Available online at www.sciencedirect.com
SCIENCE DIRECT
Archives of Biochemistry and Biophysics 448 (2006) 3–12
ABB
www.elsevier.com/locate/yabbi

Isoprenoid biosynthesis in *Artemisia annua*: Cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library

Cinzia M. Bertea [a,b], Alessandra Voster [a,b,1], Francel W.A. Verstappen [a], Massimo Maffei [b], Jules Beekwilder [a], Harro J. Bouwmeester [a,*]

[a] *Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands*
[b] *Department of Plant Biology and Centre of Excellence CEBIOVEM, University of Turin, Viale P. A. Mattioli, 25 I-10125 Turin, Italy*

Received 19 September 2005, and in revised form 17 February 2006
Available online 15 March 2006

Abstract

Artemisia annua (Asteraceae) is the source of the anti-malarial compound artemisinin. To elucidate the biosynthetic pathway and to isolate and characterize genes involved in the biosynthesis of terpenoids including artemisinin in *A. annua*, glandular trichomes were used as an enriched source for biochemical and molecular biological studies. The sequencing of 900 randomly selected clones from a glandular trichome plasmid cDNA library revealed the presence of many ESTs involved in isoprenoid biosynthesis such as enzymes from the methylerythritol phosphate pathway and the mevalonate pathway, amorpha-4,11-diene synthase and other sesquiterpene synthases, monoterpene synthases and two cDNAs showing high similarity to germacrene A synthases. Full-length sequencing of the latter two ESTs resulted in a 1686-bp ORF encoding a protein of 562 aa. Upon expression in *Escherichia coli*, the recombinant protein was inactive with geranyl diphosphate, but catalyzed the cyclization of farnesyl diphosphate to germacrene A. These results demonstrate the potential of the use of *A. annua* glandular trichomes as a starting material for studying isoprenoid biosynthesis in this plant species.

Keywords: *Artemisia annua*; Germacrene A synthase; Glandular trichomes; Asteraceae; Trichome cDNA library

Artemisia annua L., a member of the Asteraceae, is an annual herb that is widely distributed in Asia, Europe, and North America. For centuries, the aerial parts of this plant species have been used in traditional Chinese medicine as a remedy against fever. In the seventies of the last century the active compound was identified as artemisinin, an endoperoxide sesquiterpene lactone. In a number of biochemical studies the artemisinin biosynthetic pathway has been established [1,2]. In addition, several groups have reported the molecular cloning of amorpha-4,11-diene synthase, the gene encoding the first step in artemisinin biosynthesis [3–5]. However, *A. annua* contains a series of other monoterpenes and sesquiterpenes and a number of other terpene synthases catalyzing the formation of different terpene structural types have also been isolated. The first sesquiterpene synthase that was cloned from *A. annua* was the *epi*-cedrol synthase [6,7], followed by the already mentioned amorpha-4,11-diene synthase [3–5] and β-caryophyllene synthase [8]. Recently, Picaud et al. [9] cloned an (*E*)-β-farnesene synthase from *A. annua*.

All the above mentioned sesquiterpene synthases have been cloned starting from cDNA libraries or by using similarity-based approaches on total RNA isolated from *A. annua* leaves. However, in *A. annua* as in many aromatic plants, terpenoid biosynthesis is strongly coupled to glandular trichomes [10,11]. Therefore, we decided to use *A. annua* glandular trichomes as an enriched material for

* Corresponding author. Fax: +31 317 418094.
E-mail address: harro.bouwmeester@wur.nl (H.J. Bouwmeester).
[1] Present address: Department of Genetics, Biology, and Biochemistry, University of Turin, via Santena, 5 bis I-10126 Turin, Italy.

0003-9861/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2006.02.026

biochemical and molecular biological studies. The isolated secretory cells are capable of converting farnesyl diphosphate (FDP)[2] into a number of different sesquiterpenes, including amorpha-4,11-diene and they have been shown to be enriched in enzymes related to the early steps in artemisinin biosynthesis [2].

To investigate the biosynthesis of sesquiterpenes, including artemisinin in *A. annua*, we decided to isolate total RNA from the glandular trichomes and use it for the construction of a plasmid cDNA library. The evaluation of expressed sequence tags (ESTs) from this plasmid cDNA library revealed the presence of a number of clones encoding putative isoprenoid biosynthesis related enzymes. Here we report on the cloning, heterologous expression and partial characterization of a germacrene A synthase from a glandular trichome specific cDNA library of *A. annua*.

Materials and methods

Plant material, substrates and reagents

Seeds of *A. annua* L. (Asteraceae) of Vietnamese origin were obtained from Artecef BV (Maarssen, The Netherlands). Plants were grown in a greenhouse in 5 L plastic pots containing potting compost at 21/18 °C (16/8 h). Natural daylight was supplemented with artificial light (SON-T AGRO) during high-temperature period, and fertilizer was applied as required. [^{3}H]FDP was obtained from Amersham Biosciences (Piscataway, NJ, USA). Unlabelled geranyl diphosphate (GDP) and FDP were purchased from Sigma–Aldrich (Sigma–Aldrich, Chemie b.v., Zwijndrecht, The Netherlands) and were used after a buffer change as described by [12] for farnesyl diphosphate. Unless otherwise stated, reagents were obtained from Sigma–Aldrich. DNA sequences were assembled and analyzed using DNASTAR software (DNASTAR, Inc., Madison, WI, USA). Sequence reagents were supplied by Perkin-Elmer (Foster City, CA, USA). Restriction and modifying enzymes were purchased from Gibco-BRL (Invitrogen Corporation, Breda, The Netherlands). DNA fragments were isolated from agarose gel by a GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences). Amino acid alignments were done with Clustal X 1.83, with Gonnet 250 matrix and default settings.

Glandular trichome essential oil analysis

To analyse the terpenoids present in the glandular trichomes, 100 mg glands—isolated as described before [2]—were ground in 1 ml dichloromethane containing 1 mg ml^{-1} *cis*-nerolidol as an internal standard. The extract was loaded on a small glass column containing $MgSO_4$. The column was washed twice with 1 ml ether/pentane (1:4) and the eluent concentrated and analyzed by gas chromatography coupled to mass spectrometry (GC–MS) on a HP 5890 series II gas chromatograph equipped with a HP-INNO-Wax column (30 m × 0.25 mm i.d., 0.25 μm df) and coupled to a HP 5972A Mass Selective Detector (Hewlett–Packard/Agilent Technologies). The inlet temperature was 250 °C and the detection temperature 290 °C. Using a constant helium flow of 1 ml min^{-1} the oven program was as follows. Initial temperature 80 °C for 2 min followed by a ramp of 5 °C min^{-1} to 260 °C and final time of 10 min. Terpenoids were identified using authentic standards, mass spectra (and comparison with Mass spectral databases such as Wiley) and retention indices as described before [1,2].

Intact gland feeding experiments

About 15 mg glands were re-suspended in 1 ml assay buffer containing 25 mM Tris, pH 7.5, 10% (v/v) glycerol, 10 mM $MgCl_2$, 6 mM Na_3VO_4, and 2 mM DTT. Twenty micromolars unlabelled FDP were added to the assay along with 2.2×10^6 dpm [^{3}H]FDP. The samples were overlaid with 1 ml pentane and incubated for 2 h at 30 °C in a water bath with gentle shaking. Each assay was then extracted twice with 1 ml ether:pentane (1:4) and the organic phase was filtered through a small glass column containing 0.9 g silica and some anhydrous $MgSO_4$. The column was washed with 1.5 ml ether and the extract was carefully concentrated to approximately 50 μl under a stream of nitrogen. The samples were analyzed by radio-GC using a Carlo Erba 4160 series GC equipped with RAGA-93 radioactivity detector (Raytest) and FID as described before [1] but with the following temperature program: 1 min at 70 °C followed by a ramp of 5 °C min^{-1} to 240 °C and a final time of 10 min.

Total RNA isolation and cDNA library construction

Total RNA was extracted from 50 mg *A. annua* gland secretory cells using the SV Total RNA Isolation System (Promega). The quality of the isolated RNA was assessed by agarose gel electrophoresis and the quantity measured spectrophotometrically. One microgram of the resulting total RNA was employed to construct a cDNA library using the SMART cDNA Library kit and following the manufacturer's protocols for small amounts of RNA (Clontech, Palo Alto, California). After double strand cDNA synthesis and introduction of *Sfi*I restriction enzyme recognition sites, the double stranded cDNA was digested with *Sfi*I and size fractionated by column chromatography according to manufacturer's instructions. To enhance the fraction of full-length clones, only the first two, largest, fractions containing detectable cDNA were employed for library construction, using the vector pGEM-T Easy (Promega) modified with *Sfi*I sites. To introduce an *Sfi*I-site into pGEM-T Easy, a double stranded *Sfi*I linker was made. The double stranded oligo was made from *Sfi*1

[2] *Abbreviations used:* FDP, farnesyl diphosphate; GC–MS, gas chromatography coupled to mass spectrometry; IPTG, isopropyl β-D-thiogalactopyranoside; GDP, geranyl diphosphate; MEP, methylerythritol phosphate; MVA, mevalonic acid.

(5′-GGCCATTATGGCCTGCAGGATCCGGCCGCCT CGGCCGA-3′) and *Sfi*2 (5′-CGGCCGAGGCGGCCGG ATCCTGCAGGCCATAATGGCCA-3′) primers. Ten nanomoles of *Sfi*1 and *Sfi*2 were denatured by heating at 100 °C for 5 min in a block heater, and allowed to anneal by cooling to room temperature in the block. One nanogram of double stranded oligo was ligated into pGEM-T Easy using T4 DNA ligase and the ligation mix was used to transform *Escherichia coli* XL-1 blue competent cells. The purified vector digested with *Sfi* I restriction enzyme was employed for cDNA library ligation. The ligation mixture was incubated overnight at 16 °C and transformed to DH5α cells (Invitrogen).

Sequence analysis

About 900 clones from the *A. annua* glandular trichome cDNA library were randomly picked, amplified using T7 and SP6 primers and partially sequenced from their 5′-end using the T7 primer. Half a microlitre of the colony PCR products was employed directly for sequencing using Ready Reaction Dye Terminator Cycle Mix (Perkin-Elmer) and 100 ng of T7 primer. Sequencing PCR was performed according to manufactures recommendations (Perkin-Elmer) in a Gene Amp PCR System 9700 Thermocycler (Applied Biosystems). After purification the samples were sequenced on a ABI 310 capillary sequencer (Perkin-Elmer). After removal of vector and poor quality sequences, the resulting *A. annua* ESTs were compared with GenBank using the BLASTX search algorithms of the NCBI (http://www.ncbi.nlm.nih.gov/blast/blast.cig) [13]. Sequences were assembled into contigs using the Seqman II module of the DNASTAR software package (DNASTAR, Inc., Madison, WI, USA).

Cloning and functional expression of the putative germacrene A synthase in E. coli

After sequencing, two clones showed high similarity to germacrene A synthases isolated from other plant species. One of these clones proved to be full-length and it was used for heterologous expression in *E. coli*. After subcloning in a bacterial vector, full-length sequencing reactions were run across the vector/insert junction to ensure proper integration. The same analysis was done on the original clone subcloned into pGEMT-Easy plasmid. Design of several internal primers allowed the complete sequencing of the cDNA encoding the putative sesquiterpene synthase. The full-length sequences were obtained using Amplitaq DNA polymerase and FS cycle sequencing on an ABI Prism™ 373 DNA Sequencer.

To determine the biochemical function, the native clone *AaGAS* was amplified by PCR and the construct was incorporated into pRSET A vector (Invitrogen) by introducing an *Xho*I site (CTCGAG) upstream of the start codon and a *Eco*RI site (GAATTC) downstream of the stop codon. The *Xho*I and *Eco*RI sites were introduced by PCR using the primers 5′-TTCCTCGAGATGGCAGCGGTTCAAGC T-3′ (forward) and 5′-GGAATTCTTACACGGGGTAG AGAAC-3′ (reverse). One hundred microliters PCR mixture contained 0.4 μM of each primer, 0.3 mM dNTP, 50 ng of template, 10 μl of 10× PCR buffer, 5 U *PfuTurbo* DNA polymerase (Stratagene) and sterile water. The PCR was carried out using a Gene Amp PCR System 9700 Thermocycler (Applied Biosystems). The mixture was initially preheated to 94 °C for 1 min, then underwent 25 cycles at 94 °C for 45 s, 55 °C for 1 min and 72 °C for 2 min. Finally, the reaction was heated at 72 °C for 7 min. Following gel purification, PCR products were digested with *Xho*I and *Eco*RI and ligated into pRSET A digested with *Xho*I and *Eco*RI. The ligation was transformed to *E. coli* DH5α competent cells (Invitrogen). Isolated DNA from bacterial colonies was fully re-sequenced to check if the gene was integrated in the right frame without mutations. A small amount of plasmid DNA was also used for digestion with *Xho*I and *Eco*RI and checked with gel electrophoresis for correct subcloning. Plasmid DNA of the clone *AaGAS* and the control (original pRSET A vector) were transformed to BL21-(DE3)-RIL competent cells according to the manufacturer's recommendations (Stratagene).

For induction of expression, single colonies of the BL21 transformations harboring the putative sesquiterpene synthase and the empty pRSET A vector were picked from the LB 100 mg L^{-1} ampicillin plates containing 1% glucose, transferred to 1 ml LB broth supplemented with 100 mg L^{-1} ampicillin and 1% glucose and grown overnight at 37 °C (250 rpm). Aliquots of 0.5 ml were then used to inoculate 250 ml Erlenmeyer flasks containing 50 ml LB broth with ampicillin (50 μg ml^{-1}). The cultures were grown at 37 °C with vigorous shaking to $OD_{600} = 0.5$. For induction of expression 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) was added and the cultures were grown at 16 °C overnight with shaking at 250 rpm. Then the cells were centrifuged at 1000*g* for 15 min at 4 °C and the pellet was resuspended in 1 ml assay buffer containing 15 mM Mopso, pH 7.0, 10% glycerol, 1 mM ascorbic acid, 2 mM DTT, 0.1% Tween 20, 6 mM Na_3VO_4, 10 mM $MgCl_2$, and 1 mM $MnCl_2$ and then disrupted by sonication. The sonicated mixture was centrifuged at 10,000*g* for 10 min at 4 °C and the supernatant used directly for enzyme assays.

Enzyme assay

For each enzyme assay 500 μl of the lysate were diluted to 1 ml total reaction volume using the MOPSO assay buffer described above and using 5 μM FDP or 5 μM GDP as substrates. The assays were overlaid with 1 ml of pentane and incubated at 30 °C for 1 h with gentle shaking. All assays were performed in duplicate. After incubation the assays were vigorously mixed and following a short centrifugation the pentane phase from each sample was removed and loaded on a short aluminum oxide column overlaid with anhydrous $MgSO_4$. The assay mixture was re-extracted with 1 ml pentane:diethyl ether (1:4), which was also passed over

the column, and the column washed with 1.5 ml diethyl ether. The combined fractions were concentrated to 50 µl under a gentle stream of N_2 and analyzed using GC–MS, equipped with an HP-5MS column (30 m × 0.25 mm i.d., 0.25 µm film thickness) and programmed at an initial temperature of 55 °C for 4 min, with a ramp of 10 °C min^{-1} to 280 °C and final time of 10 min. To check for temperature-rearrangement, analysis was done using an injection port temperature of 250 °C as well as 150 °C [12].

Results and discussion

Artemisia annua glandular trichome isolation and essential oil analysis

Gland secretory cells were obtained in a good yield by using the glass-bead abrasion technique applied for other plant species [14,15]. Microscopic analyses revealed the presence of a fair amount of 10-celled biseriate glandular trichomes as previously described by Duke et al. [10,16]. The purity of these preparations was quite high since the only contamination was represented by fragments and some non-glandular surface hairs (data not shown). The volatile profile of pentane/ether extracts of the glandular trichomes analyzed by GC–MS revealed the presence of many monoterpenoids and sesquiterpenoids including putative intermediates in artemisinin biosynthesis. The major monoterpene was identified as camphor, followed by linalool (Fig. 1). The main sesquiterpenes were identified as selina-4,11-diene, germacrene D, β-farnesene, germacrene A (identified as a β-elemene because of Cope rearrangement at high injection port temperature [12]) and amorpha-4,11-diene. These terpenoids were demonstrated to be also present in the essential oil extracted from *A. annua* leaves [1].

Intact gland cell clusters were then incubated in the presence of [^{3}H]FDP and products of conversion were analyzed by radio-GC and GC–MS. Radio-GC analysis showed a large radio-labelled amorpha-4,11-diene peak, indicating that the conversion from FDP to amorpha-4,11-diene occurred (Fig. 2). In addition, a second peak, corresponding to germacrene A was observed (Fig. 2). These results confirm that glandular trichomes from *A. annua* contain the enzymatic machinery for essential oil production as it was also demonstrated for other aromatic plants such as mint [14] and basil [15].

EST sequence analysis and database searching

Total RNA, extracted from the secretory cells, was subsequently used to make a plasmid cDNA library. The titer of this library was 1.0×10^4 colony forming units (cfu/ml). The average insert size, based on PCR analysis of 96 individual clones, was estimated to be about 1200 bp. Sequencing of these 96 clones revealed the presence of four cDNAs with high similarity to genes involved in isoprenoid biosynthesis: geranyl diphosphate synthase small subunit (GenBank Accession No. AF182827 [17]), δ-cadinene synthase (GenBank Accession No. Y16432 [18]), linalool synthase (GenBank Accession No. AF154125 [19]), and a sesquiterpene cyclase (GenBank Accession No. AJ271792 [20]). Because of these promising results, we decided to (partially) sequence about 900 randomly selected clones from their 5′-ends. Of these, 765 clones gave high quality sequences. After editing to remove vector sequences and poor quality 3′-end sequences the average size for the resulting ESTs was approximately 600 bp. This value is comparable with the EST size calculated from other gland cDNA libraries [14,15]. Fragment assembly identified a total of 459 contigs.

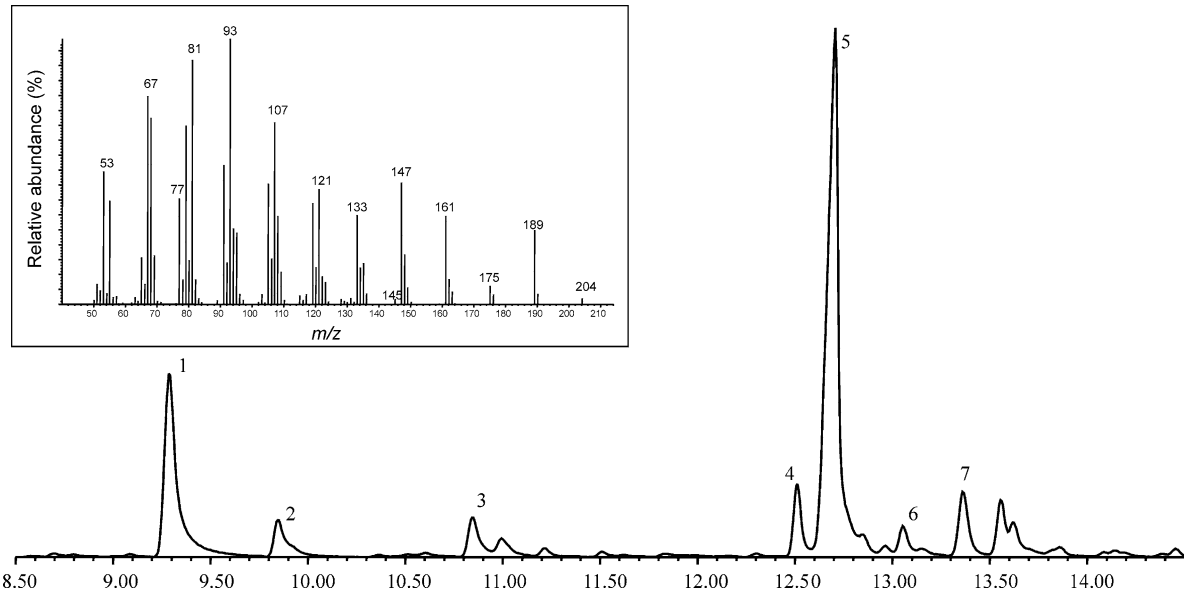

Fig. 1. GC–MS analysis (HP-INNOWax column) of essential oil extracts from *A. annua* glandular trichomes. Compounds: 1, camphor; 2, linalool; 3, β-elemene; 4, β-farnesene; 5, selina-4,11-diene; 6, amorpha-4,11-diene; and 7, germacrene D. The inset shows the mass spectrum of peak number 3, corresponding to β-elemene, the Cope rearrangement product of germacrene A.

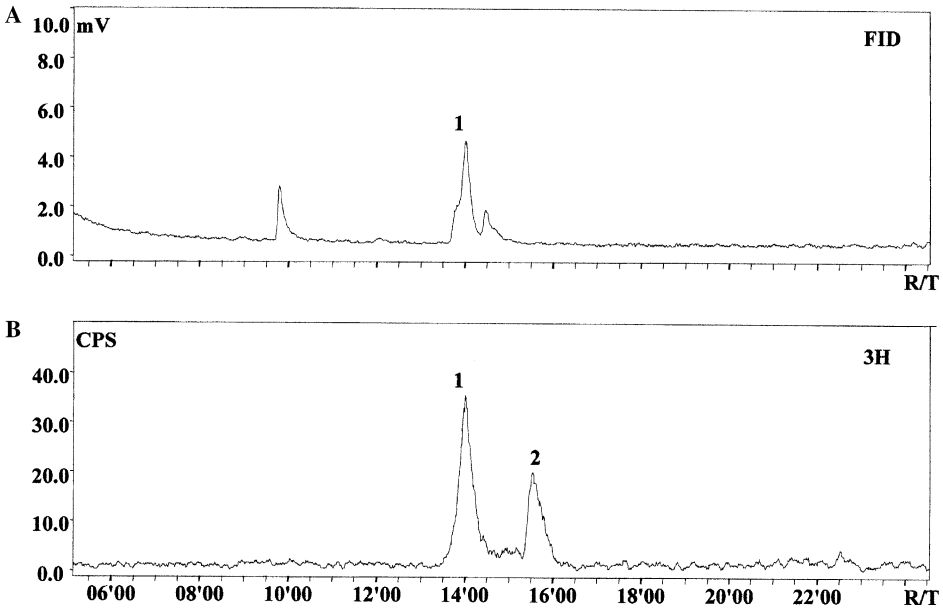

Fig. 2. Radio-GC analysis of the products formed in an assay with intact glandular cell clusters from *A. annua* with [^{3}H]farnesyl diphosphate as substrate. (A) FID trace showing a reference sample of amorpha-4,11-diene, the first intermediate in artemisinin biosynthesis (B) Radiotrace showing compounds: 1, amorpha-4,11-diene; 2, germacrene A.

Of these, 158 contigs (34.4%) contained two or more ESTs, whereas 301 (65.5%) contained only a single unique sequence (39.4% of total ESTs). One hundred forty seven contigs (32.2%) were of low redundancy (2–4 sequences per contig), 8 contigs (1.74%) were of medium redundancy (5–10 sequences per contig) and only 3 contigs (0.65%) were of high redundancy (11–43 sequences per contig). In comparison to the mint [14] and the basil gland libraries [15], the cDNA library from *A. annua* glandular trichomes contained a larger number of unique sequences and a relatively low redundancy, suggesting that the *A. annua* glands have a higher complexity at the biochemical and molecular level in comparison to the glandular trichomes of mint and basil (see also below).

ESTs were then analyzed for sequence homologies using the BLASTX algorithm and they were assigned functions based on highest similarity, and categorized into five general functional groups (Fig. 3). About 47% of the clones encoded proteins and enzymes involved in primary (amino acid metabolism, lipid metabolism, glycolysis, photosynthesis, and pentose phosphate pathway) and secondary metabolism (isoprenoid and flavonoid biosynthesis) with secondary

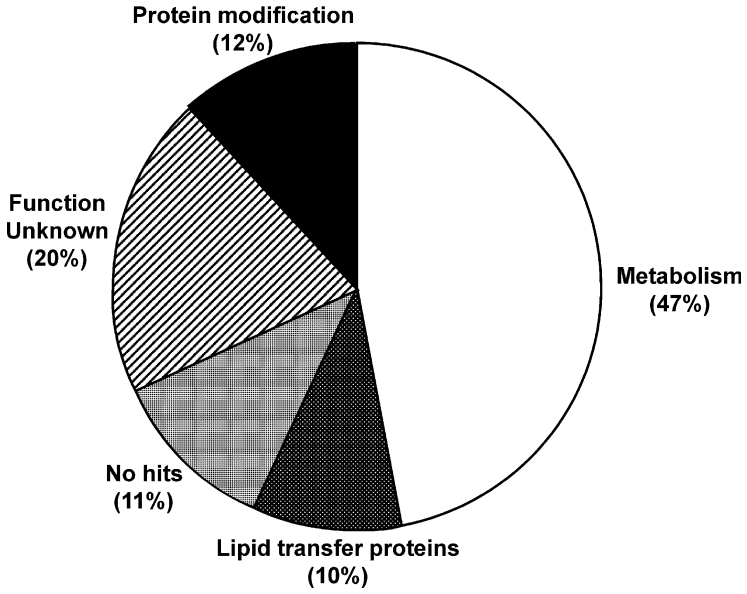

Fig. 3. Pie chart showing the representation of 900 ESTs from the *A. annua* trichome-specific cDNA library over five general functional categories. Categories are based on similarity with genes and proteins present in GenBank.

metabolism accounting for about 12% of metabolism-related clones. About 4% of the clones belonging to secondary metabolism encoded enzymes involved in flavonoid biosynthesis (4-coumarate-CoA ligase, *p*-coumarate 3-hydroxylase, flavonol synthase, flavonol-4-reductase, flavonoid-3′-5′-hydroxylase, and flavonol-*O*-methyltransferases). Up to 8% of the clones involved in metabolism encode enzymes and proteins related to isoprenoid biosynthesis (Table 1). This is a smaller fraction than in the peppermint glandular trichome library in which 25% of the ESTs were estimated to be involved in essential oil biosynthesis [14]. This is likely due to the fact that the basal cells of the *A. annua* glandular trichomes contain photosynthetically active chloroplasts—in contrast to the mint trichome cells that were completely devoid of chlorophyll [21]—which may be indicative of the presence of more general, primary, metabolism in addition to isoprenoid biosynthesis. Up to 3% of the *A. annua* trichome ESTs encoded proteins involved in photosynthesis, with chlorophyll *a*/*b*-binding protein being the highest expressed single cDNA.

Among the clones related to isoprenoid biosynthesis we found a number of sequences already isolated from this plant species before, but starting from whole leaf tissue. This includes linalool synthase [19], β-pinene synthase (GenBank Accession No. AF276072, [22]), and farnesyl diphosphate synthase (GenBank Accession No. AF112881, [23]). Amorpha-4,11-diene synthase was found as a singleton. Also a PCR using amorpha-4,11-diene synthase specific primers and the first strand cDNA employed for the library construction as a template showed a clear band of expected size, confirming that at least the first gene involved in artemisinin biosynthesis is expressed in the glandular trichomes (data not shown). In addition, we found ESTs related to the plastidic methylerythritol phosphate (MEP) pathway and the cytosolic mevalonate (MVA) pathway. The presence of both pathways that are supposed to be involved in monoterpene (MEP) and sesquiterpene (MVA) biosynthesis agrees with the presence of both groups of compounds in trichome extracts such as the monoterpenes camphor and linalool and the sesquiterpenes germacrene A, germacrene D, β-farnesene, selina-4,11-diene, amorphadiene (and, of course, artemisinin) (Fig. 1). The stronger representation of the MEP-pathway in the ESTs (Table 1) despite the fact that *A. annua* mainly produces sesquiterpenoids, may indicate that the MEP-pathway—in addition to the formation of monoterpenoids such as camphor and linalool (Fig. 1)—is also involved in the production of precursors for sesquiterpene biosynthesis in this plant species. This has been reported before for example for sesquiterpenes in chamomile [24] and lima bean [25]. More recently, the cross-talk between the two pathways was also demonstrated in *Arabidopis thaliana* [26].

Table 1
Selection of isoprenoid biosynthesis related

EST identification	No. of hits
Deoxyxylulose 5-phosphate synthase	3
Isopentenyl diphosphate isomerase	1
3-Hydroxy-3-methylglutaryl-coenzyme A reductase	1
2, 4-*C*-methyl-D-erythritol cyclodiphosphate synthase	3
Geranyl diphosphate synthase (small subunit)	1
β-Pinene synthase	3
Limonene synthase	2
Linalool synthase	8
Farnesyl diphosphate synthase	1
Germacrene A synthase	2
Amorpha-4,11-diene synthase	1
Other sesquiterpene synthases	4

ESTs and their tentative identification. ESTs are part of a collection of 900 ESTs generated from total RNA isolated from glandular trichomes of *Artemisia annua*. EST identification is based on comparison of the EST sequences with Genbank sequence information. No. of hits indicates how many copies of the particular EST were identified among the 900 clones that have been sequenced.

Our results are particularly significant when compared to those obtained with an *A. annua* total leaf cDNA library. The random sequencing of 4,224 clones from mRNA isolated from *A. annua* leaves resulted in the identification of just one clone involved in terpenoid biosynthesis [27]. Moreover, this clone was shown to encode *epi*-cedrol synthase, a sesquiterpene synthase that was already cloned from *A. annua* before but which could not be attributed any functional role *in planta* [6,7].

Another group of ESTs of high abundance was represented by lipid transfer proteins (10%) (Fig. 3). These proteins are believed to be involved in the transport of lipophilic molecules, such as essential oils, between different cell compartments and they were found in high abundance also in other glandular trichome cDNA libraries from aromatic plants [14,15]. A lipid transfer protein was also the most abundant EST in a *Medicago sativa* glandular trichome cDNA library [28]. A significant number of ESTs (31%) could not be assigned any putative function since they showed no similarity to any entry in the databases (11%) or because they showed similarity to unknown genes from other species (20%). This is in accordance with the results obtained with other gland cDNA libraries [14,15,28].

Among the isoprenoid biosynthesis related clones, two cDNAs showed significant sequence similarity with germacrene A synthases from chicory [29] and lettuce [30] reported in GenBank (Table 1). One of the two clones was full-length and it was used for heterologous expression in *E. coli*.

Sequence analysis

Full-length sequencing of the clone *AaGAS* resulted in a 1689-bp open reading frame encoding a protein of 562 amino acids showing all the motifs common to the sesquiterpene synthase family [31], such as the RPxxxxxxxxW motif at the N-terminal end of the sequence, and the (DDxxD) divalent metal ion-substrate binding motif that is highly conserved in terpene synthases [32]. According to the targeting signal prediction program PREDOTAR version 0.5 (http://www.inra.fr/Predotar) the predicted protein contained no cleavable transit peptide for plastid localization,

indicating that the germacrene A synthase has a cytosolic localization as expected for a sesquiterpene synthase. The *A. annua* germacrene A synthase falls within the *Tps-a* subfamily of the terpenoid synthases [31,33], that mainly consists of angiosperm sesquiterpene synthases [34,35].

Comparison of the deduced amino acid sequence of the *A. annua* germacrene A synthase with germacrene A synthases from other plant species showed a fairly high degree of similarity (Fig. 4A). In particular, AaGAS displays 81% similarity to germacrene A synthases from *Cichorium intybus*, CiGASsh, and *Lactuca sativa*, LsLTC1 and LsLTC2, but only 68% to the *C. intybus* CiGASlo. This is probably due to the fact that CiGASlo has a stretch of additional amino acids at the N-terminal, missing in AaGAS. The deduced amino acid sequences were further analyzed by the neighbour noining (NJ) method to infer the phylogenetic relationship between *A. annua* germacrene A synthase and sesquiterpene synthases from other plant species deposited in the NCBI GenBank. Fig. 4B shows the phylogenetic tree where the *A. annua* germacrene A synthase clusters into a separate group with the germacrene A synthases isolated from two other plant species belonging to the Asteraceae: *L. sativa* and *C. intybus*. Remarkably, AaGAS displays a higher similarity to these germacrene A synthases from different plant species than to the other sesquiterpene synthases isolated from *A. annua* whereas the germacrene A synthase cloned from another Asteraceae, *Solidago canadensis*, groups with these other *A. annua* sesquiterpene synthases. It is intriguing that the ability of *A. annua*, *L. sativa* and *C. intybus* to produce germacrene A diverged early during evolution from the ability to produce other sesquiterpenes in Asteraceae. An interesting exception is that *S. canadensis* evolved germacrene A biosynthesis independently. The other germacrene A synthases grouping with AaGAS are all involved in sesquiterpene lactone formation [12,29] whereas AaGAS is not. That is, *A. annua* does not contain any eudesmanolides, guaianolides, and germacranolides, sesquiterpene lactones that have been shown to be derived from germacrene A [12]. Interestingly, other *Artemisia* species, such as *Artemisia absinthium*, do contain the germacrene A derived guaianolides such as the extremely bitter dimer guaianolide absinthin (http://www.ggiz-erfurt.de/pdf/pub_qt_giz_absinth_2004.pdf). It is an intriguing question whether the functional role of germacrene A derived sesquiterpene lactones that are present in other Asteraceae in *A. annua* was replaced by artemisinin or artemisinin-pathway intermediates.

Pogostemon cablin germacrene A synthase (PcGAS) is grouped far away from the other germacrene A synthases. Functionally it is relevant that it groups together with Solanaceous (tobacco, pepper, potato, and tomato) sesquiterpene synthases such as *epi*-aristolochene synthase and vetispiradiene synthase that produce germacrene A as an enzyme-bound intermediate en route to their end-product [29]. The Solanaceous sesquiterpene synthases group together closely, with the exception of the tomato epidermal germacrene C synthase that clusters with the peppermint (*Mentha x piperita*) germacrene C synthase. The other two peppermint sesquiterpene synthases, β-farnesene synthase and the recently isolated *cis*-muuroladiene synthase [36], cluster together and have diverged relatively early from the other angiosperm sesquiterpene synthases (Fig. 4B).

Finally, the gymnosperm sesquiterpene synthases isolated from *Abies grandis* group in a separate cluster that diverged at an early stage from the angiosperm sesquiterpene synthases, as has been previously reported [31].

Functional expression of the putative germacrene A synthase in E. coli

The putative germacrene A synthase was functionally expressed in *E. coli* BL21-DE3-RIL cells. This strain contains the RIL-plasmid for expression of tRNA codons that are rare in *E. coli*, to give better expression and accumulation of the protein. Crude extracts of *E. coli* expressing the cloned sesquiterpene synthase displayed a high activity when incubated with FDP as substrate. GC–MS analysis of the reaction products demonstrated that the cDNA encoded an enzyme able to convert FDP to a single sesquiterpene product. At a GC–MS injection port temperature of 250 °C this product was identified as β-elemene (Fig. 5B), which was not produced in the reaction with the soluble fraction of *E. coli* harbouring the empty vector (Fig. 5A). However, from previous experience it was known that the presence of β-elemene usually is indicative of the formation of germacrene A that is rearranged upon GC–MS analysis using a high injection port temperature due to Cope rearrangement [12]. When analyzed using a lower injection port temperature of 150 °C almost no Cope rearrangement occurred and a broad peak of germacrene A was detected (data not shown), confirming that the enzyme is a germacene A synthase [12].

Many sesquiterpene synthases are characterised by the generation of one dominant product (>80% of total terpenes produced) along with traces of a varying number of side-products [4]. However, we were able to detect only a single product showing that the *A. annua* germacrene A synthase is highly specific, just as the *C. intybus* germacrene A synthases [29]. The enzyme was active only with FDP and not with GDP (data not shown).

Germacrene A is one of the constituents of the essential oil extracted from *A. annua* leaves and trichomes (Fig. 1) [1] and is one of the major products, along with amorpha-4,11-diene, in the trichome-specific enzyme assay using FDP as substrate (Fig. 2) This demonstrates that *AaGAS* is not an unimportant gene and that germacrene A is not a side product of an another sesquiterpene synthase (Fig. 5B).

Evidence for germacrene A synthase activity in plants has already been reported [12] as well as the cloning and characterization of genes encoding this enzyme from different plant species [29,30,37]. This is not surprising since there are many natural products derived from oxidative metabolism of germacrene A [12]. For example, germacrene A is the first dedicated intermediate in the biosynthesis of bitter sesquiterpenoids of chicory [12,29] and lettuce [30]. Ger-

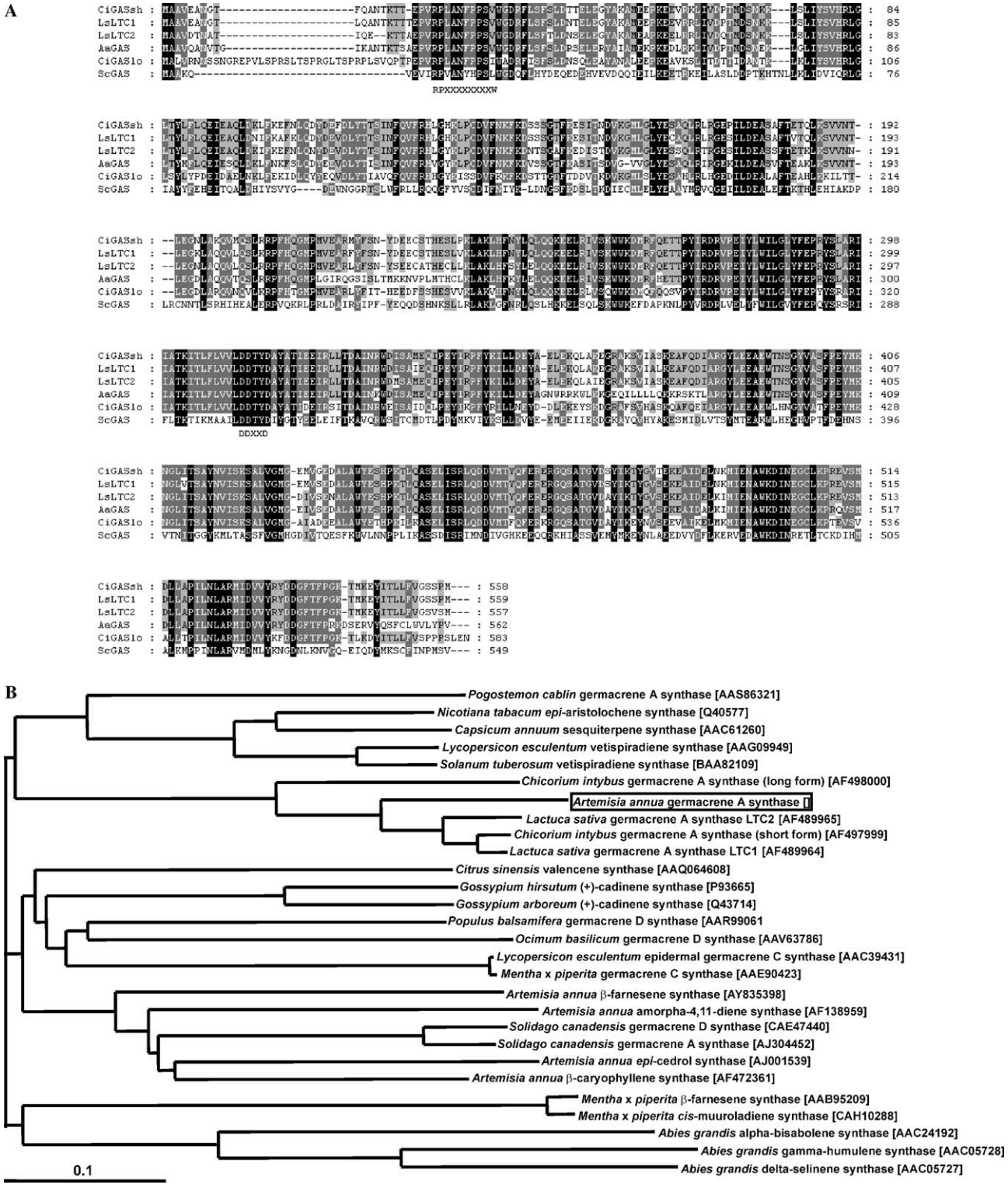

Fig. 4. (A) Aligment of deduced amino acid sequences of germacrene A synthase from *A. annua* and germacrene A synthases isolated from several other plant species. *AaGAS*: *A. annua* germacrene A synthase (GenBank Accession No. DQ447636); *CiGASsh*, *C. intybus* germacrene A synthase (short form) (GenBank Accession No. AF498000) [29]; *CiGASlo*, *Cichorium intybus* germacrene A synthase (long form) (GenBank Accession No. AF497999) [29]; *ScGAS*, *S. canadensis* germacrene A synthase (GenBank Accession No. AJ304452) [37]; *LsLTC1*, *L. sativa* germacrene A synthase (GenBank Accession No. AF489964) [30]; *LsLTC2*, *L. sativa* germacrene A synthase (GenBank Accession No. AF489965) [30] The alignment was created with the Clustal X 1.83 program and standard settings. Shading indicates conserved identity for the aligned amino acids: black background shading indicates 100% conservation, dark grey shading indicates 80% conservation, and light grey shading indicates 60% conservation. (B) Phylogenetic analysis of *A. annua* germacrene A synthase and sesquiterpene synthases isolated from other plant species. Scale bar, 0.1 is equal to 10% sequence divergence. The divergence calculations are based on amino acid sequences.

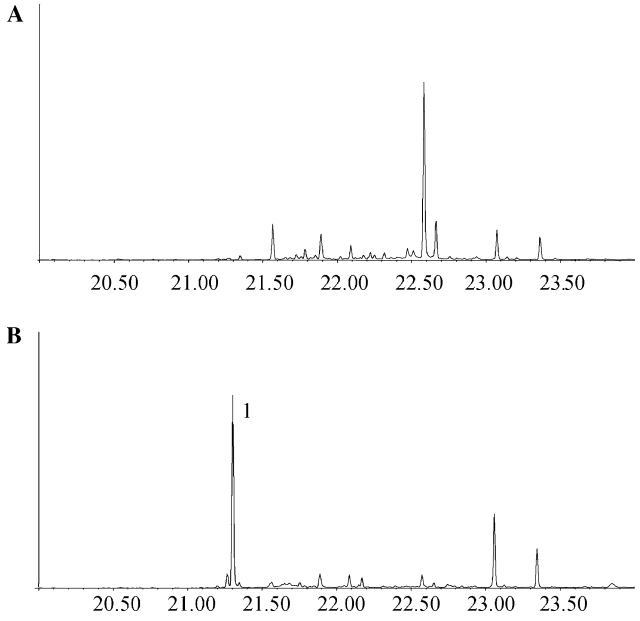

Fig. 5. GC–MS profiles of products formed by the heterologously expressed putative germacrene A synthase cloned from *A. annua* and using farnesyl diphosphate as substrate. (A) Empty pRSET A vector control, (B) pRSET A harboring the putative germacrene A synthase. Peak 1 at 21.30 min was identified as (−)-β-elemene, the Cope rearrangement product of germacrene A as induced by the high injection port temperature. The peak identity was confirmed using an authentic standard of germacrene A, kindly provided by the late prof. W.A. König, of which the mass spectrum and retention time exactly matched this product.

macrene A has also been found in extracts derived from several other species such as *S. canadensis* [37], *Carum carvi* and the liverwort *Frullania tamarisci* [12]. As described above, *A. annua* and *S. canadensis* do not contain sesquiterpene lactones in contrast to many other Asteraceae. Both species seem to miss the downstream conversion mechanism that is responsible for the formation of sesquiterpene lactones in *C. intybus* and *L. sativa* and hence can only accumulate germacrene A that is vice versa not detected in the latter two species [12]. The fact that germacrene A biosynthesis is nevertheless one of the major activities in *A. annua* (Fig. 2) suggests that the compound plays an important role, for example in plant defense. Further studies should demonstrate the significance of this role.

Conclusions

In this study, we used a combination of chemical, biochemical, and genomics approaches to show that terpenoids are not only stored but also synthesized in *A. annua* glandular trichomes. Analysis of EST databases constructed from a single type of cell in conjunction with other types of data, such as metabolic profiling, is a powerful method to identify major biochemical pathways operating in the cell. This approach was previously employed to identify genes involved in secondary metabolism by constructing and comparing EST databases from peltate gland cells of basil [15] and peppermint [14].

In the present study, we used isolated *A. annua* gland secretory cells for cDNA library construction. The random sequencing of about 900 clones indicates that *A. annua* glandular trichomes indeed express genes with functions in production of secondary metabolites (terpenoid and flavonoid biosynthesis). This material provides a superior source for the retrieval of biosynthetic genes when compared to previous studies that used whole leaves as a starting material [27]. Our approach led to the cloning of a germacrene A synthase and this is the first time that the isolation of a sesquiterpene synthase from *A. annua* glandular trichomes has been reported.

Our results show that the use of isolated gland secretory cells is a good tool for studying isoprenoid metabolism in *A. annua*, which is particularly relevant for unraveling artemisinin biosynthesis. The cloning of artemisinin biosynthetic genes should help in obtaining better production systems for this important anti-malarial medicine. The random sequencing of 900 ESTs has provided a number of candidate artemisinin biosynthetic genes that we are currently further characterizing.

Acknowledgments

We thank Anna-Margareta Ryden for helpful comments. This work was in part supported by a Marie Curie Individual Fellowship (MCFI-2000-52058) to Cinzia M. Bertea.

References

[1] H.J. Bouwmeester, T.E. Wallaart, M.H.A. Janssen, B. van Loo, B.J.M. Jansen, M.A. Posthumus, C.O. Schmidt, J.-W. de Kraker, W.A. König, M.C.R. Franssen, Phytochemistry 52 (1999) 843–854.
[2] C.M. Bertea, J.R. Freije, H. van der Woude, F.W.A. Verstappen, L. Perk, V. Marquez, J.-W. de Kraker, M.A. Posthumus, B.J.M. Jansen, Ae. de Groot, M.C.R. Franssen, H.J. Bouwmeester, Planta Med. 71 (2005) 40–47.
[3] Y.J. Chang, S.H. Song, S.H. Park, S.U. Kim, Arch. Biochem. Biophys. 383 (2000) 178–184.
[4] P. Mercke, M. Bengtsson, H.J. Bouwmeester, M.A. Posthumus, P.E. Brodelius, Arch. Biochem. Biophys. 381 (2000) 173–180.
[5] T.E. Wallaart, H.J. Bouwmeester, J. Hille, L. Poppinga, N.C. Maijers, Planta 212 (2001) 460–465.
[6] P. Mercke, J. Crock, R. Croteau, P.E. Brodelius, Arch. Biochem. Biophys. 369 (1999) 213–222.
[7] L. Hua, S.P. Matsuda, Arch. Biochem. Biophys. 369 (1999) 208–212.
[8] Y. Cai, J.-W. Jia, J. Crock, Z.-X. Lin, X.-Y. Chen, R. Croteau, Phytochemistry 61 (2002) 523–529.
[9] S. Picaud, M. Brodelius, P.E. Brodelius, Phytochemistry 66 (2005) 961–967.
[10] M.V. Duke, R.N. Paul, H.N. Elsohly, G. Sturtz, S.O. Duke, Int. J. Plant Sci. 155 (1994) 365–372.
[11] M.R. Tellez, C. Canel, A.M. Rimando, S.O. Duke, Phytochemistry 52 (1999) 1035–1040.
[12] J.-W. de Kraker, M.C.R. Franssen, Ae. de Groot, W.A. Koenig, H.J. Bouwmeester, Plant Physiol. 117 (1998) 1381–1392.
[13] S.F. Altschul, T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.
[14] M.B. Lange, M.R. Wildung, E.J. Stauber, C. Sanchez, D. Pouchnik, R. Croteau, Proc. Natl. Acad. Sci. USA 97 (2000) 2934–2939.
[15] D.R. Gang, J. Wang, N. Dudareva, K.H. Nam, J.E. Simon, E. Lewinsohn, E. Pichersky, Plant Physiol. 125 (2001) 539–555.

[16] S.O. Duke, R.N. Paul, Int. J. Plant Sci. 154 (1993) 107–118.
[17] C.C. Burke, M.R. Wildung, R. Croteau, Proc. Natl. Acad. Sci. USA 96 (1999) 13062–13067.
[18] Y.L. Meng, J.W. Jia, C.J. Liu, W.Q. Liang, P. Heinstein, X.-Y. Chen, J. Nat. Prod. 62 (1999) 248–252.
[19] J.W. Jia, J. Crock, S. Lu, R. Croteau, X.Y. Chen, Arch. Biochem. Biophys. 372 (1999) 143–149.
[20] E. van Geldre, I. de Pauw, D. Inze, M. van Montagu, E. van den Eeckhout, Plant Sci. 158 (2000) 163–171.
[21] G.W. Turner, J. Gershenzon, R.B. Croteau, Plant Physiol. 124 (2000) 665–679.
[22] S. Lu, R. Xu, J.W. Jia, J. Pang, S.P.T. Matsuda, X.-Y. Chen, Plant Physiol. 130 (2002) 477–486.
[23] D. Chen, H. Ye, G. Li, Plant Sci. 155 (2000) 179–185.
[24] K.P. Adam, R. Thiel, J. Zapp, Arch. Biochem. Biophys. 369 (1999) 127–132.
[25] J. Piel, J. Donath, K. Bandemer, W. Boland, Angew. Chem. 37 (1998) 2478–2481.
[26] O. Laule, A. Fürholz, H.-S. Chang, T. Zhu, X. Wang, P.B. Heifetz, W. Gruissem, M.B. Lange, Proc. Natl. Acad. Sci. USA 100 (2003) 6866–6871.
[27] B.H. Nam, S.-M. Kim, Y.-J. Chang, S.-U. Kim, Abstract Meeting Phytochemistry in the Genomics and Post-genomics Era (2001). Phytochemical Society of North America (PSNA), Oklahoma, USA.
[28] A. Naveed, N.L. Paiva, G.D. May, R.A. Dixon, Planta 221 (2005) 28–38.
[29] H.J. Bouwmeester, J. Kodde, F.W.A. Verstappen, I.G. Altug, J.-W. de Kraker, T.E. Wallaart, Plant Physiol. 129 (2002) 134–144.
[30] M.H. Bennett, J.W. Mansfield, M.J. Lewis, M.H. Beale, Phytochemistry 60 (2002) 255–261.
[31] J. Bohlmann, G. Meyer-Gauen, R. Croteau, Proc. Natl. Acad. Sci. USA 95 (1998) 4126–4133.
[32] E.M. Davis, R. Croteau, in: F. Leeper, J.C. Vederas (Eds.), Topics in Current Chemistry: Biosynthesis-Aromatic Polyketides, Isoprenoids, Alkaloids, Springer-Verlag, Heidelberg, 2000, pp. 53–95.
[33] S.C. Trapp, R.B. Croteau, Genetics 158 (2001) 811–832.
[34] J. Crock, M. Wildung, R. Croteau, Proc. Natl. Acad. Sci. USA 94 (1997) 12833–12838.
[35] S. Aubourg, A. Lecharny, J. Bohlmann, Mol. Gen. Genet. 267 (2002) 730–745.
[36] I.M. Prosser, R.J. Adams, M.H. Beale, N.D. Hawkins, A.L. Phillips, J.A. Pickett, L.M. Field, Phytochemistry, (2006), in press.
[37] I. Prosser, A.L. Phillips, S. Gittings, M.J. Lewis, A.M. Hooper, J.A. Pickett, M.H. Beale, Phytochemistry 60 (2002) 691–702.