Cytochrome P450s from Cynara cardunculus L. CYP71AV9 and CYP71BL5, catalyze distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway

Cytochrome P450s from *Cynara cardunculus* L. CYP71AV9 and CYP71BL5, catalyse distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway

Kaouthar Eljounaidi[a] , Katarina Cankar[b,c] , Cinzia Comino[a] , Andrea Moglia[a] , Alain Hehn[d,e] , Frédéric Bourgaud[d,e], Harro Bouwmeester[c] , Barbara Menin[f] , Sergio Lanteri[a] , Jules Beekwilder [b]

[a] Dipartimento di Scienze Agrarie, Forestali e Alimentari, University of Torino, Via L. da Vinci 44, 10095 Grugliasco (Italy)
[b] Plant Research International, P.O. Box 16, 6700 AA Wageningen (The Netherlands)
[c] Laboratory of Plant Physiology, Wageningen University, P.O. Box 658, 6700AR Wageningen (The Netherlands)
[d] Université de Lorraine, UMR 1121 Agronomie et Environnement, 2 avenue de la Forêt de Haye, TSA 40602, 54518 Vandoeuvre-lès-Nancy (France)
[e] INRA, UMR 1121 Agronomie et Environnement, 2 avenue de la Forêt de Haye, TSA 40602, 54518 Vandoeuvre-lès-Nancy (France)
[f] PTP/Rice Genomics Unit, Via Einstein, 26900 Lodi (Italy)

Corresponding author:

Dr. Jules Beekwilder

e-mail: jules.beekwilder@wur.nl

Fax:+31317418094. Tel: 0317-480979

Plant Research International, P.O. Box 16, 6700 AA Wageningen (The Netherlands)

ABSTRACT

Cynara cardunculus (Asteraceae) is a cross pollinated perennial crop which includes the two cultivated taxa globe artichoke and cultivated cardoon. The leaves of these plants contain high concentrations of sesquiterpene lactones (STLs) among which cynaropicrin is the most represented, and has recently attracted attention because of its therapeutic potential as anti-tumor and anti-photoaging agent. Costunolide is considered the common precursor of the STLs and three enzymes are involved in its biosynthetic pathway: i.e. the germacrene A synthase (GAS), the germacrene A oxidase (GAO) and the costunolide synthase (COS). Here we report on the isolation of two P450 genes, (i.e. CYP71AV9 and CYP71BL), in a set of ~19,000 *C. cardunculus* unigenes, and their functional characterization in yeast and *in planta*. The metabolite analyses revealed that the co-expression of CYP71AV9 together with GAS resulted in the biosynthesis of germacra-1(10),4,11(13)-trien-12-oic acid in yeast. The co-expression of CYP71BL5 and CYP71AV9 with GAS led to biosynthesis of the free costunolide in yeast and costunolide conjugates in *N. benthamiana,* demonstrating their involvement in STLs biosynthesis as GAO and COS enzymes. The substrate specificity of CYP71AV9 was investigated by testing its ability to convert amorpha-4,11-diene, (+)-germacrene D and cascarilladiene to their oxidized products when co-expressed in yeast with the corresponding terpene synthases.

Keywords

Cynara cardunculus, sesquiterpene lactones, P450 enzymes, germacrene A, *N. benthamiana* transient expression, costunolide.

1. Introduction

Cynara cardunculus L (2n = 2x = 34) is a cross-pollinated *Asteraceae* species which comprises three taxa: globe artichoke (var. *scolymus* L.), exploited for the production of immature inflorescences (heads or capitula), cultivated cardoon (var. *altilis* DC), exploited for the production of fleshy stems, and their common progenitor the wild cardoon [var. *sylvestris* (Lamk) Fiori] [1,2]. *Cynara cardunculus* leaves are rich in secondary metabolites, and its extracts have been used since ancient times as anti-oxidants and hepatoprotectors [3,4,5] and anticholesteremic agents [6].

The main secondary metabolites in *C. cardunculus* leaves are caffeoylquinic acid derivatives (5-O-caffeoylquinic acid commonly known as chlorogenic acid, and several dicaffeoylquinic acid isomers), flavonoids (luteolin and apigenin) and sesquiterpene lactones (STLs), such as cynaropicrin [7] and, at lower concentration, grosheimin and its derivatives [8,9,10,11]. STLs play several ecophysiological roles as allelochemicals, insect repellents and animal allergens [12,13] and have been identified in several plant families. STLs are highly abundant in some members of the *Asteraceae* family [14].

The STLs are responsible of the peculiar bitter taste of both globe artichoke heads and cultivated cardoon stems [15], and have been found to represent the major family of lipophilic components in cultivated cardoon leaves (95 g kg^{-1} dry weight) [16]. Furthermore, it has been demonstrated they possess a broad range of pharmacological properties [17,18,19,20,21], among which cytotoxicity against several types of cancer cells [22] as well as anti-photoaging [23] and anti-inflammatory activities [24]. Recently the effectiveness of cynaropicrin against acute gastritis has been proved as well [15].

Although the detailed structure of the STLs varies across the *Asteraceae* family, their backbones, derived from the farnesyl diphosphate (FPP), are constrained to a limited set of core skeletons, such as germacranolide, eudesmanolide and guaianolide of which costunolide is generally considered the common precursor [25,26,27,28,29].

The biosynthetic pathway of costunolide was proposed by de Kraker et al. [30, 31] upon isolation of enzymes in chicory roots able to convert the FPP into costunolide (Fig. 1). The authors highlighted that FPP is cyclized to C15 hydrocarbon germacrene A by germacrene A synthase (GAS); afterwards three-steps oxidation of the methyl group at C12 of germacrene A,

by the cytochrome P450 germacrene A oxidase (GAO), leads to the formation of germacrene A acid (germacra-1(10),4,11(13)-trien-12-oic acid) [32]. The latter is then hydroxylated by (+)-costunolide synthase (COS) into an unstable intermediate that undergoes a non-enzymatic lactonization of C6 hydroxyl and C12 carboxylic group, yielding costunolide [33].

The GAS, GAO and COS enzymes involved in this pathway have been isolated and characterized in some *Asteraceae* species like lettuce [32,33], sunflower [32,33], pyrethrum [34] and chicory [35,36,37]. In globe artichoke we previously isolated the GAS enzyme [11], but the downstream pathway, leading to the synthesis of cynaropicrin, has still to be elucidated.

In the present work we investigated the catalytic activity of *C. cardunculus* candidate P450s involved in the STLs pathway. The candidate GAO and COS genes (CcGAO and CcCOS) were identified, cloned and their sequences were compared to the ones of the previously isolated orthologs in other *Asteraceae* species. Through their expression in both yeast and tobacco, followed by GC-MS and LC-MS metabolic profiling, we performed their functional characterization. Additionally, the cross-reactivity of the CcGAO toward other terpenes, such as the cascarilladiene, (+)-germacrene D and the amorpha-4,11-diene, was assessed through its heterologous expression in yeast alongside the corresponding terpene synthases.

2. Materials and methods

2.1 Gene cloning and sequence analysis

Cichorium intybus germacrene A oxidase (GAO, Accession Number GU256644) and costunolide synthase (COS, JF816041.1) were used as BLAST search terms to query the *Cynara cardunculus* L. unigene database containing ~19,000 sequences [38]. The full length of GAO and COS candidate genes were amplified from globe artichoke (*Cynara cardunculus* L. var. *scolymus*) leaves cDNA using high fidelity Phusion polymerase (Finnzymes) with the addition of restriction sites *Bam*HI/*Eco*RI and *Not*I/*Pac*I, respectively. The amplification products (CcGAO and CcCOS) were cloned into pGEM-T Easy (Promega), using standard restriction and ligation methods, and sequenced. The CYP number of the isolated genes was provided by Dr Nelson (http://drnelson.uthsc.edu/CytochromeP450.html). Gene isolation was also performed from cultivated cardoon (*Cynara cardunculus* L. var. *altilis*) leaves (CcaGAO and CcaCOS). The cytochrome P450 protein sequences were compared through the ClustalW algorithm

(www.ebi.ac.uk/Tools/clustalw2/index.htmL) [39] using standard settings and aligned by BioEdit software. A phylogenetic analysis was conducted using MEGA5.2 software [40]. The neighbour joining method was used to construct a guided tree, supported by bootstrapping based on 5000 replicates.

2.2 Gene expression in yeast

The full length CcGAO was sub-cloned, using the *Bam*HI/*Eco*RI restriction sites, into the yeast expression vector pYeDP60 with adenine (ADE2) and uracil (URA3) as auxotrophic selection markers [41] generating the pYeDP60-CcGAO plasmid. This vector was co-transformed into yeast strain WAT 11 expressing the Arabidopsis ATR1 NADPH-cytochrome P450 reductase [42] together with different terpene synthases already characterized in other species: *i.e.* the chicory germacrene A synthase (CiGAS) [35], the *Artemisia annua* amorpha-4,11-diene synthase (AaAMDS) [43], the Goldenrod (+)-germacrene D synthase (ScGDS) [44] or goldenrod cascarilladiene synthase (ScCS) [45]. CiGAS was cloned into pYeDP80 vector [41] with a tryptophan (TRP1) auxotrophic selection marker using *Bam*HI and *Eco*RI restriction sites generating pYeDP80-CiGAS plasmid. AaAMDS was cloned into pYES3/CT yeast expression vector (Invitrogen) with TRP1 selection marker using *Bam*HI and *Not*I restriction sites leading to the formation of AaAMDS-pYES3/CT vector. ScGDS and ScCS were both cloned into pYES3/CT using *Sac*I/*Not*I restriction sites originating ScGDS-pYES3/CT and ScCS-pYES3/CT, respectively. After the transformation, yeast clones containing the CcGAO and the terpene synthases were selected on Synthetic Dextrose (SD) minimal medium (0.67% Difco yeast nitrogen base medium without amino acids, 2% d-glucose, 2% agar) supplemented with amino acids, but omitting uracil, adenine sulphate and L-tryptophan for auxotrophic selection of transformants.

The full length CcCOS was sub-cloned, using *Not*I/*Pac*I restriction sites, into the yeast expression vector pYeDP60 which was modified in order to contain *Pac*I and *Not*I sites at the polylinker originating pYEDP60-CcCOS vector. This plasmid was co-transformed with the vector pESC-Trp-TpGAS-CiGAO, that contains *Tanacetum parthenium* germacrene A synthase (TpGAS) and *Cichorium intybus* germacrene A oxidase (CiGAO), previously described [37], into the WAT11 yeast strain [42] and the clones were selected on SD minimal medium (0.67%

Difco yeast nitrogen base medium without amino acids, 2% D-glucose, 2% agar) supplemented with amino acids, but omitting uracil, adenine sulphate and L-tryptophan for auxotrophic selection of transformants.

2.3 Yeast culture and sample preparation for the metabolic profiling

For the induction of gene expression in yeast, the transformed yeast strain of interest was inoculated into 5 mL of Synthetic Galactose (SG) minimal medium (0.67% Difco yeast nitrogen base medium without amino acids, 2% D-galactose) omitting the appropriate amino acids. The inocula were cultured overnight at 30°C at 300 rpm. The starter culture was then diluted to an OD_{600} value of 0.05 in SG minimal medium and incubated for 48 h at 30°C and 300 rpm. Two conditions were tested: acidic conditions (the medium pH is initially 6 and then decreases after 48 h of cultivation) and an adjusted pH medium (addition of 100 mM of HEPES/NaOH pH 7.5, to maintain the culture medium pH>6). All yeast induction experiments were performed in triplicates in 50 mL of culture volume. After 48 h of fermentation, the culture medium was extracted with 20 mL ethyl acetate. The ethyl acetate fractions were evaporated with a stream of N_2 to a final volume of 1 mL which was dehydrated using anhydrous Na_2SO_4 and analysed by GC-MS. For analysis of artemisinic acid, samples were first methylated using diazomethane.

2.4 GC-MS analyses

Metabolites from yeast cultures were analysed through a gas chromatography system (7890A, Agilent Technologies, USA) equipped with ZB-5 column (30 m×0.25 mm inner diameter, 0.25 mm film thickness, Phenomenex) using helium as carrier gas at flow rate of 1 mL min^{-1}. The sample was injected into the device chamber at 250°C; the injection was performed in splitless mode and the ZB5 column was maintained at 45°C for 1 min, after which the temperature was increased at a rate of 10°C per min until reaching 300°C and then held for 5 min at 300°C. The GC was coupled to a Triple-Axis detector (5975C, Agilent Technologies, USA).

Quantification of sesquiterpenoids was conducted by determination of total ion count (TIC) area of the sesquiterpenoid peaks from three independent fermentation experiments. Germacrene A and germacra-1(10),4,11(13)-trien-12-oic acid thermally converted into β-elemene and elematrien-12-oic acid, respectively, at the routine injection port temperature of 250°C as discussed by de Kraker et al. [30,31,46].

2.5 Plasmid construction for gene expression in *Nicotiana benthamiana*

For the expression of CcGAS in *Nicotiana benthamiana*, three different constructs were prepared, each of these three allows for targeting of the CcGAS to a particular host plant sub-cellular compartment within the plant cell. CcGAS coding sequence was first cloned into the Impact Vector 1.1 (http://www.impactvector.com) that has no signaling peptide allowing for cytosolic accumulation. Then, it was cloned in the Impact Vector 1.4 that harbors a signal peptide from *Chrysanthemum morifolium* small subunit which targets the CcGAS into the stroma of the chloroplast and finally in the Impact Vector 1.5 that uses the yeast CoxIV (cytochrome oxidase subunit IV) secretion signal to target the CcGAS into the mitochondrial matrix. For the analysis of enzymatic activity, the putative CcGAO and CcCOS coding sequences were cloned into the Impact Vector 1.1. The different candidate genes sequences were then transferred by Gateway LR reaction using LR Clonase Enzyme to pBin-Dest binary vector (Invitrogen). Each of the prepared expression vectors was introduced into *Agrobacterium tumefaciens* AGL0 strain by electro-transformation.

2.6 Transient expression in *Nicotiana benthamiana*

The transient expression of the candidate genes in *N. benthamiana* was performed by agro-infiltration as described in van Herpen *et al.* [47]. Recombinant *A. tumefaciens* strains were grown at 28°C and 220 rpm for 24 h in LB media containing kanamycin (50 mg L^{-1}) and rifampicin (34 mg L^{-1}). After 20 min of centrifugation at 4000 g at 20°C, harvested cells were resuspended into 10 mM MES buffer containing 10 mM $MgCl_2$ and 100 µM acetosyringone (4′-hydroxy-3′,5′-dimethoxyacetophenone, Sigma) to a final OD_{600} of 0.5, and incubated at room temperature under gentle shaking at 50 rpm for 2 h. For co-infiltration, equal volumes of different recombinant *Agrobacterium* strains were mixed. Strain mixtures were infiltrated into leaves of 5 weeks old *N. benthamiana* plants using a 1 mL syringe. Bacteria were slowly injected into the abaxial side of the leaf. After agro-infiltration the plants were grown under greenhouse conditions for 5 additional days and then harvested for analysis.

2.7 Headspace analysis and GC-MS thermodesorption

Volatile collection from agro-infiltrated *N. benthamiana* leaves and GC-MS analysis were performed according to van Herpen et al. [47]. Steel sorbent cartridges (89 mm×6.4 mm O.D.; Markes) containing Tenax for volatile collection were conditioned at 280°C for 40 min under a nitrogen flow of 30 psi in a TC-20 multi-tube conditioner and were kept airtight with brass caps until use. Freshly collected *N. benthamiana* leaves were placed in water in a small vial and were enclosed in a glass container. To trap the leaf-produced volatiles, air was sucked through the Tenax cartridge with a flow rate of 100 mL min^{-1} for 4 h. Incoming air was purified with a second Tenax cartridge. Cartridges were dried for 15 min at room temperature with a nitrogen flow of 15 psi before GC-MS analysis. Headspace samples were analyzed with a Thermo Trace GC Ultra connected to a Thermo TraceDSQ quadruple mass spectrometer (Thermo Fisher Scientific, USA).

Cartridges were placed in an automated thermodesorption unit (Ultra; Markes, Llantrisant) in which they were flushed with helium at 50 mL min^{-1} for 2 min to remove moisture and oxygen just before thermodesorption. After flushing, the collected volatiles were desorbed from the Tenax cartridge by heating at 220°C for 5 min with a helium flow of 50 mL min^{-1}. The released compounds were focused on an electrically cooled sorbent trap (Unity; Markes, Llantrisant) at a temperature of 5°C. Subsequently, the trapped volatiles were injected on the analytical column (ZB-5MSI, 30 m×0.25 mm ID, 1.0 µm – film thickness, Zebron, Phenomenex) in splitless mode by ballistic heating of the cold trap to 250°C for 3 min. The temperature program of the GC started at 40°C (3 min hold) and rose 10°C min^{-1} to 280°C (2 min hold). The column effluent was ionized by electron impact (EI) ionisation at 70 eV. Mass scanning was done from 33 to 280 *m/z* with a scan time of 4.2 scans s^{-1}. The eluted compounds were identified using Xcalibur software (Thermo, Waltham) by comparing the mass spectra with those of authentic reference standards.

2.8 Glycosidase treatment

To verify if the germacrene A was modified by glycosylation, a treatment with glycosidase to release germacrene A related compounds was performed according to the protocol of Herpen et

al. [47]. After grinding in liquid nitrogen 200 mg of infiltrated leaf material from each experiment and extracting it with 1 mL citrate phosphate buffer, pH 5.4, the extracts were briefly vortexed and sonicated once for 15 min. Then 200 mL of ViscozymL (Sigma) were added and the mixture was vortexed, incubated overnight at 37°C, and subsequently extracted three times with 1 mL of ethyl acetate. Extracts were then dehydrated using Na_2SO_4, concentrated to approximately 250 mL and methylated using diazomethane right before the GC-MS analysis.

3. Results

3.1 Cloning and isolation of P450s genes

A bioinformatic screening of the *Cynara cardunculus* unigene database, using *Cichorium intybus* germacrene A oxidase (CiGAO, GU256644) and costunolide synthase (CiCOS, JF816041.1) as query, was performed. This analysis resulted in the identification of homologous genes with >60% amino acid identity: for CiGAO four genes were identified, while for CiCOS only one was observed. For each, the ortholog with the highest identity was further investigated. The CcGAO open reading frame was defined to be 1,461 bp long (486 residue protein, CYP71AV9, accession number KF752448), while the CcCOS coding sequence has 1,485 bp (494 residue protein, CYP71BL5, accession number KF752451).

The predicted amino acid sequences of the isolated genes present the heme-binding region at its C-terminal region, a characteristic motif of eukaryotic P450s. Both GAO and COS genes have been also fully isolated in cultivated cardoon genotype (Accession Numbers KF752449 and KF752452, respectively) as well.

The deduced amino acids of the candidate genes shared significant sequence identities with orthologous genes from *Asteraceae* family [32,33,34,35,36,37]. CcGAO has 95% degree of homology with GAO from *Saussurea costus* (ADF43081.1) and 89% with GAO from *Lactuca sativa* (ADF32078.1) (Table S1,2; Fig. S1).

The CcCOS protein shows the highest homology with both COS of *Cichorium intybus* (AEG79727.1) and COS of *Lactuca sativa* (AEI59780.1) with 88% amino acid identity. Fig. 2 reports the phylogenetic relationship of CcGAO and CcCOS with orthologous sequences in other *Asteraceae* species.

3.2 *In vitro* functional characterization of GAO and COS

The full length cDNA sequence of CcGAO was co-expressed in *S. cerevisiae* with the previously isolated germacrene A synthase of chicory (CiGAS) under both acidic conditions or in buffered yeast medium. The metabolites were extracted and subjected to gas chromatograph mass spectroscopy (GC-MS) analysis.

In acidic conditions, the biologically synthesized terpenoids were trapped by a dodecane overlay on the yeast cultures and two new compounds were detected, with respect to yeast cultures expressing the empty vector or CiGAS alone. The mass fragmentation patterns of the two detected peaks matched with those of elema-1,3,11(13)-trien-12-al (retention time of 15.48 min) and the elema-1,3,11(13)-trien-12-ol (retention time of 17.41 min), characterized by a parental mass of *m/z* 218 and *m/z* 220, respectively (Fig. 3A). On the other hand, in buffered conditions (in the ethyl acetate extracts), a unique compound: elema-1,3,11(13)-trien-12-oic acid (retention time of 17.41 min), with parental mass of *m/z* 234,was identified (Fig. 3B,3D). The formation of elemene-derivatives is known to occur in the GC-MS instrument from the corresponding germacrene A derivatives [31]. Therefore the observed MS spectra were interpreted to represent germacra-1(10),4,11(13)-trien-12-ol, germacra-1(10),4,11(13)-trien-12-al and germacra-1(10),4,11(13)-trien-12-oic acid.

The open reading frame of CcCOS was cloned into the pYeDP60 expression vector and tested in the yeast expression system through co-transformation with TpGAS and CiGAO from *Tanacetum parthenium* and *Cichorium intybus*, respectively. The LC-MS analyses of the yeast culture organic extracts revealed a product that co-eluted with a standard (+)-costunolide at 39.21 min, while no signal was detected in extracts of the yeast transformed with the empty vector, TpGAS and CiGAO (Fig. 3C,3D).

3.3 Germacrene A production *in planta*

The previously isolated globe artichoke germacrene A synthase (CcGAS) [11] was expressed *in N. benthamiana.* In order to maximize production of germacrene A, the CcGAS open reading frame was fused to different subcellular targeting signals. Its open reading frame was cloned into

a binary expression vector under the control of the Rubisco promoter using cytosol (cCcGAS), mitochondria (mCcGAS), or chloroplast (cpCcGAS) targeting signals (Fig. 4A) and the vector introduced into *A. tumefaciens*. The metabolic content of *N. benthamiana* leaves was assessed 5 days after the agro-infiltration [47].

Regardless of the sub-cellular compartment targeted, in leaves agro-infiltrated with *A. tumefaciens* carrying the CcGAS, we detected a peak whose mass spectrum matched with this of β-elemene, the heat-induced cope rearrangement product of germacrene A with a retention time of 13.40 min and a parental mass of *m/z* 204 (Fig. 4B). This peak was not observed in the headspace analysis of empty vector agro-infiltrated leaves.

In leaves infiltrated with *A. tumefaciens* carrying mCcGAS and cpCcGAS the germacrene A production was comparable (Fig. 4C), while leaves infiltrated with cCcGAS showed, on average, 3 fold higher germacrene A production. Thus cCcGAS was used for the further experiments in *N. benthamiana.*

3.4 Heterologous expression of STLs related genes *in Nicotiana benthamiana*

Agrobacteria transformed with an expression vector containing cCcGAS and CcGAO were infiltrated into *N. benthamiana* leaves. Plants agro-infiltrated only with CcGAO or with the empty vector were used as negative controls. Five days after the inoculation, the agro-infiltrated leaves were collected for the headspace trapping and GC-MS analysis. In the leaves agro-infiltrated with CcGAS and CcGAO, the peak of germacrene A, although detectable was strongly decreased, suggesting that the CcGAO converted the germacrene A to other products (Fig. 5A). However, no detectable new product peaks appeared in the headspace (Fig. 5A) or in the dichloromethane extracts of *N. benthamiana* agro-infiltrated leaves.

In order to check the presence of germacrene A conjugates in the leaves of *N. benthamiana*, a glycosidase treatment was performed to release glycosyl conjugates. After this treatment no new peaks were detected by GC-MS by comparing the leaves agro-infiltrated with CcGAS/CcGAO with those agro-infiltrated with only CcGAS.

CcGAS, CcGAO and CcCOS were transiently co-expressed through agro-infiltration in 5-week-old *N. benthamiana* leaves. After 5 days, methanol extracts were prepared and analyzed using LC-QTOF-MS. Chromatograms obtained from leaves infiltrated with CcGAS, CcGAO and

CcCOS showed two new compounds at 22.16 and 22.50 min (Fig. 5B), which were not observed when the infiltration was performed with the empty plasmid or with CcGAS and CcGAO. The MS spectra suggest that these compounds are most likely conjugates of costunolide. The parent masses of the two peaks, 354.14 (at 22.16 min) and 540.21 (at 22.50 min), were within 6 and 22 ppm from the elemental formulas of costunolide-cysteine ($C_3H_7NO_2S$) and costunolide-glutathione ($C_{10}H_{17}N_3O_6S$) conjugates, respectively (Fig. 5C). Similar costunolide conjugates were found in previous studies through the heterologous expression in tobacco of chicory and pyrethrum orthologs of these genes [34, 37].

3.5 Specificity of CcGAO

The CcGAO gene was co-expressed in yeast with either goldenrod (+)-germacrene D synthase (ScGDS), goldenrod cascarilladiene synthase (ScCs) or *Artemisia annua* amorphadiene synthase (AaAMDS) (Fig. 6D). After 48 h of incubation in an adjusted pH medium, the organic extracts of each culture were analyzed by GC-MS. Upon the co-expression of CcGAO and ScGDS, novel products were observed at retention time of 16.08 min, 16.33 min and 16.82 min with parental mass of *m/z* 220, 220 and 222, respectively. These oxidized products were not observed in the empty vector or the ScGDS extracts (Fig. 6A, Fig. S2). In the methylated extracts of the yeast expressing CcGAO and AaAMDS, several products were detected and identified as artemisinic alcohol, artemisinic acid and dihydroartemisinic aldehyde with retention time of 17.50 min, 17.62 min and 17.39 min and parental masses of 220, 234, 236 *m/z* respectively (Fig. 6B, Fig. S2). The co-expression of CcGAO with the ScCs led to a partial conversion of the cascarilladiene to an oxidized product with a retention time of 16.73 min and parental mass of 220 *m/z* (Fig. 6C, Fig. S2). In these three different reactions the oxidation performed by CcGAO was considered partial due to the fact that the non-oxidized sesquiterpenes were still visible in the yeast extracts expressing CcGAO and the corresponding terpene synthase.

4. Discussion

In this work the germacrene oxidase (CcGAO; CYP71AV9) and the costunolide synthase (CcCOS; CYP71BL5) from globe artichoke were identified and characterized. The co-

expression in yeast of CcGAO gene with the previously characterized GAS from *Cichorium intybus* converted germacrene A to different oxidized products, depending on the culture conditions. In acidic conditions the formation of germacra-1(10),4,11(13)-trien-12-ol and germacra-1(10),4,11(13)-trien-12-al, which correspond to the oxidized products of germacrene A was observed. In the buffered culture the germacrene A was efficiently converted to germacra-1(10),4,11(13)-trien-12-oic acid. It has been previously reported that co-expression in yeast and in acidic conditions of the GAS with GAO genes from other *Asteraceae* species like pyrethrum [34], sunflower, lettuce and chicory [32] leads to the production of costic acids and ilicic acid, which are cyclization products of germacrene A acid. However the synthesis of these cyclization products was not detected in our experiments, possibly due to a lower activity of CcGAO in comparison to the ones of the previously isolated in respected to the GAOs isolated.

In *Nicotiana benthamiana* no products were detected following headspace analyses of agro-infiltrated leaves co-expressing the cytosolic cCcGAS and CcGAO, although the disappearance of germacrene A suggested the efficient conversion of germacrene A by the CcGAO. Likely, the expected new compound produced by CcGAO reacted with endogenous metabolites to form conjugates undetectable with the metabolic profiling approaches used [37]. Indeed, analogous results were previously obtained in *N. benthamiana* with GAO from chicory [37].

In tobacco agro-infiltration experiments, the enzymatic activity of both *Cynara cardunculus* GAO and COS was confirmed: the co-expression of CcCOS with CcGAS and CcGAO led to the formation of two costunolide conjugates: costunolide-glutathione and costunolide-cysteine, although free costunolide could not be detected. These results are in accordance with those reported by [37], who detected formation of costunolide conjugates upon expression of *C. intybus* GAO and *T. parthenium* COS *in planta*.

The biosynthesis of terpenoids can occur in different plant subcellular compartments, including the cytosol, mitochondria and plastids [48,49]. The ectopic expression of several sesquiterpene synthases in mitochondria or plastids has been shown to improve plant sesquiterpene production [37,47]. When we evaluated the effect of the CcGAS targeting in agro-infiltrated *N. benthamiana* leaves it became clear that chloroplastic and mitochondrial localization did not improve germacrene A biosynthesis by CcGAS in this heterologous plant

(Fig. 5D). Our results are in contrast with the ones reported by Liu et al. [37] in tobacco, as they observed a 15 fold-higher production of germacrene A in mitochondria, compared to cytosolic expression. Since the plant host (*N. benthamiana*) and the targeting signals were identical in both experiments, the most logical explanation of this discrepancy should be found in the properties of the CcGAS. Possibly it is not very compatible with subcellular transport, or it may be more readily adapted to cytosolic conditions than to the conditions found in plastids or mitochondria.

Sesquiterpene synthases catalyze the first committed step in sesquiterpene biosynthesis through the conversion of the farnesyl diphosphate (FPP) to a large variety of sesquiterpene skeletons. The creation of C15 sesquiterpene hydrocarbon skeletons as well as the regio- and stereo-selective formation of the lactone ring is essential for the structural diversity of STLs observed in different plant species. The coordinated reactions of the sesquiterpene synthases and sesquiterpene-modifying P450s enzymes elaborate the diverse structures of STLs [33]. To assess the substrate specificity of the CcGAO enzyme, its cross-reactivity toward other sesquiterpene lactones was tested. The CcGAO enzyme was shown to convert different non-natural substrates: amorphadiene, cascarilladiene, and (+)-germacrene D to oxidized products. The reaction of the GAO from different species, including chicory, lettuce and sunflower, with amorphadiene has been demonstrated before [32]. Here we show the ability of CcGAO to oxidize also cascarilladiene and (+)-germacrene D (Fig. 6). These results suggest that CcGAO has broad substrate specificity and may react with a wide range of related substrates. As expected the oxidation activity of CcGAO of the non-natural substrates seems to be less than the one of the native substrate: germacrene A.

STLs are major components of the leaves of *Cynara* species, and knowledge of the mechanisms involved in the biosynthesis of these pharmaceutically important compounds can improve the therapeutic value of an important edible crop like globe artichoke.

In the present study two enzymes from the biosynthetic pathway towards STLs in *Cynara cardunculus* have been elucidated. The present results are of interest for developing microbial and plant-based platforms for the production of these pharmaceutically relevant secondary metabolites.

Acknowledgements

The authors wish to acknowledge Qing Liu for kindly providing pESC-Trp-TpGAS-CiGAO. Bert Schipper and Roland Mumm are acknowledged for technical support.

References

[1] S. Lanteri, A. Acquadro, E. Saba, E. Portis, Molecular fingerprinting and evaluation of genetic distances among selected clones of globe artichoke (*Cynara cardunculus* L. var. *scolymus* L.), J. Hortic. Sci. Biotech. 79 (2004) 863-870.

[2] S. Lanteri & E. Portis, Globe Artichoke and Cardoon. In Vegetables I, ed. Springer, 49-74. New York: Springer, 2008.

[3] R. Gebhardt, Antioxidative and protective properties of extracts from leaves of the artichoke (*Cynara scolymus* L) against hydroperoxide-induced oxidative stress in cultured rat hepatocytes, Toxicol. Appl. Pharm. 144 (1997) 279-286.

[4] D. Peshev, R. Vergauwen, A. Moglia, E. Hideg, W. Van den Ende, Towards understanding vacuolar antioxidant mechanisms: a role for fructans? J. Exp. Bot. 64 (2013) 1025-38.

[5] G. Aktay, D. Deliorman, E. Ergun, F. Ergun, E. Yeşilada, C. Çevik, Hepatoprotective effects of Turkish folk remedies on experimental liver injury, J. Ethnopharmacol. 73(2000) 121-129.

[6] R. Bundy, A.F. Walzer, R.W. Middelton, C. Wallis, H.C. Simpson, Artichoke (*Cynara scolymus*) reduces plasma cholesterol in otherwise healthy hypercholesterolemic adults: a randomized, double blind placebo controlled trial, Phytomedicine 15 (2008) 668–675.

[7] A. J. MacLeod, N. M. Pieris , N. G. DeTroconis, Aroma volatiles of *Cynara scolymus* and *Helianthus tuberosus*, Phytochemistry 7 (1982) 1647–1651.

[8] G. Schneider & K. Thiele, Distribution of the bitter principle cynaropicrin in *Cynara*, Planta Med, 26(1974) 174-83.

[9] J. Fritsche, C.M. Beindorff, M. Dachtler, H. Zhang, J.G., Lammers, Isolation, characterization and determination of minor artichoke (*Cynara scolymus* L.) leaf extract compounds, Eur.Food.Res.Technol. 215(2) (2002) 149-157.

[10] G. Cravotto, G. M. Nano, A. Binello, P. Spagliardi, G. Seu, Chemical and biological modification of cynaropicrin and grosheimin: a structure–bitterness relationship study, J.Sci.Food.Agr. 85(10) (2005) 1757-1764.

[11] B. Menin, C. Comino, E. Portis, A. Moglia, K. Cankar, H. Bouwmeester, S. Lanteri, J. Beekwilder, Genetic mapping characterization of the globe artichoke (+)-germacrene A synthase gene, encoding the first dedicated enzyme for biosynthesis of the bitter sesquiterpene lactone cynaropicrin, Plant.Sci. 190 (2012) 1-8.

[12] E. Rodriguez, G. H. N. Towers, J. C., Mitchell, Biological activities of sesquiterpene lactones, Phytochemistry 15(11) (1976) 1573-1580.

[13] A. K. Picman, Biological activities of sesquiterpene lactones, Biochem Syst Ecol, 14(3) (1986) 255-281.

[14] F. C. Seaman, Sesquiterpene lactones as taxonomic characters in the *Asteraceae,* Bot Rev 48(2) (1982) 121-594.

[15] K. Ishida, R. Kojima, M. Tsuboi, Y. Tsuda, M. Ito, Effects of artichoke leaf extract on acute gastric mucosal injury in rats, Biol. Pharm. Bull. 33(2) (2010) 223-229.

[16] P. A. Ramos, A. R. Guerra, O. Guerreiro, C. S. Freire, A. M. Silva, M. F. Duarte, A. J. Silvestre, Lipophilic Extracts of *Cynara cardunculus* L. var. *altilis* (DC): a Source of Valuable Bioactive Terpenic Compounds, J. Agr. Food. Chem. 61(35) (2013) 8420-8429.

[17] R. Paduch, M. Kandefer-Szerszeń, M. Trytek, J. Fiedurek, Terpenes: substances useful in human healthcare, Arch. Immunol. Ther. Ex. 55(5) (2007) 315-327.

[18] H. Shimoda, K. Ninomiya, N. Nishida, T. Yoshino, T. Morikawa, H. Matsuda, M. Yoshikawa, Anti-Hyperlipidemic sesquiterpenes and new sesquiterpene glycosides from the leaves of artichoke (*Cynara scolymus* L.): structure requirement and mode of action, Bioorg. Med. Chem .Lett. 13(2) (2003) 223-228.

[19] J. Y. Cho, A. R. Kim, J. H. Jung, T. Chun, M. H. Rhee, E. S. Yoo, Cytotoxic and pro-apoptotic activities of cynaropicrin, a sesquiterpene lactone, on the viability of leukocyte cancer cell lines, Eur. J. Pharmacol. 492(2) (2004) 85-94.

[20] E. C. Schinor, M. J. Salvador, I. Y. Ito, S. De Albuquerque, D. A. Dias, Trypanocidal and antimicrobial activities of *Moquinia kingie,* Phytomedicine 11(2) (2004) 224-229.

[21] F. Emendörfer, F. Emendörfer, F. Bellato, V.F. Noldin, V. Cechinel-Filho, R.A. Yunes, F. Delle Monache , A. M. Cardozo, Antispasmodic activity of fractions and cynaropicrin from *Cynara scolymus* on *guinea-pig ileum*, Biol. Pharm. Bull 28(5) (2005) 902-904.

[22] K. Yasukawa, H. Matsubara, Y. Sano, Inhibitory effect of the flowers of artichoke (*Cynara cardunculus*) on TPA-induced inflammation and tumor promotion in two-stage carcinogenesis in mouse skin, J. Nat. Med-Tokyo 64(3) (2010) 388-391.

[23] Y. T. Tanaka, K. Tanaka, H. Kojima, T. Hamada, T. Masutani, M. Tsuboi, Y. Akao, Cynaropicrin from *Cynara scolymus* L. suppresses photoaging of skin by inhibiting the transcription activity of nuclear factor-kappa B, Bioorg. Med. Chem. Lett. 23(2) (2012) 518-523.

[24] J. Y. Cho, K. U. Baik, J. H. Jung, M. H. Park, In vitro anti-inflammatory effects of cynaropicrin, a sesquiterpene lactone, from *Saussurea lappa,* Eur. J. Pharmacol 398 (2000) 399–407.

[25] N. H. Fischer, Biochemistry of the mevalonic acid pathway to terpenoids; Towers GHN, Stafford HA, eds. New York: Plenum Press. 1990

[26] T. A. Geissman, The biogenesis of sesquiterpene lactones of the *compositae*. in Recent Advances in Phytochemistry, eds Runeckles VC, Marby TJ (Academic Press, New York), 6:65–95, 1973

[27] W. Herz, S. V. Bhat, H. Crawford, H. Wagner, G. Maurer, L. Farkas, Bahifolin, a new sesquiterpene lactone, and 5, 7-dihydroxy-3, 3′, 4′, 6-tetramethoxyflavone, a new flavone, from *Bahia oppositifolia*, Phytochemistry 11(1) (1972) 371-375.

[28] G.A. Cordell, Biosynthesis of sesquiterpenes, Chem. Rev. 76(4) (1976) 425-460.

[29] W. Parker, J. C. Roberts, R. Ramage, Sesquiterpene biogenesis, Chem. Soc. Rev. 21(3) (1967) 331-363.

[30] J. W. De Kraker , M. C. Franssen, A. De Groot, T. Shibata, H. J. Bouwmeester, Germacrenes from fresh costus roots, Phytochemistry 58(3) (2001) 481-7.

[31] J.W. De Kraker, M. C. Franssen, M. Joerink, A. De Groot, H.J. Bouwmeester, Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory, Plant. Physiol. 129(2002) 257-268.

[32] D. T. Nguyen, J. C. Gopfert, N. Ikezawa, G. MacNevin, M. Kathiresan, J. Conrad, O. Spring, and D.K. Ro, Biochemical conservation and evolution of germacrene A oxidase in *Asteraceae,* J. Biol. Chem. 285(2010) 16588–16598.

[33] N. Ikezawa, J. C. Göpfert, D. T. Nguyen, S. U. Kim, P. E. O'Maille, O. Spring, D. K. Ro, Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism, J. Biol. Chem. 286 (24) (2011) 21601-11.

[34] A. M. Ramirez, N. Saillard, T. Yang, M. C. R. Franssen, H. J. Bouwmeester, M. A. Jongsma, Biosynthesis of Sesquiterpene Lactones in Pyrethrum (*Tanacetum cinerariifolium*), PLoS one 8(5)(2013) e65030.

[35] H. J. Bouwmeester, J. Kodde, F. W. A. Verstappen, I. G. Altug, J. W. De Kraker, T. E. Wallaart, Isolation and characterization of two germacrene A synthase cDNA clones from chicory, Plant Physiol. 129 (2002) 134–144.

[36] K. Cankar, A. V. Houwelingen, D. Bosch, T. Sonke, H. Bouwmeester, A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)- valencene, Febs. Lett. 585(2011) 178–182.

[37] Q. Liu, M. Majdi, K. Cankar, M. Goedbloed, T. Charnikhova, F. W. Verstappen, R. C. H. De Vos J. Beekwilder, S. Van der Krol, H. J. Bouwmeester, Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*, PloS one 6(8), (2011), e23255.

[38] D. Scaglione, A. Acquadro, E. Portis, M. Tirone, S.J. Knapp, S. Lanteri, RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L., BMCGenomics 13(2012) 3.

[39] J. Thompson, T. Gibson, F. Plewniak, F.Jeanmougin, D. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25(1997) 4876–4882.

[40] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, Mol. Biol. Evol 28(2011) 2731-2739.

[41] D. Pompom, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments, Methods.Enzymol 272(1996) 51-64.

[42] P. Urban, C. Mignotte, M. Kazmaier, F. Delorme, and D. Pompon, Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5, J. Biol. Chem. 272 (1997) 19176–19186.

[43] P. Mercke, M. Bengtsson, H. J. Bouwmeester, M. A. Posthumus, P. E. Brodelius, Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua* L., Arch. Biochem. Biophys. 381(2)(2000)173-80.

[44] I. Prosser, I. G. Altug, A. L. Phillips, W. A. König, H. J. Bouwmeester, M. H. Beale, Enantiospecific (+)-and (−)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif, Arch. Biochem. Biophys. 432(2) (2004) 136-144.

[45] I. Altug, Isolation of cDNAs encoding (+)- and (–)-Germacrene D-Synthase, α-Gurjunene-and Cascarilladiene Synthase from *Solidago canadensis* L. and their functional expression in *E. coli* (Doctoral dissertation, Universitätsbibliothek) 2004.

[46] J. W. De Kraker, M. C. Franssen, A. De Groot, W. A. König, H. J. Bouwmeester, (+)-Germacrene A Biosynthesis, the Committed Step in the Biosynthesis of Bitter Sesquiterpene Lactones in Chicory, Plant.Physiol 117(4) (1998) 1381-1392.

[47] T. W. Van Herpen, K. Cankar, M. Nogueira, D. Bosch, H. J. Bouwmeester, J. Beekwilder, *Nicotiana benthamiana* as a Production Platform for Artemisinin Precursors, PLoS. one 5(12): e14222. (2010).

[48] M. Gutensohn, D. A. Nagegowda, N. Dudareva, Involvement of compartmentalization in monoterpene and sesquiterpene biosynthesis in plants. In: Bach TJ, Rohmer M, eds. Isoprenoid synthesis in plants and microorganisms, New York, NY, USA: Springer, 155–169 2013.

[49] B. M. Lange, A. Ahkami, Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes current status and future opportunities, Plant. Biotechnol.J. 11(2) (2013)169-196.

FIGURE CAPTIONS

Fig. 1. Proposed biosynthetic pathway of STLs. Enzymatic reactions: I cyclization by the germacrene A synthase; II oxidation by the germacrene A oxidase; III hydroxylation by the costunolide synthase; IV subsequent spontaneous lactonization.

Fig. 2. Phylogenetic tree of the different P450s ortholog proteins GAO (A) and COS (B). This tree was obtained with MEGA software version 5.2 and based on the neighbour-joining method (5000 bootstrap replicates). The length of the lines indicates the relative distances between nodes. Cc: *Cynara cardunculus var. scolymus; Cca:Cynara cardunculus var. altilis; Ci: Cichorium intybus; Ls: Lactuca sativa; Ha: Helianthus annuus; Sc: Saussurea costus; Bs: Barnadesia spinosa; Tc: Tanacetum cinerariifolium.*

Fig. 3. Enzymatic characterization of *Cynara cardunculus* GAO (CcGAO) and COS (CcCOS) in yeast.
(A) GC-MS chromatograms of yeast extract expressing chicory GAS (CiGAS) and CcGAO, only CiGAS or empty plasmid in acidic conditions. Peak 1: elema-1,3,11(13)-trien-12-al; peak 2: elema-1,3,11(13)-trien-12-ol.
(B) GC-MS chromatograms of yeast extract expressing chicory GAS (CiGAS) and globe artichoke GAO (CcGAO), only CiGAS or empty plasmid in buffered medium. Peak 3: elema-1,3,11(13)-trien-12-oic acid.
(C) LC-MS metabolic profile of WAT11 yeast strain extracts expressing *Tanacetum parthenium* GAS (TpGAS), *Cichorium intybus* GAO (CiGAO) with CcCOS, or the empty vector extracts as a negative control. Metabolic profile of (+)-costunolide standard is also reported.
(D) Mass spectrum of peaks obtained from yeast extracts expressing CiGAS and CcGAO: (a) elema-1,3,11(13)-trien-12-oic acid; and yeast expressing: TpGAS, CiGAO and CcCOS (b) (+)-costunolide.

Fig. 4. Globe artichoke GAS (CcGAS) expression in *Nicotiana benthamiana.*
(A) Constructs used for the globe artichoke GAS (CcGAS) infiltration experiments. LB: left border; Pnos/Tnos: promoter and terminator of nopaline synthase gene; NptII: neomycin phosphotransferase gene; PRbcS: Rubisco promoter; TRbcS: Rubisco terminator; coxIV: yeast mitochondrial targeting signal; Chloro: chloroplast targeting signal; RB: right border.
(B) GC-MS chromatograms of the volatiles emitted from *N. benthamiana* leaves infiltrated with the indicated genes: a: Empty vector; b: CcGAS cytosol; c: CcGAS chloroplast; d: CcGAS mitochondria.
(C) Average Peak Area (TIC) of β-elemene (on-column heat-induced Cope rearrangement product of germacrene A) observed in infiltrated leaves of *N. benthamiana* with CcGAS using cytosol targeting, chloroplast targeting or mitochondrial targeting.

Fig. 5. Enzymatic characterization of globe artichoke GAO (CcGAO) and globe artichoke COS (CcCOS) *in planta*.
(A) Headspace chromatograms of volatile metabolites in *N. benthamiana leaves* agro-infiltrated with empty vector, CcGAS or CcGAS/CcGAO.
(B) LC-MS chromatograms of non-volatile metabolites in *N. benthamiana leaves* agro-infiltrated with empty vector, CcGAS/CcGAO or CcGAS/CcGAO/CcCOS. Peak 1: costunolide-Cys; peak 2: costunolide-GSH.
(C) Chemical structures of costunolide conjugates: 1. costunolide-cysteine, 2. costunolide-glutathione.

Fig. 6. Activity of globe artichoke GAO (CcGAO) on other substrates in yeast.
(A) GC-MS chromatograms of yeast extracts expressing empty plasmid, goldenrod (+)-germacrene D synthase (ScGDS) alone or ScGDS/CcGAO.
(B) Selective ions GC-MS chromatograms at *m/z* 220 of the sesquiterpenoid products obtained with the expression of empty vector, *Artemisia annua* amorphadiene synthase (AaAMDS) alone and co-expression in yeast of AaAMDS and CcGAO. Peak 1: dihydroartemisinic aldehyde; peak 2: artemisinic alcohol; peak 3: artemisinic acid.
(C) GC-MS chromatographs at a selective ion *m/z* 220 of yeast extracts expressing empty vector, goldenrod cascarilladiene synthase (ScCs) alone and ScCs/CcGAO.
(D) Structures of (+)-germacrene D, amorpha-4,11-diene, cascarilladiene, germacrene A acid.

Supplementary data

Table S1. Comparison of germacrene A oxidase (GAO) proteins from different plant species: percentage of identity, full length size and molecular weight (www.ebi.ac.uk/Tools/clustalw2/index.htmL) (Thompson et al., 1997) were indicated. *Cynara cardunculus* var. *scolymus (Cc); Cynara cardunculus* var. *altilis (Cca); Cichorium intybus (Ci); Lactuca sativa (Ls); Helianthus annuus (Ha); Saussurea costus (Sc); Barnadesia spinosa (Bs); Tanacetum cinerariifolium (Tc).*

Table S2. Comparison of costunolide synthase (COS) proteins from different plant species: percentage of identity, full length size and molecular weight (www.ebi.ac.uk/Tools/clustalw2/index.htmL) (Thompson et al., 1997) were indicated. *Cynara cardunculus* var. *scolymus (Cc); Cynara cardunculus* var. *altilis (Cca); Cichorium intybus (Ci); Lactuca sativa (Ls); Tanacetum cinerariifolium (Tc).*

Fig. S1. Alignment of protein sequences of GAO (A) and COS (B) from different plant species. Alignments were performed using ClustalW (Bio Edit software). Dark and light shading indicate identical and similar residues, respectively. The red box presents the heme-binding region, a conserved region characteristic of cytochrome P450 proteins. Cc: *Cynara cardunculus var scolymus, Cca: Cynara cardunculus var altilis,Ci: Cichorium intybus; Ls: Lactuca sativa, Ha: Helianthus annuus, Sc: Saussurea costus, Bs: Barnadesia spinosa and Tc: Tanacetum cinerariifolium.*

Fig. S2. Mass spectra of the oxidized products obtained by CcGAO.
Oxidized products obtained from the co-expression of AaAMDS and CcGAO with (a) artemisinic alcohol; (b) artemisinic acid, (c) dihydroartemisinic aldehyde
Oxidized compound observed with co-expression of ScGDS and CcGAO: (d) compound observed at 16.08 min, (e) oxidized compound observed at 16.33 and (f) oxidized compound observed at 16.82min
Oxidized compound obtained with the co-expression of ScCs and CcGAO at 16.73 min (g).

Table S1

	CcGAO	CcaGAO	ScGAO	LsGAO	CiGAO	HaGAO	TcGAO	full length	Molecular Weight
CcGAO	100.00	99.59	95.06	89.09	89.09	87.24	89.51	486 aa	54.71 Da
CcaGAO	99.59	100.00	94.65	89.09	89.09	86.83	89.09	486 aa	54.72 Da
ScGAO	95.06	94.65	100.00	91.39	90.98	88.52	90.16	488 aa	55.04 Da
LsGAO	89.09	89.09	91.39	100.00	97.34	89.14	89.75	488 aa	54.97 Da
CiGAO	89.09	89.09	90.98	97.34	100.00	89.14	88.93	488 aa	54.98 Da
HaGAO	87.24	86.83	88.52	89.14	89.14	100.00	88.93	488 aa	55.09 Da
TcGAO	89.51	89.09	90.16	89.75	88.93	88.93	100.00	488 aa	55.07 Da
BsGAO	83.33	83.33	82.99	84.02	83.61	80.94	83.81	496 aa	55.71 Da

Table S2

	CcCOS	CcaCOS	LsCOS	CiCOS	TcCOS	full length	Molecular Weight
CcCOS	100.00	99.80	88.57	88.03	87.45	494 aa	55.88 Da
CcaCOS	99.80	100.00	88.78	88.24	87.65	494 aa	55.92 Da
LsCOS	88.57	88.78	100.00	96.73	89.39	490 aa	55.27 Da
CiCOS	88.03	88.24	96.73	100.00	89.07	494 aa	55.69 Da
TcCOS	87.45	87.65	89.39	89.07	100.00	496 aa	56.26 Da

Figure 1

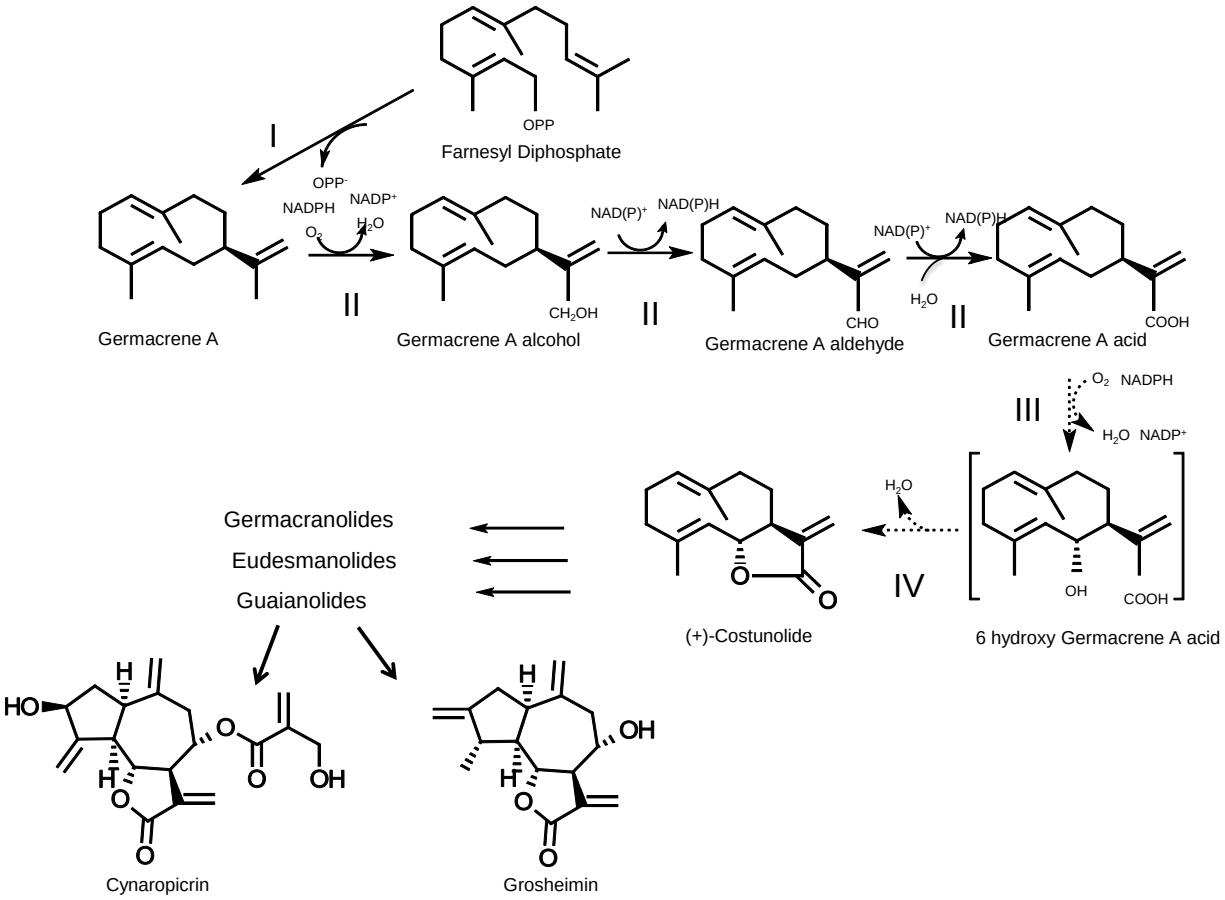

OPP
Farnesyl Diphosphate
I
OPP-
NADPH
NADP+
O2
H2O
II
Germacrene A
CH2OH
Germacrene A alcohol
NAD(P)+
NAD(P)H
II
CHO
Germacrene A aldehyde
NAD(P)+
NAD(P)H
H2O
II
COOH
Germacrene A acid
III
O2 NADPH
H2O NADP+
H2O
IV
OH
COOH
6 hydroxy Germacrene A acid
O
O
(+)-Costunolide
Germacranolides
Eudesmanolides
Guaianolides
HO
H
O
O
O
OH
Cynaropicrin
H
OH
O
O
Grosheimin

Figure 2

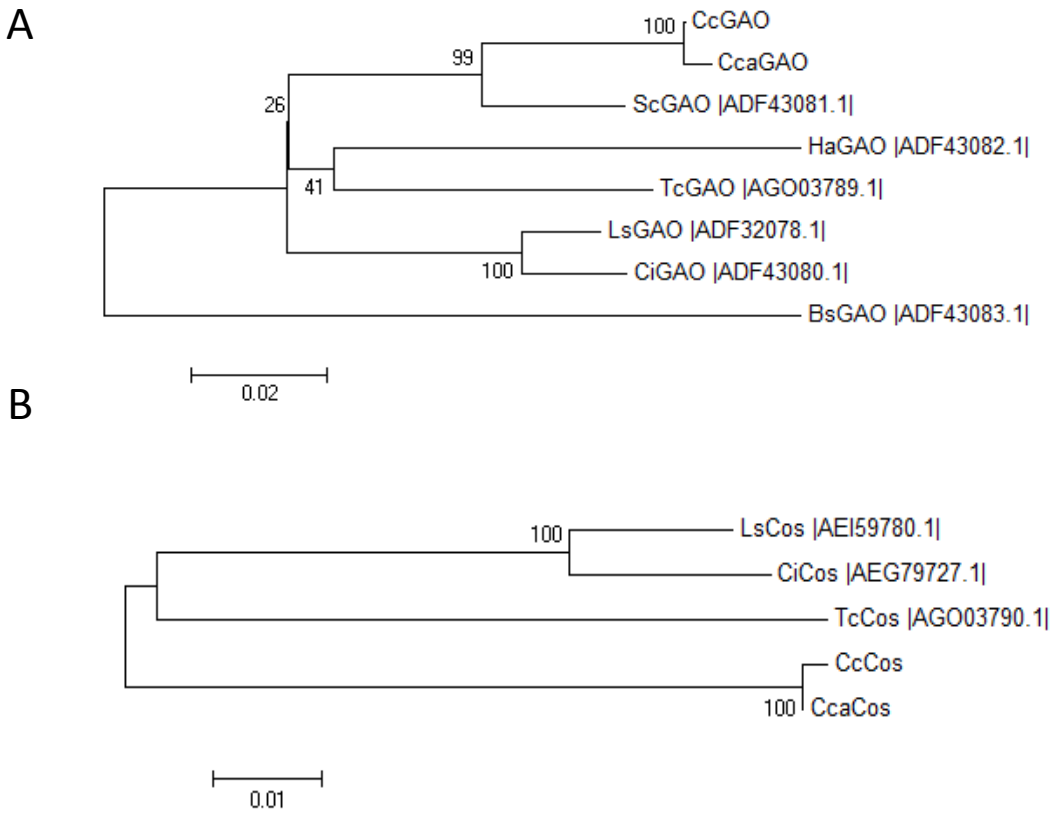

A
100
CcGAO
99
CcaGAO
26
ScGAO |ADF43081.1|
HaGAO |ADF43082.1|
41
TcGAO |AGO03789.1|
LsGAO |ADF32078.1|
100
CiGAO |ADF43080.1|
BsGAO |ADF43083.1|
0.02
B
100
LsCos |AEI59780.1|
CiCos |AEG79727.1|
TcCos |AGO03790.1|
CcCos
100
CcaCos
0.01

Figure 3

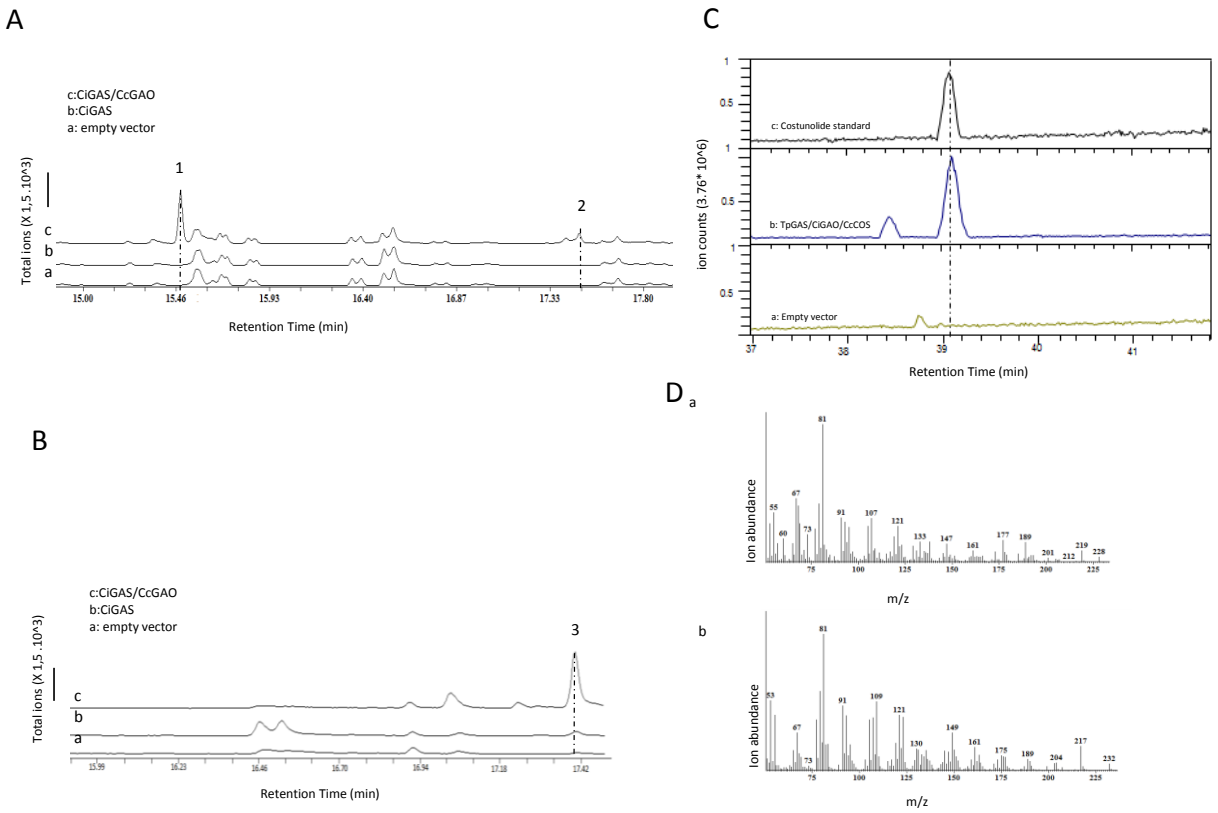

A
c:CiGAS/CcGAO
b:CiGAS
a: empty vector
Total ions (X 1,5 .10^3)
1
2
c
b
a
15.00
15.46
15.93
16.40
16.87
17.33
17.80
Retention Time (min)
B
c:CiGAS/CcGAO
b:CiGAS
a: empty vector
Total ions (X 1,5 .10^3)
3
c
b
a
Retention Time (min)
C
ion counts (3.76* 10^6)
c: Costunolide standard
b: TpGAS/CiGAO/CcCOS
a: Empty vector
37
38
39
40
41
Retention Time (min)
D a
Ion abundance
m/z
b
Ion abundance
m/z

Figure 4

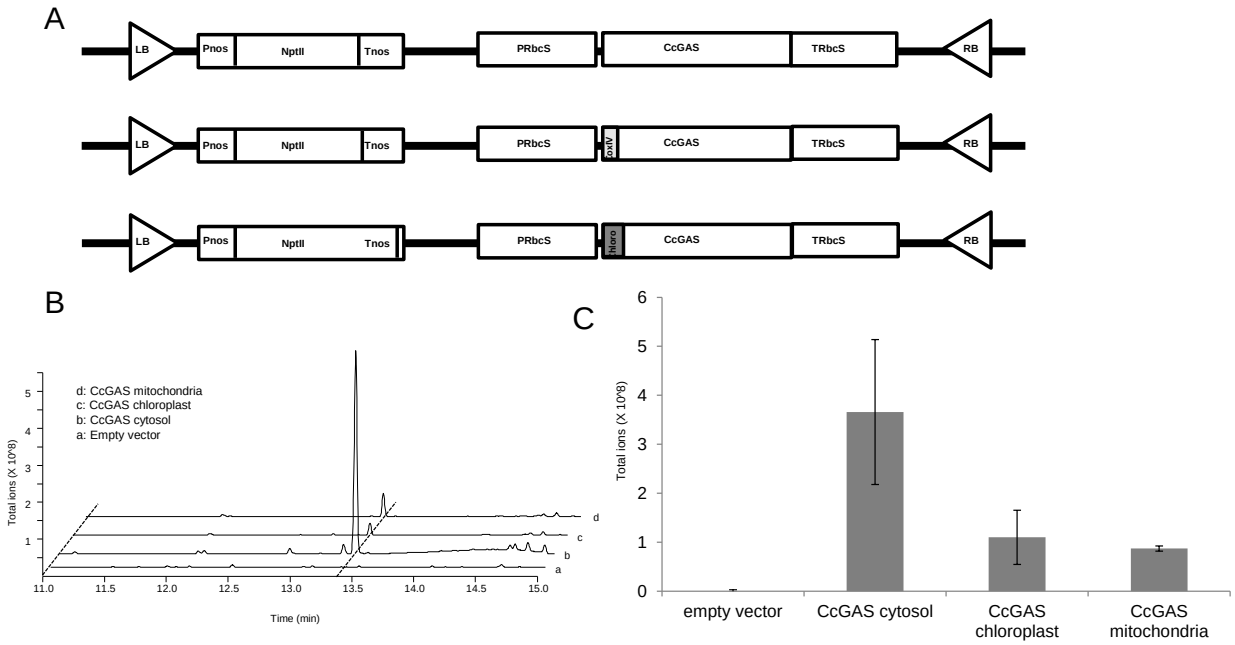

A
LB
Pnos
NptII
Tnos
PRbcS
CcGAS
TRbcS
RB
pMXC
rbcS chloro
B
d: CcGAS mitochondria
c: CcGAS chloroplast
b: CcGAS cytosol
a: Empty vector
Total ions (X 10^8)
Time (min)
C
Total ions (X 10^8)
empty vector
CcGAS cytosol
CcGAS chloroplast
CcGAS mitochondria

Figure 5

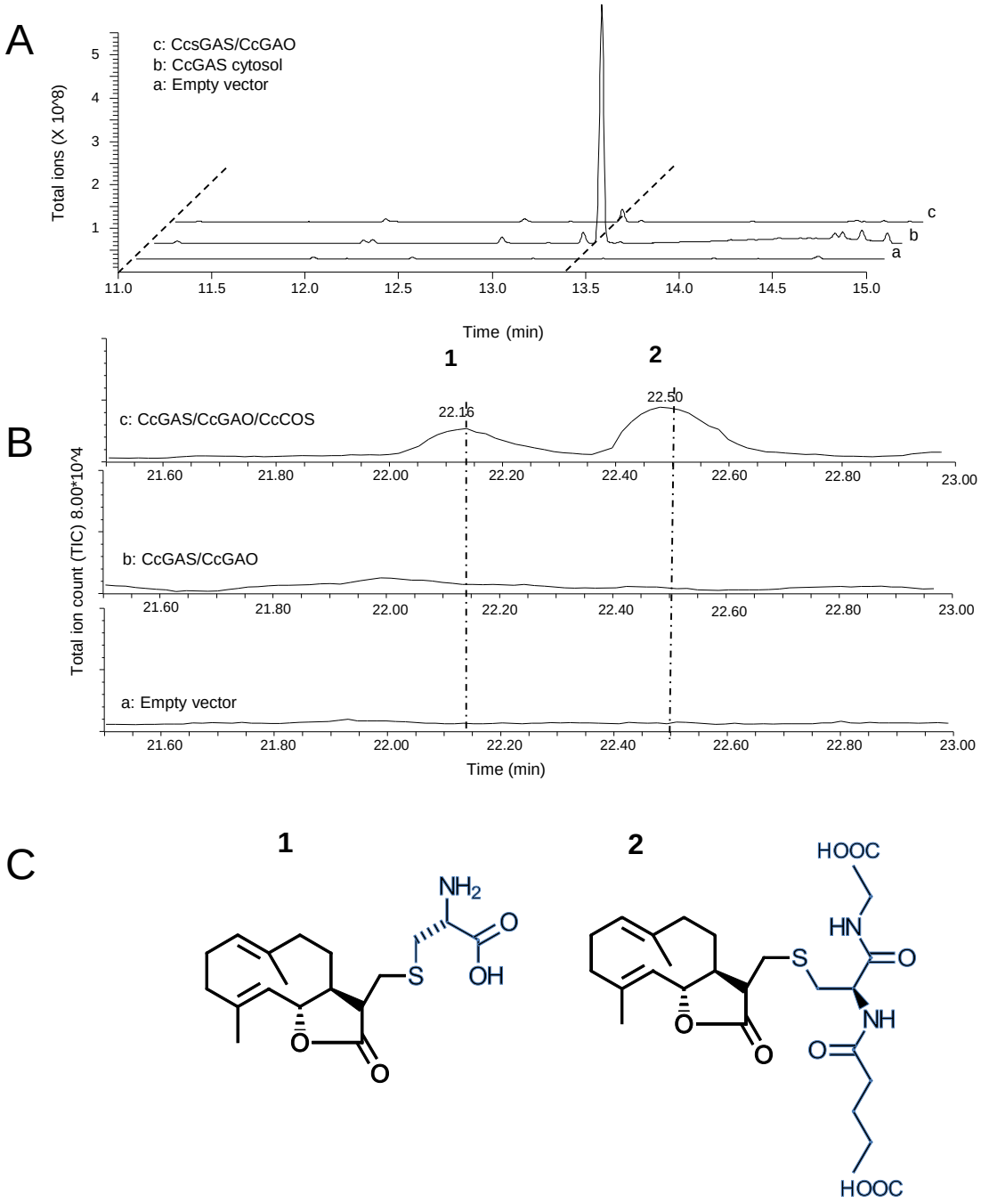

A
c: CcsGAS/CcGAO
b: CcGAS cytosol
a: Empty vector
Total ions (X 10^8)
Time (min)
B
1
2
c: CcGAS/CcGAO/CcCOS
22.16
22.50
b: CcGAS/CcGAO
a: Empty vector
Total ion count (TIC) 8.00*10^4
Time (min)
C
1
2
NH2
OH
HOOC
HN
NH
S
O

Figure 6

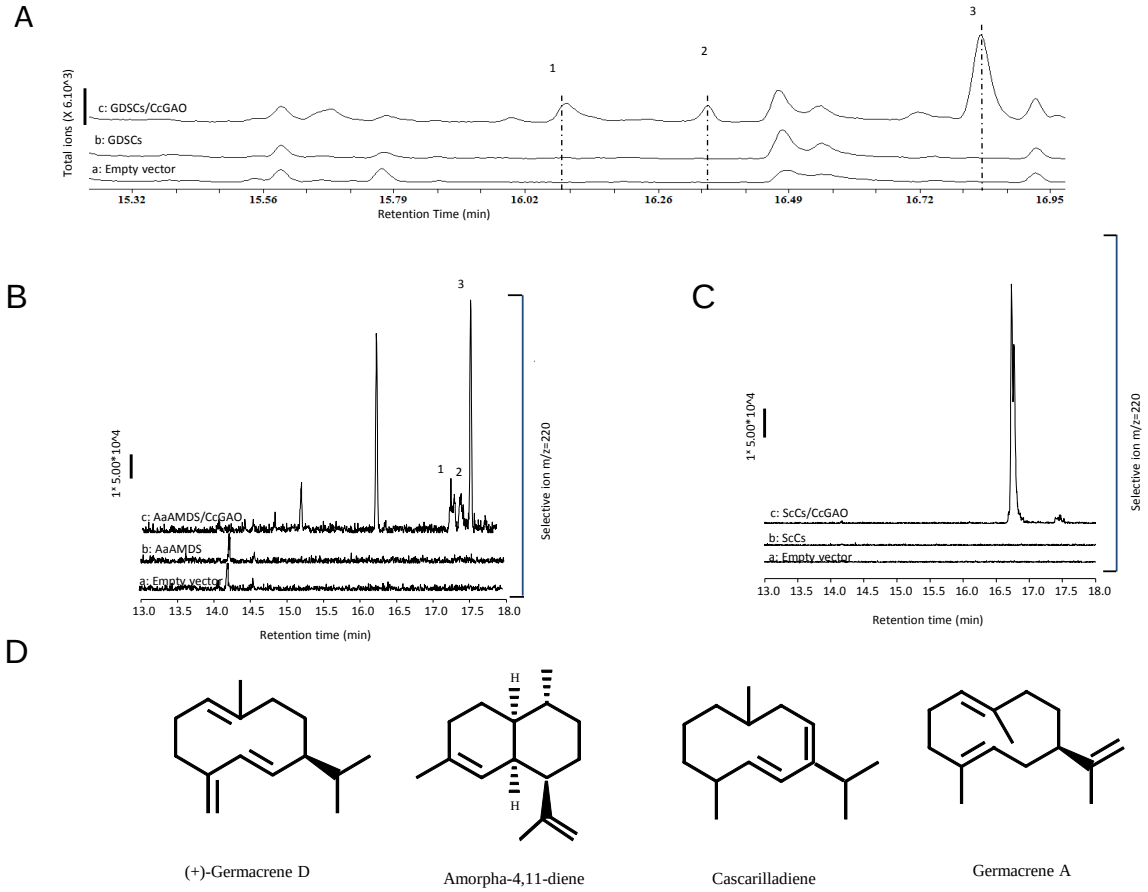
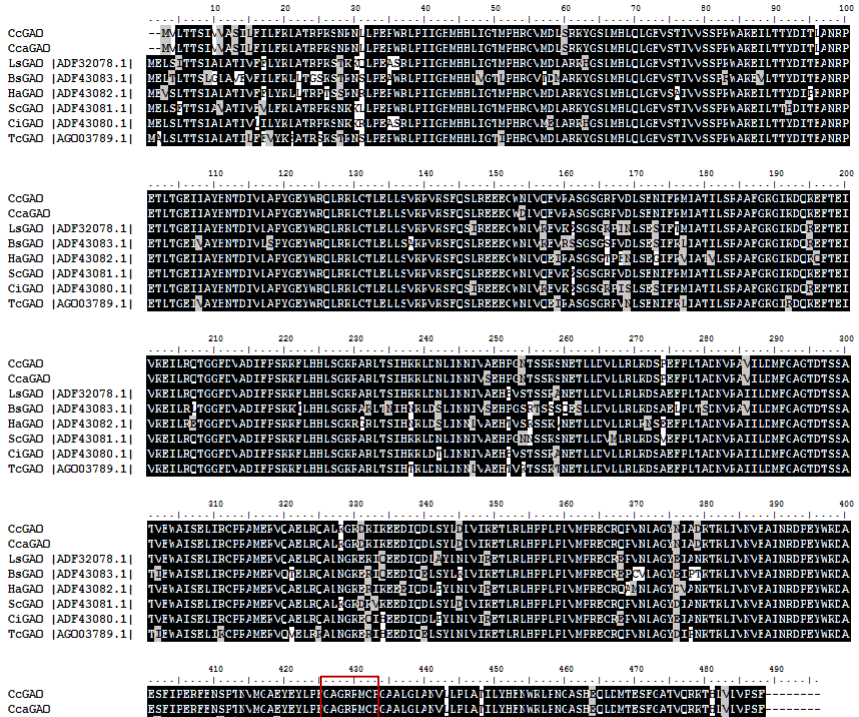

A
Total ions (X 6.10^3)
c: GDSCs/CcGAO
b: GDSCs
a: Empty vector
1
2
3
15.32
15.56
15.79
16.02
16.26
16.49
16.72
16.95
Retention Time (min)
B
1* 5.00*10^4
c: AaAMDS/CcGAO
b: AaAMDS
a: Empty vector
Selective ion m/z=220
13.0 13.5 14.0 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0
Retention time (min)
C
1* 5.00*10^4
c: ScCs/CcGAO
b: ScCs
a: Empty vector
Selective ion m/z=220
13.0 13.5 14.0 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0
Retention time (min)
D
H
H
(+)-Germacrene D
Amorpha-4,11-diene
Cascarilladiene
Germacrene A

Figure S1

A


CcGAO
CcaGAO
LsGAO |ADF32078.1|
BsGAO |ADF43083.1|
HaGAO |ADF43082.1|
ScGAO |ADF43081.1|
CiGAO |ADF43080.1|
TcGAO |AGO03789.1|

B

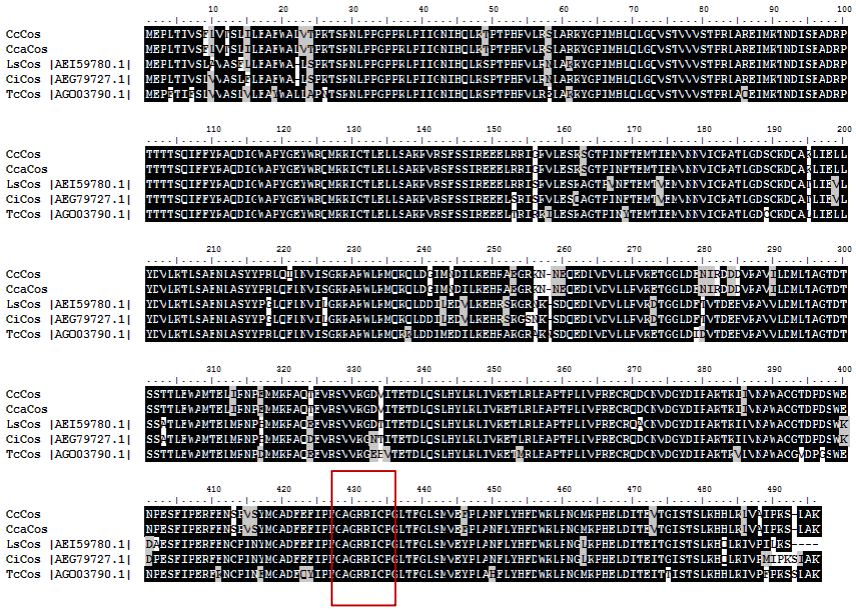

CcCos
CcaCos
LsCos |AEI59780.1|
CiCos |AEG79727.1|
TcCos |AGO03790.1|

Figure S2