258.6 (7.4)	296.0 (20.6)	11.7 (0.8)	71.6 (1.9)	526.3 (7.0)	952.0 (66.4)	232.5 (26.9)
carvacrol methyl ether	29.4 (0.5)	35.7 (2.1)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
thymoquinone	39.9 (0.9)	124.3 (7.2)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
thymol	1641.4 (48.9)	63.6 (3.7)	n.d.	n.d.	n.d.	160.5 (18.9)	172.2 (17.2)	89.2 (12.3)	146.4 (7.7)	n.d.	2.4 (1.2)	165.3 (9.5)	57.8 (4,.0)	12.3 (0.6)	n.d.	n.d.
carvacrol	51.1 (1.7)	6297.6 (108.8)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
carvacrol acetate	n.d.	33.3 (1.4)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
(E)-β-caryophyllene	125.0 (2.3)	83.8 (13.9)	134.0 (0.8)	49.5 (2.6)	35.9 (3.2)	69.2 (4.1)	82.8 (5.9)	197.1 (12.8)	231.3 (5.2)	55.1 (1.9)	35.5 (5.7)	55.7 (4.1)	54.4 (1.9)	22.2 (1.0)	54.4 (4.8)	51.2 (2.9)
α-humulene	3.4 (0.1)	9.2 (1.4)	20.1 (0.2)	4.1 (1.3)	2.1 (1.1)	9.9 (0.6)	12.1 (1.1)	12.0 (0.9)	13.7 (0.3)	3.7 (1.5)	n.d.	8.0 (1.2)	n.d.	n.d.	7.1 (1.4)	7.7 (0.3)
(-)-germacrene D	31.8 (0.6)	34.8 (6.5)	376.9 (3.2)	79.1 (4.6)	60.2 (5.0)	228.8 (10.7)	265.0 (17.0)	525.3 (24.5)	610.9 (14.3)	279.7 (10.2)	152.7 (29.7)	189.9 (14.6)	239.7 (7.4)	77.5 (3.3)	121.5 (10.9)	256.0 (13.8)
bicyclo-germacrene	4.2 (0.0)	1.6 (1.6)	16.2 (0.1)	n.d.	n.d.	17.1 (0.8)	18.9 (1.7)	13.7 (1.2)	21.1 (0.7)	2.9 (1.5)	n.d.	n.d.	16.0 (0.6)	n.d.	37.0 (3.8)	5.3 (3.8)
cis-*α*-bisabolene	7.1 (0.2)	60.8 (9.9)	2.7 (0.1)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

cov. = peaks of *α*-thujene in GC analysis were covered by a non-terpene compound and could not be calculated; n.d. = not detectable

Table S6 Oligonucleotide sequences used for gene expression, synthesis of RNA hybridization probes and qRT-PCR

Oligo sequences for gene expression, Start and stop codons are underlined	
CYP71D178-182-fwd	5'- GATGGATATTTCAATTTCATGGGT -3'
CYP71D178-182rev	5'- ATTATGAGGTTGGATTGTGGATT-3'
CYP71D178-182-Not1-fwd	5'- TAAAAGCGGCCGCGATGGATATTTCAATTTC -3'
CYP71D178-182-Spe1-rev	5'- TTAAACTAGTTTATGAGGTTGGATT -3'
Oligo sequences for RNA hybridization analysis	
CYP71D178-fwd	5'- GGCTTCACCCTCCTTTCCCGATTATACCAAGAC -3'
CYP71D178-rev	5'- GGCCTGGCTTCTCCGTCATGTCAATGTC -3'
Oligo sequences for qRT-PCR analysis	
CYP71D178-fwd	5'- CAAGGAATGACTGCTGCTGAC -3'
CYP71D178-rev	5'- TTGGATTGTGGATTGTTGGAACC -3'
CYP71D179/182-fwd	5'- CGTGGCTTCTCAACCTTCTC -3'
CYP71D179/182-rev	5'- CGCTCTTCTTCACCCTATGC -3'
CYP71D180-fwd	5'- GCAAAGAAGAATGCGAGGTC -3'
CYP71D180-rev	5'- GATTGAACGTGTCGGGATCT-3'
CYP71D181-fwd	5'- TACTGGAAAGACCCCGACAC -3'
CYP71D181-rev	5'- CGAACGGGATTAACTCGAAA -3'

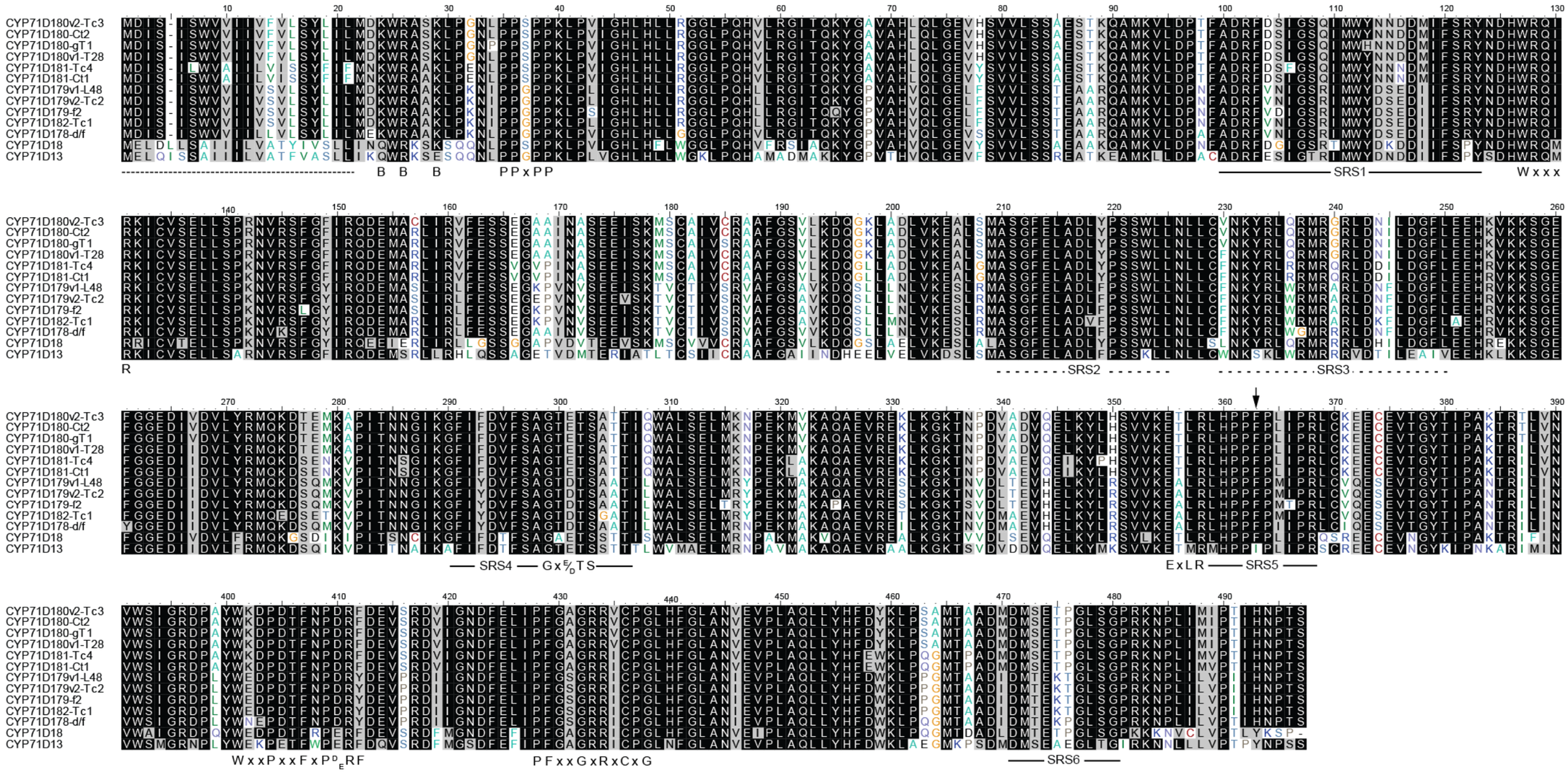
Table S7 Copy numbers per µg total RNA (normalized for mg plant material) from absolute qRT-PCR. Mean values (±SE), (n=9 except for plant lines *Ov*-Ct, *Tv*-Tc and *Om*-gT n=3)

	CYP71D178	*CYP71D178/179/182*	*CYP71D180*	*CYP71D181*	*Ovtps*2
Ov-d2	862.9 (152.0)	1349.2 (201.2)	10.1 (3.8)	3.5 (1.1)	7622.9 (1215.4)
Ov-f5	11850.5 (386.0)	17385.5 (373.3)	0.1 (0.1)	5.1 (1.0)	10937.7 (627.0)
Ov-ff4	25965.4 (3767.1)	32086.2 (4651.4)	1.7 (0.8)	22.4 (20.1)	19715.3 (3149.5)
Ov-ff7	17089.0 (1126.1)	21789.9 (1160.7)	4.1 (2.2)	0.1 (0.1)	290.0 (77.8)
Ov-ff8	12.3 (4.2)	18.3 (3.7)	5.3 (3.0)	33.5 (5.8)	539.7 (90.9)
Ov-df6	10008.4 (1520.1)	13677.2 (1971.4)	1.7 (1.1)	3.4 (2.9)	58443.8 (8126.0)
Ov-Ct	2360.4 (49.6)	9344.8 (124.1)	381.0 (29.2)	85007.0 (1880.0)	198714.1 (9640.0)
Tv-Tc	399.2 (10.4)	146901.5 (314.0)	205.6 (16.5)	17.6 (0.6)	129.7 (7.6)
Om-gT	0.1 (0.1)	13.3 (6.7)	n.d.	7.7 (1.8)	21071.1 (1982.8)

Figure S2 Amino acid alignment of eleven cytochrome P450s all five named cytochrome P450s, CYP71D178-CYP71D182. CYP71D18 from mint is added for comparison. Common sequence motifs of cytochrome P450s are shown: the P450 'signature' sequence PFxxGxRxcxG; WxxxR motif; ExLR motif; proline rich hinge (PPxPP); the membrane anchor is underlined with a dotted line. Putative substrate recognition sites are underlined and named from SRS1 to SRS6. SRS2 and SRS3 are likely found within the markings of the broader dotted regions.

Figure on next page as fold-out.

Figure S2


CYP71D180v2-Tc3
CYP71D180-Ct2
CYP71D180-gT1
CYP71D180v1-T28
CYP71D181-Tc4
CYP71D181-Ct1
CYP71D179v1-L48
CYP71D179v2-Tc2
CYP71D179-f2
CYP71D182-Tc1
CYP71D178-d/f
CYP71D18
CYP71D13
B B B
PPxPP
SRS1
Wxxx
R
SRS2
SRS3
SRS4
GxϕTS
ExLR
SRS5
WxxPxxFxPφRF
PFxxGxRxCxG
SRS6

9.3 Supplementary Material for Chapter III

Table S8 Data of Figures 5, 7, 8, 10 and 11. Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180v1*) that had been fed with different monoterpenes. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. All plants were put in one glass vessel for monoterpene feeding. Amounts presented are mean values (± standard error) in µg g^{-1} dry weight (n = 9).

		Plant line			
Terpene fed	Terpene products	CYP71D178	CYP71D180v1	Vector Ctrl	Col-0 WT
γ-terpinene					
	p-cymen-8-ol	4.92 (0.68)	5.72 (0.75)	4.02 (1.37)	5.12 (0.98)
	thymol	6.51 (0.45)	n.d.	n.d.	n.d.
	cuminol	11.19 (0.72)	11.17 (0.38)	9.08 (1.01)	8.59 (0.44)
	carvacrol	152.52^{a} (8.04)	443.78^{b} (12.09)	13.85^{c} (1.50)	23.58^{c} (2.77)
α-terpinene					
	p-cymen-8-ol	6.80^{b} (0.37)	18.12^{a} (0.66)	10.04^{b} (0.58)	9.72^{b} (0.94)
	mt3	3.41^{b} (3.41)	34.67^{a} (1.94)	5.63^{b} (0.73)	2.84^{b} (1.48)
	thymol	n.d.	2.81 (2.81)	n.d.	n.d.
	cuminol	11.08^{b} (0.36)	20.58^{a} (0.80)	10.99^{b} (0.46)	11.53^{b} (0.51)
	carvacrol	81.55^{b} (3.71)	476.93^{a} (59.89)	25.93^{b} (1.55)	93.08^{b} (19.66)
	mt10	9.04 (0.74)	43.92 (2.66)	n.d.	n.d.
p-cymene					
	p-cymene-8-ol	10.77^{a} (0.46)	13.83^{b} (0.30)	15.11^{b} (0.49)	19.99^{c} (0.63)
	mt2	1.39^{a} (0.56)	8.89^{ab} (0.26)	2.20^{a} (0.76)	4.19^{a} (0.16)
	mt3	4.32 (1.21)	2.98 (0.14)	2.20 (0.42)	3.22 (0.42)
	cuminol	12.99^{a} (2.03)	26.12^{b} (1.31)	22.97^{b} (0.68)	24.73^{b} (0.74)
	carvacrol	n.d.	2.78 (0.13)	n.d.	n.d.

Values followed by different letters (a-c) are significantly different ($P < 0.05$) in one-way ANOVA followed by a Tukey's test for all pairwise comparisons. mt1, mt2, mt3, mt10 are unidentified monoterpenes.

Table S8 continuation of previous page

Terpene fed	Terpene products	Plant line			
		CYP71D178	CYP71D180v1	Vector Ctrl	Col-0 WT
(-)-*R*-*α*-phellandrene					
	p-cymene-8-ol	1.87^{ab} (0.60)	13.69^{ab} (1.32)	0.44^{b} (0.29)	24.86^{ab} (1.91)
	mt1	n.d.	11.04 (2.55)	n.d.	n.d.
	mt3	2.57 (0.85)	0.32 (0.32)	3.32 (0.93)	5.82 (0.23)
	cuminol	6.28 (1.88)	5.40 (0.34)	3.56 (0.69)	5.66 (0.23)
	carvacrol	n.d.	2.57 (0.55)	n.d.	n.d.
	mt10	n.d.	250.44 (11.01)	n.d.	0.37 (0.37)
(+)-*R*-limonene					
	trans-*p*-menth-2,8-dien-1-ol	66.17 (1.54)	48.72 (3.58)	60.25 (2.66)	57.03 (4.79)
	cis-*p*-menth-2,8-dien-1-ol	109.07^{a} (2.49)	42.05^{b} (3.10)	71.09^{b} (2.57)	56.32^{b} (3.98)
	(+)-*trans*-isopiperitenol	386.22^{a} (30.66)	68.04^{b} (4.38)	112.52^{b} (11.43)	52.09^{b} (2.96)
	(+)-*trans*-carveol	44.22^{b} (1.65)	231.25^{a} (7.45)	47.61^{b} (2.21)	48.05^{b} (3.15)
	(+)-*cis*-carveol	102.07^{b} (7.71)	829.16^{a} (32.35)	88.70^{b} (7.54)	187.53^{c} (12.95)
	carvone	0.27^{b} (0.27)	9.38^{a} (0.31)	1.20^{b} (0.38)	3.78^{b} (0.32)
	limonene-10-ol	23.08 (1.06)	31.65 (2.40)	32.82 (3.98)	29.43 (2.05)
	perillyl-alcohol	18.11^{b} (1.12)	34.32^{a} (2.52)	15.28^{b} (0.86)	17.28^{b} (1.14)
(-)-*S*-limonene					
	trans-*p*-menth-2,8-dien-1-ol	8.90 (1.42)	7.09 (0.99)	6.94 (0.62)	8.72 (0.33)
	cis-*p*-menth-2,8-dien-1-ol	11.69 (2.08)	6.18 (0.80)	6.30 (0.48)	7.20 (0.19)
	(-)-*trans*-isopiperitenol	47.35 (7.49)	10.49 (1.33)	13.18 (0.35)	11.16 (0.48)
	(-)-*trans*-carveol	258.13^{b} (47.38)	689.26^{a} (23.91)	138.75^{b} (5.25)	446.52^{c} (29.24)
	(-)-*cis*-carveol	29.29^{ab} (5.10)	49.92^{a} (5.96)	10.78^{b} (0.11)	21.20^{ab} (1.36)
	carvone	0.29^{a} (0.19)	4.61^{a} (0.52)	n.d.^{b}	2.28^{a} (0.10)
	limonene-10-ol	2.47^{ab} (0.51)	12.51^{a} (1.25)	n.d.^{b}	3.90^{ab} (0.18)
	perillyl-alcohol	10.56^{b} (1.09)	22.70^{a} (0.97)	13.98^{b} (0.39)	11.12^{b} (0.35)

Values followed by different letters (a-c) are significantly different ($P < 0.05$) in one-way ANOVA followed by a Tukey's test for all pairwise comparisons. mt1, mt2, mt3, mt10 are unidentified monoterpenes.

Table S9 Total amounts of hydroxylated products obtained from feeding of different monoterpene substrates to *A. thaliana* plants. All plant lines were fed together with one substrate in a single vessel (n = 3 per plant line, 12 plants in total per vessel). Approximately 65 mg of each monoterpene were available in the volatile phase for conversion over the 5 day-period of the experiment.

Plant line Substrate	CYP71D178 [µg]	CYP71D180v1 [µg]	Vector Ctrl [µg]	Col-0 WT [µg]	Total [µg]
γ-terpinene	36.7	96.8	4.0	8.2	145.7
p-cymene	6.2	10.4	9.3	11.4	37.3
α-terpinene	23.9	121.5	10.0	25.8	181.2
(-)-*R*-*α*-phellandrene	1.8	41.9	1.5	9.1	54.3
(+)-*R*-limonene	143.3	282.5	64.5	99.7	589.9
(-)-*S*-limonene	62.1	164.8	21.7	110.1	358.6

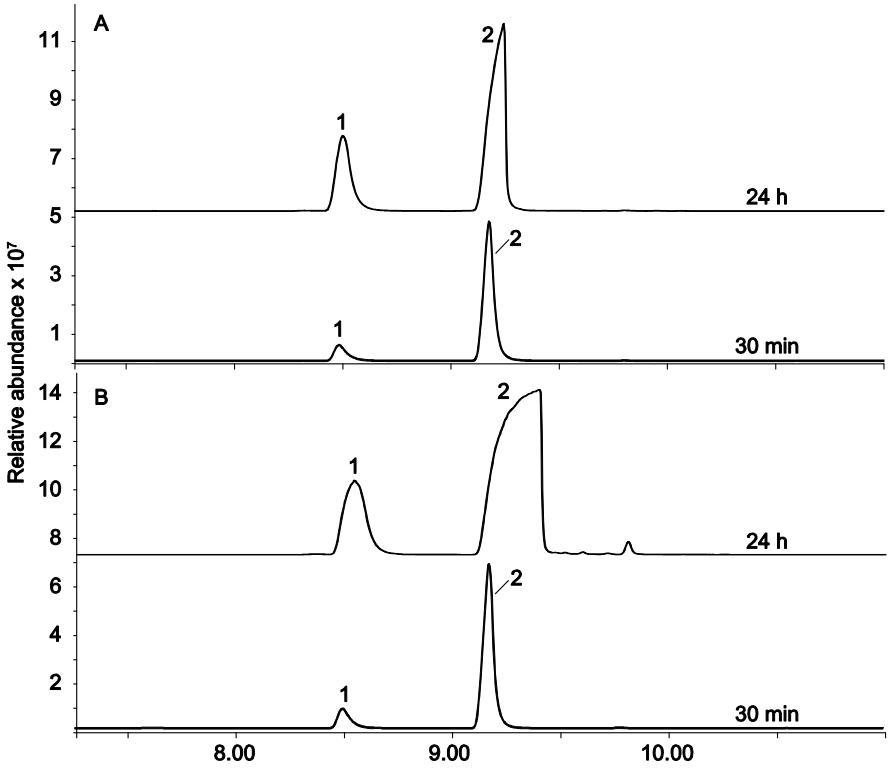


Figure S3 Spontaneous formation of *p*-cymene from *γ*-terpinene after 24 h. 20 µl *γ*-terpinene was applied in 1 l glass vessels with sealed lids and the conversion was monitored after 30 min and 24 h. (**A**) Conversion in the presence of *A. thaliana* Col-0 plants , 1, *p*-cymene; 2, *γ*-terpinene. (**B**) Spontaneous conversion in an empty glass vessel, 1, *p*-cymene; 2, *γ*-terpinene.

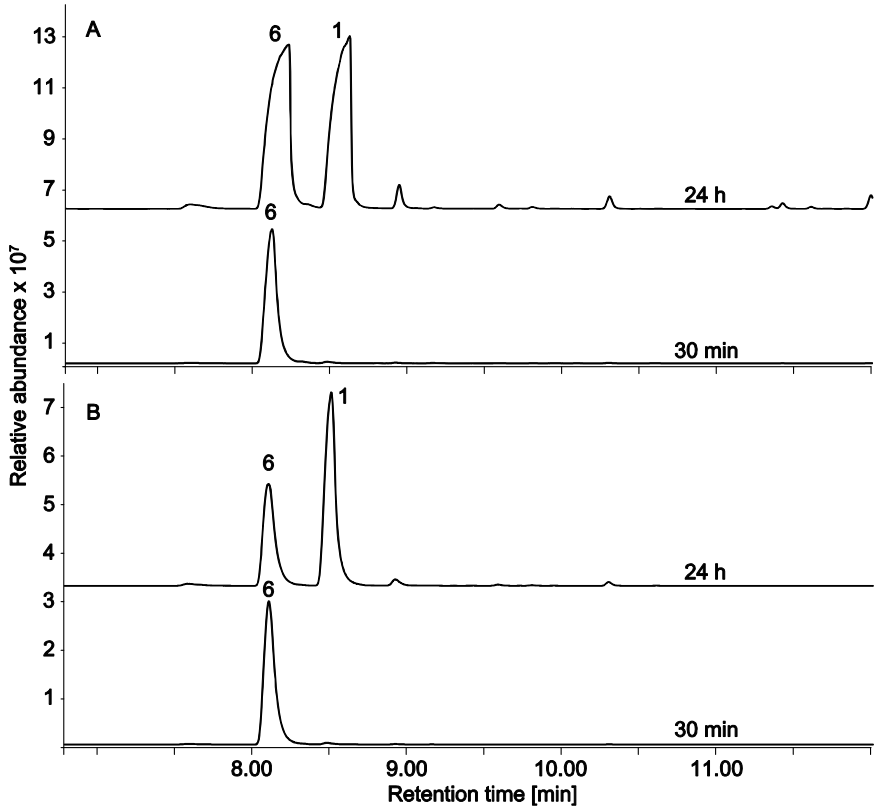


Figure S4 Spontaneous formation of *p*-cymene from (-)-*R*-*α*-phellandrene after 24 h. 20 µl (-)-*R*-*α*-phellandrene was applied in 1 l glass vessels with sealed lids and the conversion was monitored after 30 min and 24 h. (**A**) Conversion in the presence of *A. thaliana* Col-0 plants , 1, *p*-cymene; 6, (-)-*R*-*α*-phellandrene. (**B**) Spontaneous conversion in an empty glass vessel, 1, *p*-cymene; 6, (-)-*R*-*α*-phellandrene.

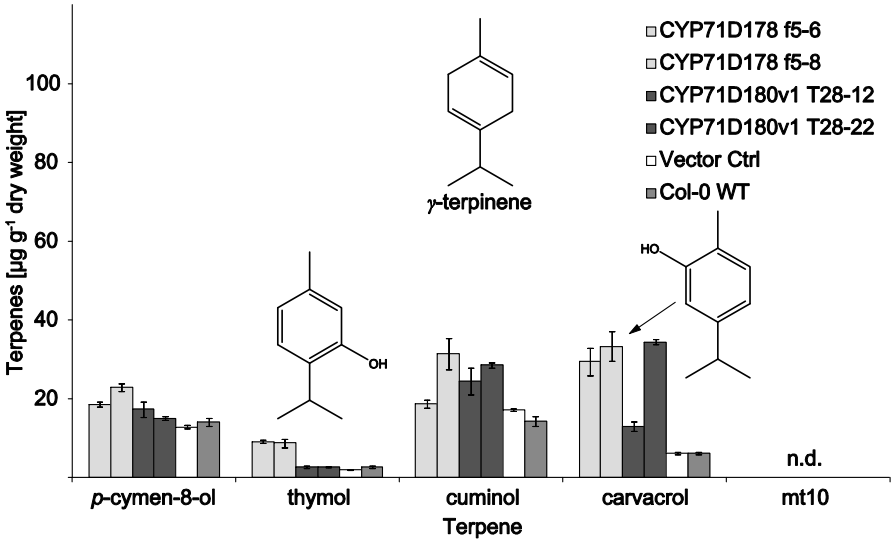


Figure S5 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *γ*-terpinene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. Structures of substrate, *γ*-terpinene, and important products, thymol and carvacrol, are shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable.

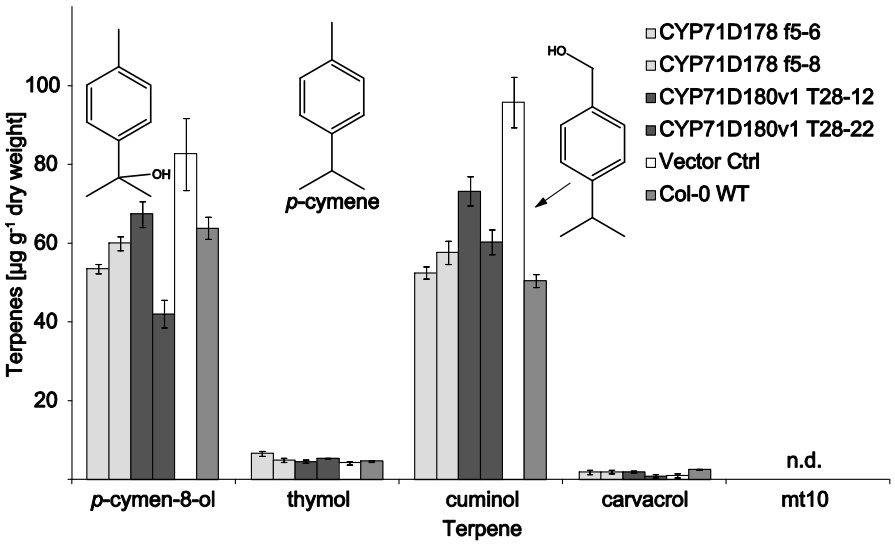


Figure S6 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *p*-cymene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. Structures of substrate, *p*-cymene, and important products, *p*-cymene-8-ol and cuminol, are shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable.

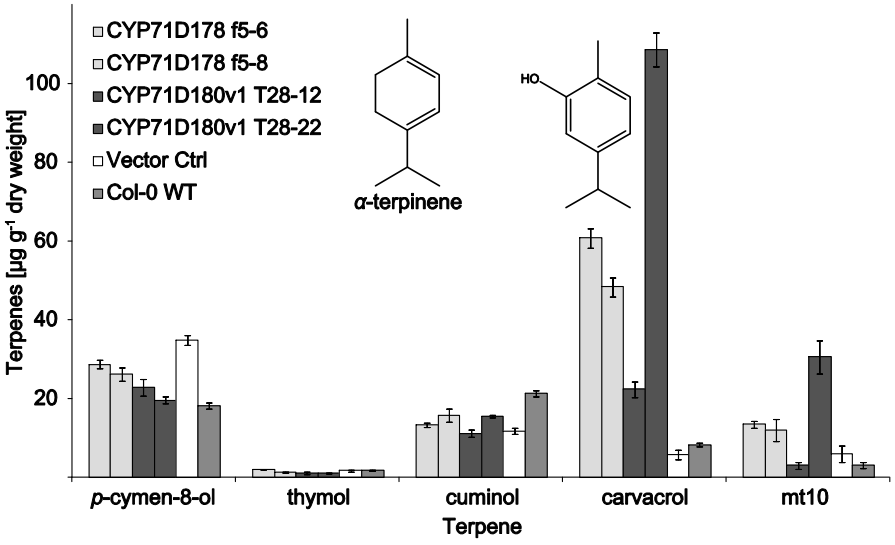


Figure S7 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *α*-terpinene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. Structures of substrate, *α*-terpinene, and the most abundant product, carvacrol, are shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable. mt10 = unidentified monoterpene.

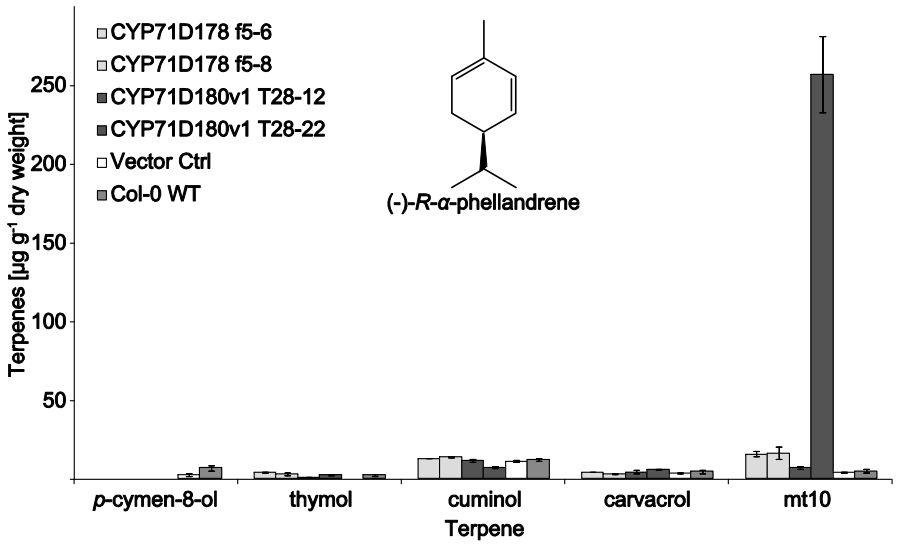


Figure S8 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with *α*-phellandrene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. The structure of substrate, *α*-phellandrene, is shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable. mt10 = unidentified monoterpene.

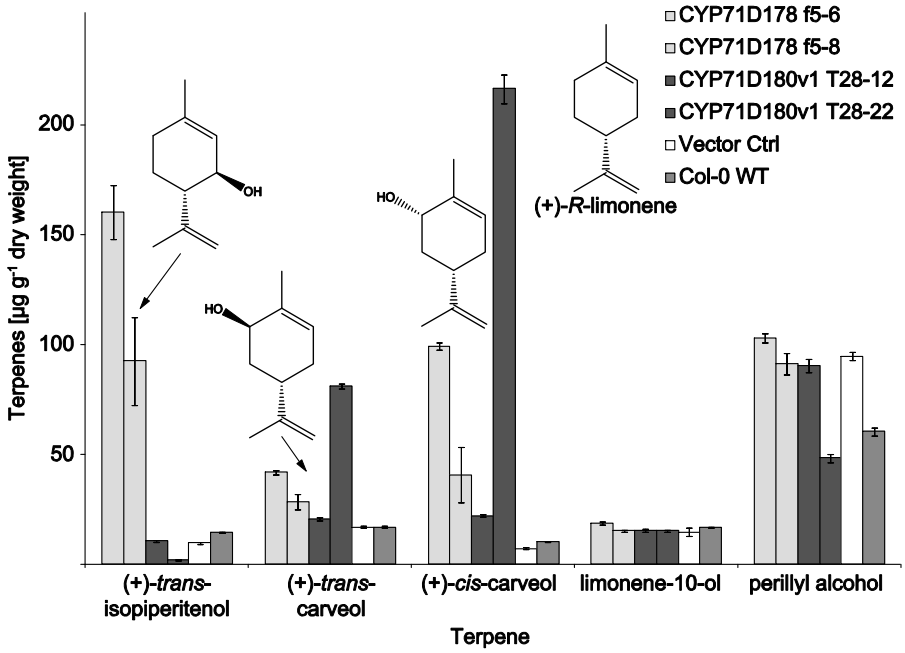


Figure S9 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with (+)-*R*-limonene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. Structure of the substrate, (+)-*R*-limonene, and major products, (+)-*trans*-isopiperitenol, (+)-*trans*-carveol and (+)-*cis*-carveol, are shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable.

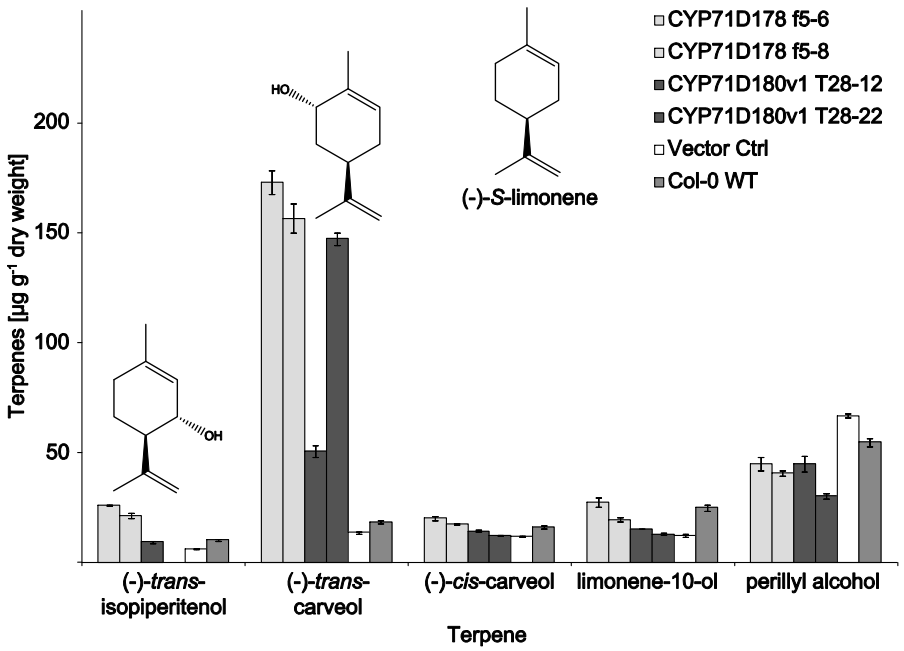


Figure S10 Amounts of terpenes released from β-glucosidase-treated extracts of *A. thaliana* transformed with CYP71D genes (*CYP71D178*, *CYP71D180*v1) that had been fed with (+)-*R*-limonene. Controls include plants transformed with an empty vector (Vector Ctrl) and wild-type *A. thaliana* Col-0 (Col-0 WT) plants. Structure of the substrate, (-)-*S*-limonene, and major products, (-)-*trans*-isopiperitenol and (-)-*trans*-carveol, are shown. Plant lines were fed separately. Amounts presented are mean values ± standard error (n = 6). n.d. = not detectable.

10 Danksagung

First of all I'd like to thank my supervisor Professor Jonathan Gershenzon for letting me work on this fascinating topic and for his faith in my work over the years. I am very thankful for his excellent supervision and lots of fruitful discussion and all the suggestions which helped me to solve minor and major problems.

I'd like to thank the community of the cytochrome P450 scientists I've met on two meetings about these fascinating enzymes in Nice and Woods Hole. All the questions and suggestions helped me to finish this project successfully. My special thanks go to Danièle Werck-Reichhart and Jean-François Ginglinger to teach me all the little secrets of P450 expression and microsomal extractions in yeast. Many thanks to David Nelson for the proper naming of „my" P450s.

Bei Julia bedanke ich mich für Zeit an der geteilten Laborbank (die ich häufig komplett belagert habe) und für das Korrekturlesen der Arbeit.

Ich danke Jörg Degenhardt für die Einführung in die Welt der Terpensynthasen und die hilfreichen Anmerkungen und Korrekturen zu den einzelnen Kapiteln.

Johannes Novak danke ich für die Auswahl der Oregano-Kultursorten mit denen alles angefangen hat.

Ich danke den Gärtnern für die Aufzucht, Vermehrung und Pflege von unzähligen Versuchspflanzen. Mein Dank geht an Udo Kornmesser, Andreas Weber, Birgit Hohmann, Jana Zitzmann und Elke Goschala und natürlich an Tamara Krügel, daß sie meinen Kräutern einen Unterschlupf gewährt hat.

John und Alex möchte ich für den Crash Kurs in qRT-PCR danken und der Abteilung Bioorganische Chemie dafür, daß ich ihre qRT-PCR Maschine mehrere Wochen im Dauerbetrieb blockieren durfte.

Ich danke Michael für die Unterstützung und Anleitung im Analytiklabor.

Katrin möchte ich für die Hilfe bei der Ernte und besonders dem Mörsern der Oregano-Blätter und beim Pipettieren der qRT-PCR-Platten danken, ohne die ich die Deadline für das erste Kapitel niemals geschafft hätte.

Claudia danke ich für hunderte Kolonie-PCRs und Minipreps, die zum Gelingen des zweiten Kapitels beigetragen haben.

Bettina danke ich für die Aufreinigung und das Laden der unzähligen Sequenzier-PCRs, die ich im Laufe der Jahre produziert habe.

Ich danke Angela und Ramona die mir bei der Bewältigung der Bürokratie stets mit Rat und Tat zur Seite standen.

Der gesamten Arbeitsgruppe möchte ich für die großartige Arbeitsatmosphäre danken und die Hilfe und Unterstützung, die sie mir während dieser Zeit haben zukommen lassen. Es hat mir immer sehr viel Spaß gemacht, mit Euch zu arbeiten. Besonders bedanken möchte ich mich bei Katharina, Diana, Alex, Julia, Meike, Almuth, Lawrie, Andreas, Holger, Sybille, Mengsu, Axel, Grit, Christine und Stefan und allen, die ich vergessen haben sollte.

Ein spezieller Dank geht noch ein meine Büromitbewohner Andy, Alex, Almuth und Kim. Der Freitag naht immer früher als man denkt.

Ein besonderer Dank gilt Ina, die die Höhen und Tiefen meiner Doktorarbeit ertragen hat und für ihre Hilfe bei der Formatierung in letzter Sekunde.

Ganz besonders danke ich meinen Eltern, die mich immer unterstützt haben!

11 Curriculum Vitae

Personal

Name:	Christoph Crocoll
Date of birth:	11.02.1977
Place of birth:	Kassel, Germany
Adress:	Erich-Weinert-Str. 21 07749 Jena Germany
Citizenship:	German
Email:	christophcrocoll@gmx.de, ccrocoll@ice.mpg.de

Scientific career

1/2006 - 2/2011	PhD Thesis at the Max Planck Institute for Chemical Ecology, Jena and the Friedrich-Schiller-University, Jena. Topic: Biosynthesis of the phenolic monoterpenes, thymol and carvacrol, by terpene synthases and cytochrome P450s in oregano and thyme
01/2005 - 12/2005	PhD Thesis at the Max Planck Institute for Chemical Ecology, Jena and the Friedrich-Schiller-University, Jena. Topic: Identification of enzymes related to terpene biosynthesis after herbivorie in maize (work was stopped due to irresolvable methodical problems).
09/2003 - 08/2004	Diploma Thesis at the Max Planck Institute for Chemical Ecology, Jena.
10/1998 - 12/2004	University Studies in biology at the Friedrich-Schiller-University, Jena. Major: ecology, microbiology, botany.

Civilian service 08/1997 - 09/1998

Education

08/1988 - 06/1997	Albert-Schweitzer-Schule, Gymnasium, Kassel
08/1984 - 07/1988	Grundschule Obervellmar, Vellmar

Work experience

8/2003 - 10/2003	student worker at the Max Planck Institute for Chemical Ecology, Jena.
10/2002 - 12/2002	student tutor for lectures in botany at the (FSU Jena, Institut für Allgemeine Botanik)
8/2002 - 11/2002	student worker at „The Jena Experiment"
8/2002 - 12/2002	student worker fort the Project „Untersuchung des Einflusses der Versalzung von Fließgewässern durch Kalihalden in Nordthüringen" (FSU Jena, AG Limnologie)

12 Publications

Scientific Articles

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J., Degenhardt, J. (2010). Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Molecular Biology, 73, 587-603.

Inderjit, Evans H., **Crocoll C.**, Bajpai D., Kaur R., Feng Y., Silva C., Carreón J.T., Valiente-Banuet A., Gershenzon J., and Callaway R.M. (*in press*). Volatile chemicals from leaf litter are associated with invasiveness of a neotropical weed in Asia. *Ecology*

Oral Presentations

Crocoll, C ., Degenhardt, J., Gershenzon, J. (2010). The route to thymol and carvacrol formation: CYP71D178-D182 from oregano, thyme and marjoram; *10th International Symposium on Cytochrome P450 Biodiversity and Biotechnology*, Woods Hole, MA, USA.

Crocoll, C., Degenhardt, J., Gershenzon, J. (2008). Molecular characterization characterization of monoterpene biosynthesis in oregano: Pathways to thymol and pizza sauce; *ICE Symposium / Max-Planck-Institut für chemische Ökologie*, Jena, Germany.

Crocoll, C., Gershenzon, J., Degenhardt, J. (2008). Functional characterization of a cytochrome P450 from *Origanum vulgare* involved in thymol biosynthesis; *6. Kurt-Mothes-Doktoranden-Workshop Sekundärstoffwechsel*, Jena, Germany.

Crocoll, C., Gershenzon, J., Degenhardt, J. (2008). Functional characterization of a cytochrome P450 from *Origanum vulgare* involved in thymol biosynthesis; *9th International Symposium on Cytochrome P450 Biodiversity and Biotechnology*, Nice, France.

Crocoll, C., Gershenzon, J., Degenhardt, J. (2008). Modifications of secondary compounds – *Origanum vulgare* terpene biosynthesis in *Arabidopsis thaliana*; *7th IMPRS Symposium / Max-Planck-Institut für chemische Ökologie*, Dornburg, Germany.

Crocoll, C., Gershenzon, J., Degenhardt, J. (2006). Spicy Yeast and Arabidopsis - expression of terpenoid biosynthesis genes; *5th IMPRS Symposium / Max-Planck-Institut für chemische Ökologie*, Jena, Germany.

Poster Presentations

Crocoll, C., Gershenzon, J., Degenhardt, J. (2009). Thymol and carvacrol formation catalyzed by a cytochrome P450 from *Origanum vulgare*; *Botanikertagung 2009* "Plants for the future" / Deutsche Botanische Gesellschaft, Universität Leipzig, Leipzig, Germany.

Crocoll, C. , Gershenzon, J., Degenhardt, J. (2009). Thymol and carvacrol formation by transgenic *Arabidopsis* overexpressing a cytochrome P450 from *Origanum vulgare*; *25th ISCE Meeting/International Society of Chemical Ecology*, Neuchâtel, Switzerland.

Crocoll, C., Gershenzon, J., Degenhardt, J. (2009). P450s @ work: Key steps in thymol and carvacrol formation: *ICE Symposium / Max-Planck-Institut für chemische Ökologie*, Jena, Germany.

Crocoll, C ., Gershenzon, J., Degenhardt, J. (2007). Elucidation of terpene biosynthesis in *Origanum vulgare* (2007). TERPNET 2007 – *8th international meeting on biosynthesis and function of isoprenoids in plants, microorganisms and parasites*, Straßburg, France.

Crocoll, C. , Degenhardt, J., Gershenzon, J. (2006). Elucidation of Terpene Biosynthesis in *Origanum vulgare*; *SAB Meeting 2006 / Max-Planck-Institut für chemische Ökologie*, Jena, Germany.

Jena, den

_____

Christoph Crocoll

13 Selbständigkeitserklärung

Die zurzeit gültige Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena ist mir bekannt. Die vorliegende Arbeit wurde von mir selbst und nur unter Verwendung der angegebenen Hilfsmittel erstellt. Alle verwendeten Quellen wurden angegeben. Alle Personen, die an der experimentellen Durchführung, Auswertung des Datenmaterials oder bei der Verfassung der Manuskripte beteiligt waren, sind benannt.

Es wurde weder bezahlte noch unbezahlte Hilfe eines Promotionsberaters in Anspruch genommen.

Die vorliegende Arbeit wurde bisher weder als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung noch bei einer anderen Hochschule als Dissertation eingereicht.

Jena, den _____

Christoph Crocoll