Plant Physiology and Biochemistry 144 (2019) 73–84

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

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

Research article

Unravelling triterpene biosynthesis through functional characterization of an oxidosqualene cyclase (OSC) from *Cleome arabica* L

Ladhari Afef*, Chappell Joseph

Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY, 40536, USA

ARTICLE INFO

Keywords:
Cleome arabica
Heterologous expression
Lupeol
Pentacyclic triterpene
Oxidosqualene cyclase

ABSTRACT

Cleome arabica is a medicinal plant contains diverse bioactive compounds and terpenoids are the major components. However, the isolation and purification of the active triterpenes from this plant involve long and complicated procedures. The present work investigates the triterpenes profiles of different tissues, besides that, describes the isolation, heterologous expression and functional characterization of *C. arabica* gene coding for triterpenes synthases. The phytochemical investigation through GC-MS revealed significant accumulation of pentacyclic triterpenes in leaves and siliques at mature stage compared to the stems and roots of *C. arabica*. Among the pentacyclic triterpenes, the lupeol reached the highest level of 320 μg/g DW in leaves at maturity stage compared to the other tissues. The biosynthesis of a pentacyclic triterpene was investigated through isolation and cloning of a full-length oxidosqualene cyclase cDNA (CaOSC) from mature leaves of *C. arabica*. The bioinformatic analyses revealed that CaOSC was highly homologous with the characterized lupeol synthases and shared 79.3% identity to camelliol C synthase from *A. thaliana*. Heterologous expression of CaOSC gene in *Saccharomyces cerevisiae* synthesized lupeol as a single product. The lupeol biosynthesis was exponentially increased after induction through the fermentation process reaching the maximum of 2.33 μg/ml for 240 h. Furthermore, organ-specific expression of lupeol gene was exactly matched the accumulation pattern in different tissues of *C. arabica* during phenological cycle. Thus, the identified CaOSC will be useful in enhancing triterpene yield for industrial purposes.

1. Introduction

Cleome arabica (Capparidaceae) is an appreciated plant in North Africa for its medicinal and biological attributes (Tsichritzis et al., 1993; Djeridane et al., 2010; Takhi et al., 2011), that is used to heal rheumatic pains. It has been reported that this plant produces a wide range of natural products mainly terpenoids that are involved in fundamental physiological and ecological processes (Ladhari et al., 2013, 2014, 2018). Some terpenoids from *C. arabica* have been found to possess an anti-inflammation (Ahmed et al., 1990; Tsichritzis et al., 1993) antimicrobial, antifungal (Takhi et al., 2011), antioxidant (Selloum et al., 1997; Djeridane et al., 2010), phytotoxic (Ladhari et al., 2013) and cytotoxic activities (Nagaya et al., 1997), which led to an increased interest in this plant. Another important aspect of this plant is that the siliques and seeds contain new active triterpenes, which are the main compounds that could influence *C. arabica* distribution in nature (Ladhari et al., 2018). However, the isolation and purification of the active triterpenes from this medicinal plant involve long and complicated procedures. Due to the low amounts of these compounds in naturally growing plant, a comprehensive understanding of the biosynthetic pathway is extremely important to overcome low production of triterpenes for further use in industrial production.

Triterpenoids are ubiquitous isoprenoids produced by all eukaryotes, with different uses in agriculture, in medicine, and in industry. However, terpenes production in plant changes with growth and environmental conditions, making it difficult to obtain large amounts of any one of terpene constituent (Wu et al., 2006). Through the recent advances in biotechnology, the metabolic engineering for isoprenoid biosynthesis has been the subject of extensive studies over the last decade in order to increase the production and yields of desired products (Chappell et al., 1995; Kempinski et al., 2019). In fact, an annual terpenoids production of 10^7 tons (Van Beilen and Poirier, 2007), including menthol, monoterpenoid extracted from peppermint are used in the fragrance industry (Bohlmann and Keeling, 2008). Similarly, artemisinin, a sesquiterpene isolated from the plant *Artemisia annua*, and it has been developed as a key of pharmacological agent to control malaria (Tu, 2011), with annual global demand exceeding 100 metric tons in 2006 (Kindermans et al., 2007). Despite these important productions,

* Corresponding author.
E-mail addresses: afef.ladh@yahoo.fr, afef.ladhari@uky.edu (A. Ladhari).

https://doi.org/10.1016/j.plaphy.2019.09.035
Received 26 July 2019; Received in revised form 9 September 2019; Accepted 20 September 2019
Available online 21 September 2019

the demand and supply imbalance in terpenoids could hinder their extensive application. In fact, specific terpenes are highly required, but the availability of single terpene entities is limited in nature (Zhuang and Chappell, 2015). In plants, isoprenoid biosynthesis can proceed via two independent pathways, the MVA pathway which occurs in the cytosol in association with the endoplasmic reticulum, or the methylerythritol phosphate (MEP) pathway, which occurs within plastids (Kempinski et al., 2019). Most of the triterpenoids are derived from the cyclization of 2,3-oxidosqualene, catalyzed by oxidosqualene cyclases (OSCs), which can generate and modify the chemical structures of terpenoids (Xu et al., 2004). Approximately 50 OSCs have been cloned from diverse plant species and have been characterized, typically using heterologous expression in yeast. Many types of OSCs (Basyuni et al., 2010) among higher plants were found to catalysis the formation of the most popular pentacyclic triterpene, such as β-amyrin synthase (Liu et al., 2016; Sun et al., 2017), lupeol synthase (Guhling et al., 2006) and cycloartenol synthase (Calegario et al., 2016), but some were also reported to be multifunctional triterpene synthase (Kushiro et al., 1998; Basyuni et al., 2006; Brendolise et al., 2011).

Thus, despite these amounts of studies on many plants, the literature contains little information about the chemistry of terpenes in *Cleome arabica*. Besides, its molecular biology is basically unknown and there is no sequence or structural information on oxidosqualene cyclase in *C. arabica*. Therefore, it's crucial to understand the biosynthetic pathway of its active compounds for further industrial uses. In this context, the main purpose of this research is to clone and functionally characterize an oxidosqualene cyclase involved in the biosynthesis of an important active pentacyclic triterpene in *Cleome arabica.*

2. Material and methods

2.1. Plant material and RNA extraction

Seeds of *C. arabica* were obtained from a natural population growing in south area of Tunisia. Inflorescences were removed from plants and seeds were collected at dehiscence of the siliques. Seeds were disinfected using 70% ethanol for 2 min, and then soaked in sodium hypochlorite (3%) for 10 min at room temperature, rinsed by several times with sterile distilled water. The seeds were germinated in Petrie plate for 5 days and then were planted in pots in a greenhouse (Photoperiod 16 h: 8 h/light: dark at 30 °C) setting at the University of Kentucky, USA. Fresh roots, stems, leaves, siliques of *C. arabica* from 15 to 130 days were harvested in 5 phenological stages: vegetative, floral budding, full flowering, fresh fruiting, and mature fruiting. Plants with newly emerged shoots with leaves were classified as belonging to the vegetative stage, and samples were collected after 2 weeks. Only shoots with floral buds were selected to represent the floral budding stage after 6 weeks. At the full flowering stage, only shoots with full opened flowers were considered after 10 weeks. For the fresh and mature fruiting stages, the samples were collected when the shoots had siliques with green and black seeds, respectively, after 14 and 18 weeks. The harvested plant materials at different development stages were frozen immediately with liquid nitrogen and stored at −80 °C for further use.

Total RNA was extracted from the mature leaves of *C. arabica* using TRIzol reagent according to the manufacturer's protocol (Invitrogen, USA) and treated with DNase I (at 37 °C for 30 min to remove any traces of genomic DNA. The final RNAs concentration was determined by measuring the absorbance at 260 nm by Nano Drop ND-2000c spectrophotometer. The quality of RNA was checked by determining the ratio of absorbance at 260 and 280 nm ($A_{260/280}$) and its integrity was examined by electrophoresis in 1% (w/v) agarose gel.

2.2. cDNA synthesis and cloning of oxidosqualene cyclase

First strand cDNA was synthesized from 1 μg DNase I-treated total RNA using superscript III Reverse Transcriptase (Invitrogen, USA) with Adapter- oligo (dt) primer. The reaction was set in total volume of 20 μL of total RNA, 1 μM oligo(dT) primer, first strand buffer (250 mM Tris-HCl, pH 8.3, 250 mM KCl, 25 mM $MgCl_2$, 0.1 M DTT), 10 mM dNTPs, and RNase out (40 mM). The amplification was carried out at 42 °C for 90 min and final denaturation step at 70 °C for 10 min to inactivate the reverse transcriptase.

Primers were designed based on highly conserved regions of amino acid sequences of reported OSCs by multiple sequence alignment of retrieved sequences from the GenBank™ database at National Center for Biotechnology Information (NCBI). The PCR was done using cDNA as template for partial CaOSC gene amplification with the following conditions: 1 cycle of 95 °C (30s), 30 cycles of 95 °C (30 s), 55 °C (1 min) and 68 °C (3 min), and final extension at 68 °C (5 min). The PCR product amplified with primers OSC F1 5′-ACACATAGAAGGTCACAGCACG ATG-3′ and OSC R1 5′- CAACCTCCATCGTCTTTCTGTGTTC-3′ was examined on 1% agarose gel and cloned in T/A cloning vector pGEM-T easy (Promega, USA). The ligated product was transformed into *Escherichia coli* host strain (DH5α) and plated on LB agar plates containing ampicillin (100 μg/mL). Then, the plasmids were isolated from selected clones and subjected to sequencing.

The RACE PCR protocol was implemented to isolate full length gene. Two forward and two reverse gene specific primers were designed from the partial CaOSC sequence are as follows: Race OSC F 5′ -GGA ACTCCGCAAGGAACTGTACTTAC -3′, Race OSC Nested F 5′-CATCGTG CTGTGACCTTCTATGTGT -3′ for 3′ RACE and Race OSC R 5′-GTAAG ACATACGAAAAGTGTGTGGCCATG -3′, Race OSC Nested R 5′- GAAC ACAGAAAGACGATGGAGGTTG-3′ for 5′ RACE. Gene Racer Kit (Invitrogen, USA) was used for RACE PCR and 3′ RACE and 5′ RACE ready cDNA was prepared from total RNA. The PCR amplification procedures for 3′ RACE and 5′ RACE was done using primers Race OSC F/Gene Racer 3′ primer (provided with the kit) and Race OSC R/Gene Racer 5′ primer (provided with the kit) respectively. The 3′ and 5′ RACE nested PCR products were cloned in pGEM-T easy vector and sequenced. Based on 3′ and 5′ RACE sequence data, the primers from start and stop codon, CaOSC–F (5′-ATGTGGAAGTTGAAGATAGGAGA GGG-3′) and CaOSC-R (5′- TAATTTGACAAGGCAAATGGAACTCG-3′) were used for PCR amplification under the conditions: 10 s at 98 °C, 30 s at 55 °C, 3 min at 72 °C for 30 cycles. Additionally, the actin gene was amplified as a positive control using the primer Actin-F (5′- ATG GCCAACGCAGCTTCAGGAATGGCTGTT-3′) and Actin- R (5′-TTAGTTG GCGCGGTTTCTGAAAACATCAAG-3′). The PCR product was purified using a PCR purification kit (Qiagen), and the final volume adjusted to 30 μl with sterilized water. Resulting PCR product were cloned into pGEM-T Easy vector and transformed into *E. coli*. Plasmid DNAs from several transformants were isolated and sequenced using a Big Dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA). Blastx analysis was performed to ensure the homology of the obtained nucleotide sequences.

2.3. Sequence analysis

The sequence homology and the deduced amino acid sequence comparisons were performed in the GenBank database using the Blastn and Blastx algorithms. The deduction of the amino acid sequences, calculation of theoretical molecular mass and pI, was performed with ExPASy Proteomic tools provided at http://www.expasy.ch/tools/. The full-length nucleotide sequences obtained were translated using the Translate tool. The sequences were translated to identify open reading frames using the ORF Finder in the NCBI database (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Predicted amino acid sequences were phylogenetically analysed against OSC sequences of different plant species recouped from the GenBank™ through Blastp algorithm. Multiple alignments of the deduced amino acid sequences were carried out with the ClustalW program (http://www.ebi.ac.uk/clustalw/), and a phylogenetic tree was constructed using MEGA 7.0 program by Neighbor-Joining method (Kumar et al., 2018), and reliability of nodes

has been tested with 1000 bootstrap replicates.

2.4. Functional expression of OSC cDNA in Saccharomyces cerevisiae mutant GIL77

The full-length cDNA of CaOSC was obtained by a forward primer with a Not I site (5′-AAGGAAAAAAGCGGCCGCAAAAACAATGTGGAA GTTGAAGATAGGAGAGGG -3′) and a reverse primer with a Spe I site (5′- GGACTAGT TAATTTGACAAGGCAAATGGAACTCG -3′). The PCR amplification program was designed as follows: at 98 °C 98 °C for 10 s, at 55 °C for 30 s, 72 °C for 3 min (30 cycles). The purified PCR product was digested with Not I and Spe I restriction enzymes (New England Biolabs, Hitchin, UK) and ligated into yeast expression vector pYES derived from GAL1 promoter with T4 DNA ligase (Promega, Heidelberg, Germany). The plasmid was transformed into yeast mutant GIL77 by electroporation. The transformation of yeast was conducted by the lithium acetate method with selection on synthetic complete agar lacking uracil (SC-Ura) (Kawai et al., 2010). The transformed yeast was inoculated through two sets of experiments, the first set was carried out in shaking flask and the other on a large volume in a bioreactor.

2.4.1. Shake flask cultivations

The transformant colonies were grown on plates containing selected medium at 28 °C. A single colony was used to inoculate 2 mL of the selected medium, which was incubated overnight at 28 °C and 250 rpm. Subsequently, 250 mL shake flasks containing 20 mL of SC-Ura medium supplemented with ergosterol (20 μg/mL), were inoculated at an OD_{600} of 0.1, and incubated at 28 °C and 250 rpm. After 2 days, cells were collected and resuspended to a new SC-Ura medium supplemented with 2% galactose for induction and further cultured in shake flasks, at 28 °C and 250 rpm, for another 15 days.

2.4.2. Bioreactor cultivations

Similar to the shake flask cultivations described above, a single colony from a fresh plate containing selective medium was used to inoculate 5 mL of selected medium and incubated overnight at 28 °C and 250 rpm. Cells were then transferred to 50 mL of liquid medium in 500 mL shake flasks and cultivated overnight at 28 °C and 250 rpm. Subsequently, the preculture was harvested by centrifugation at 4000 rpm for 5 min and resuspended in 10 mL of fresh medium. The OD_{600} was measured using a Genesis20 spectrophotometer (Thermo Fisher Scientific) to inoculate the bioreactors with an initial OD_{600} of 0.1. All cultivations were performed in 5 L of SC-Ura, containing ergosterol (20 g/L), and incubated in a a bioreactor of Applikon vessels (Applikon, Delft, The Netherlands) using the DasGip Microbiology PD system (DasGip, Jülich, Germany). Homogenous mixing was enabled using two six-blade Rushton turbines at 600 rpm. The temperature was controlled at 28 °C and the pH was adjusted to 5 using 2M KOH and 2M HCl. The vessels were sparged with air at an initial rate of 60 L/h to assure constant aeration at 1 vvm. After 2 days, the bioreactor cell culture was supplemented with 2% galactose for induction and further cultured in bioreactor conditions was carried out for another 15 days. Two to 3 mL sample was removed from the reactor at regular interval for the measurement of OD_{600} and for GC-MS analysis.

2.4.3. Metabolite analysis

After galactose induction, the yeast cells were collected by centrifugation at 1750*g* for 5 min and refluxed with 0.5% pyrogallol in MeOH at 80 °C for 2 h. The saponification extract was partitioned with the same volume of petroleum ether twice. The solvent was dried out by a gentle stream of N_2 and derivatized with MSTFA +1% TMCS silylation in pyridine (1:1, 100 μL) (60 min at 80 °C), then directly analysed with GC-MS. The lupeol biosynthesis was determined for every 24 h during 15 days in both shake flasks and in bioreactor and the quantitative analysis was calculated based on an external standard curve run in tandem with the experimental samples.

2.5. Extraction and analysis of triterpenoid from C. arabica

The fresh leaves, siliques, stems and roots tissues were powdered with liquid N_2 then 50 mg of different plant parts were extracted with 5 mL of ethanol and 5 μL of an internal standard of cholestane (1 μg) was added for 1 h at room temperature. The ethanol extracts were partitioned with an equal volume of petroleum ether and water (1:1 (v/v)). The extraction with petroleum ether was repeated three times to maximize triterpene recovery. The petroleum extracts were then dried down and prepared with N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) for GC-MS analysis.

2.6. GC-MS analysis

The chemical profiles of plant and yeast petroleum ether extracts were performed by Agilent GC-MS system. This instrument consisted of a 7683 series auto sampler, a 7890A GC system, and a 5975C inert XL mass selective detector with a Triple-Axis Detector (Agilent Technologies). An aliquot of 1 μL of derivatized extract was analysed by GC-MS equipped with AGC column HP5-MS (30 m × 0.25 mm × 0.25 μm film thickness) and helium was used as the carrier gas at a flow rate of 0.9 mL/min. The GC parameters were employed to separate the terpenoids using the following temperature program: 200 °C for 0.5 min, ramp of 10 °C for 1 min to 270 °C followed by 3 °C per min ramp to 320 °C 10 min hold. The mass spectrometer was operated in electron impact mode at 70eV and in the scan range *m/z* of 30–400 Amu. The transfer line temperature was set at 270 °C and ion source temperature at 230 °C.

The GC/MS data was processed with ChemStation software (Agilent, Santa Clara, CA, USA) and Automated Mass Spectral Deconvolution and Identification System (AMDIS) (National Institute of Standards and Technology, United States Department of Commerce, Washington, DC, USA). The peaks for the various triterpenes and phytosterols were deconvoluted using AMDIS after baseline correction and identified by their relative retention time to authentic standards and compared with the mass spectra from commercial mass database (NIST17 mass spectral library, http://nistmassspeclibrary.com/). Quantification was based on measuring the area of peaks and performed with the use of external standard method, i.e., calibration curves prepared separately for known concentrations of the authentic standards of campesterol, stigmasterol, β-Sitosterol, β-amyrin, α-amyrin, lupeol (Sigma-Aldrich, Germany). Final data were counted as the means ± standard deviation of three independent samples.

2.7. Tissues expression analysis

Plant tissues (leaves, siliques, stems and roots) grown at different developmental stages were harvested after 2, 6, 10, 14 and 18 weeks, immediately frozen in liquid N_2 and stored at −80 °C. Total RNA was extracted using Trizol reagent (Invitrogen) from these samples and used for quantitative reverse-transcription (RT) PCR. The RNA (5 μg) was used for cDNA synthesis by superscript III Reverse Transcriptase (Invitrogen) following standard protocols. Quantitative Real-time PCR (qRT-PCR) was performed in triplicates by means of 2x SYBR Green PCR Master Mix in optical plates using ABI Step one real time quantitative PCR system (Applied Biosystems). SYBR green PCR (20 μL) contained 5 μL of cDNA template, 100 nM each of the primers, and 10 μL of 2x SYBR Green PCR Master Mix under following cycling conditions: 30 cycle of 98 °C for 1 min and 40 cycles of 94 °C for 10 s, 60 °C for 20 s, and 72 °C for 25 s. Actin was used as endogenous control using two primers: Atn-F (5′-ATGGCCGTCGTTGGAGTTCATCGATTCGC-3′) and Atn-R (5′- TCACAAGTGATGGTTGAGATCTTGTTCCAT -3′) and specifically designed internal primers for the CaOsc1-F (5′-ACACATAGAAGG TCACAGCACGATG-3′) and CaOsc1-R (5′-CAACCTCCATCGTCTTTCTG TGTTC-3′). Based on the comparative C_t method ($2^{-\Delta\Delta Ct}$), the gene expression levels of CaOSC, was calculated for each of the tissue and

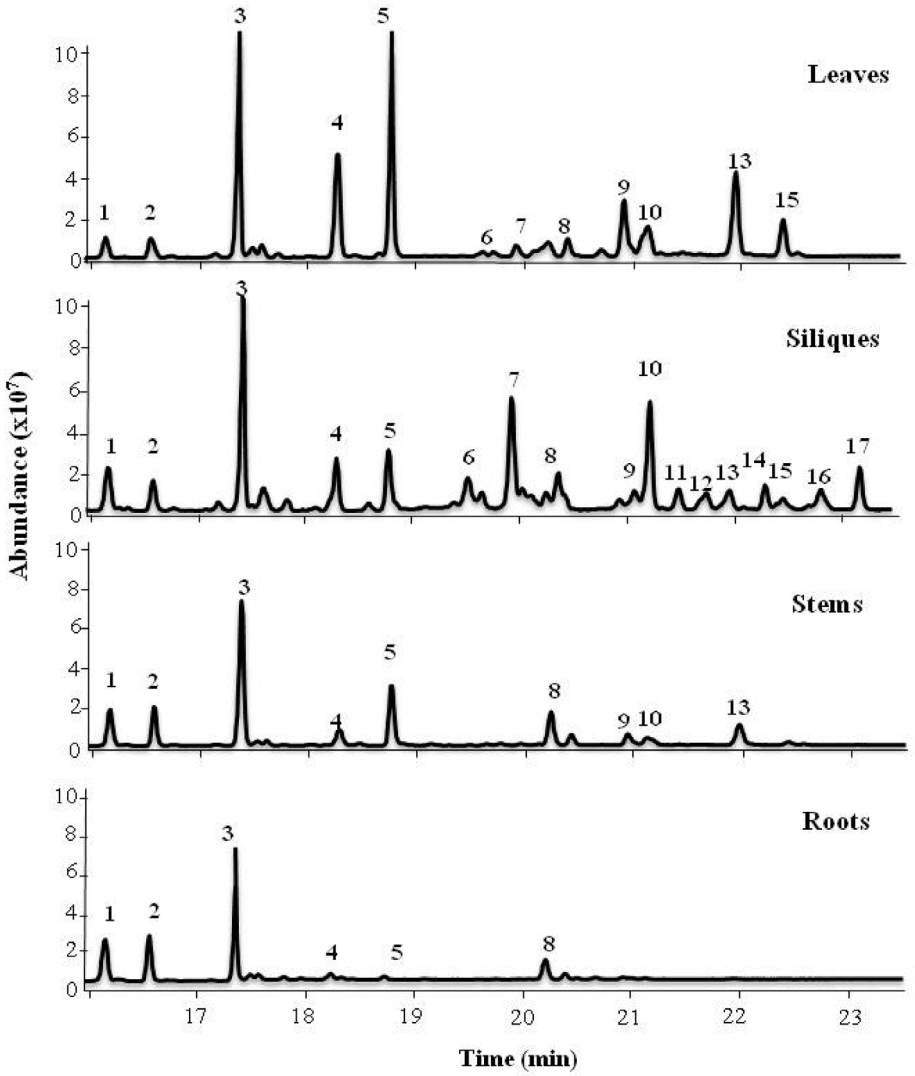

Fig. 1. Chemical profile of petroleum ether extracts of different tissues of *C. arabica* at mature fruiting stage (Peaks: **1**-Campesterol; **2**-Stigmasterol, **3**- β-Sitosterol; **4**-lupeol; **5-** β-amyrin; **6-** α-amyrin; (**7–17** are unknown terpenes).

evaluated on a comparison basis.

2.8. Statistical analysis

Data for campesterol, stigmasterol, β-Sitosterol, lupeol, β-amyrin and α-amyrin contents in *C. arabica* materials including root, stem, leaf and siliques parts were subjected to ANOVA and a posthoc LSD tests using PC software package SPSS (version 20.0; SPSS Inc., Chicago, IL, USA), to analyse content differences. The means were separated on the basis of least significant differences at 0.05 probability level.

3. Results

3.1. Terpenoids accumulation during phenological cycle of C. arabica

The terpenoids profile in leaves, stems, siliques and roots of *C. arabica* were determined through GC-MS analyses (Fig. 1). The terpenoids accumulation in *C. arabica* was significantly different among tissues during development stages. Depending on the plant material, the phytosterols (campesterol, stigmasterol and β-sitosterol) are mainly presented in all different plant tissues, while the pentacyclic triterpenes (lupeol, β-amyrin and α-amyrin) were accumulated with a significant disparity in different plant parts (Fig. 1). In total, seventeen terpenoids were mostly detected in silique, whereas leaf was marked by the absence of five triterpenoids (11–17). The chemical profile was completely different in stems and roots, which were only distinguished through nine and six compounds, respectively. Among the pentacyclic triterpenes, the lupeol and β-amyrin were the most abundant components in leaves and siliques compared to the other tissues. In addition, leaves and siliques possess α-amyrin that was not detected in either stems or roots.

Thus, the pentacyclic triterpenes were mainly influenced in the absolute amount during the development stage of the plant (Fig. 2). In fact, the β-sitosterol, α-amyrin, lupeol and β-amyrin levels were increased by 5-fold at fruiting stage, while the campesterol and stigmasterol were slightly increased by 2-fold. The highest level of β-sitosterol was revealed particularly in leaves and siliques (300 and 260 μg/g DW, respectively) at fruiting stage. In addition, leaves contained the highest levels of lupeol (320 μg/g DW) and β-amyrin (202 μg/g DW) at fruiting stage, while the lowest levels were noted in roots and ranged from 3.4 to 17.2 μg/g DW during development stage. Therefore, these triterpenes are promising compounds for plant development. Due to the significant potential of these compounds, it is crucial to study the triterpenes biosynthesis in *C. arabica* through oxidosqualene cyclase.

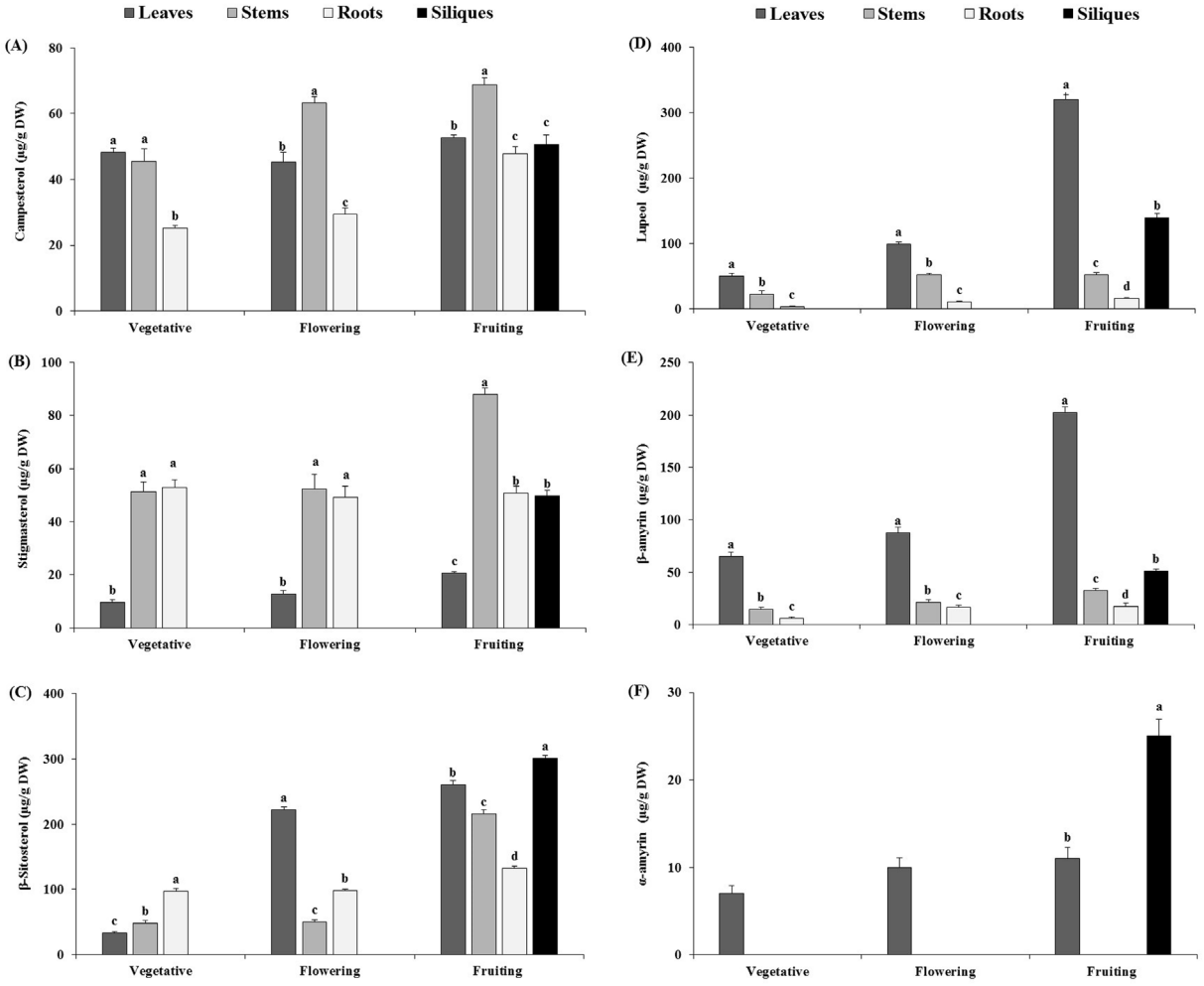

Fig. 2. Phenologic changes of campesterol (A), stigmasterol (B), β-sitosterol (C), lupeol (D), β-amyrin (E) and α-amyrin (F) in leaves, stems, roots and siliques of *C. arabica*. Values with different small letters-a, b, c, d-within columns for each development stage differ significantly at the level of $p < 0.05$ (LSD test) (bars are ± s.e.).

3.2. Cloning of OSC cyclase from C. arabica

Chemical confirmation of triterpenes production in different plant parts of *C. arabica* led us to prepare the cDNA from mature leaves. Based on highly conserved regions among reported plant OSCs, specific primers were designed using multiple sequence alignment principle by ClustalW.

The designed specific primers were used to amplify these sequences approximately 2289 bp open reading frame (ORF), encoding a 762 amino acid protein with a predicted molecular mass of 87.7 kDa. The full-length sequence of CaOSC is available in GenBank with accession number KX897945. The deduced amino acid sequences of CaOSC sharing 79.3%, and 75% identity with camelliol C synthase (CAMS1, At1g78955) and LUP1 (At1g78970) from *Arabidopsis thaliana,* respectively. This homology reveals also 78.5% and 75.4% with LUP4 (At1g78950) and LUP 2 (At1g78960), respectively (Fig. 3). This result provides evidence that CaOSC might in fact be a lupeol synthase. More detailed sequence analyses exhibited that CaOSC contains highly conserved DCTAE motif, which is implicated in substrate binding and four repeats of the QW motifs typical of the triterpene synthase superfamily (Fig. 3). The QW motifs may strengthen the structure of the enzyme and stabilize the carbocation intermediates during cyclization. To elucidate the phylogenetic relationship of deduced amino acid sequences of CaOSC proteins with other known members of the OSC superfamily (GenBank™) a phylogenetic analysis was performed with MEGA 7 software, phylogeny based on the neighbor-joining method (Fig. 4). The CaOSC grouped in accord with the amino acid correspondence, constituting four phylogenetic clusters: cycloartenol synthase, multifunctional triterpenes synthases, lupeol synthase, and β-amyrin synthase. Phylogenetic analysis showed that CaOSC was grouped with lupeol synthase and is more closely related to camelliol C from *A. thaliana* than to any other OSC (Fig. 4).

3.3. Characterization of the encoding OSC in C. arabica

The functional validation of lupeol synthase was cloned into the expression vector pYES-CaOSC under the control of galactose (GAL) promoter and transformed into the erg7-deficient yeast mutant GIL77, which synthesizes 2,3-oxidosqualene endogenously. The empty vector was also transformed into GIL77 as a negative control. The yeast was cultivated in YPDE medium, which was firstly inoculated into 15 mL cultures tubes containing 2 mL medium during the overnight and then inoculated to flasks in 20 mL of media shaking at 28 °C. The production of lupeol in flasks was compared to huge culture of 5 L in a fermenter with the same condition cultivated in SCE-ura for 15 days and was induced after 24 h with galactose (2%). The equivalent of two OD_{600} of yeast transformants expressing the pYES-CaOSC construct and the empty vector were harvested at different time points during induction, then extracted with petroleum ether. The extracts, which were analysed through GC-MS, showed a peak that was absent in the empty vector

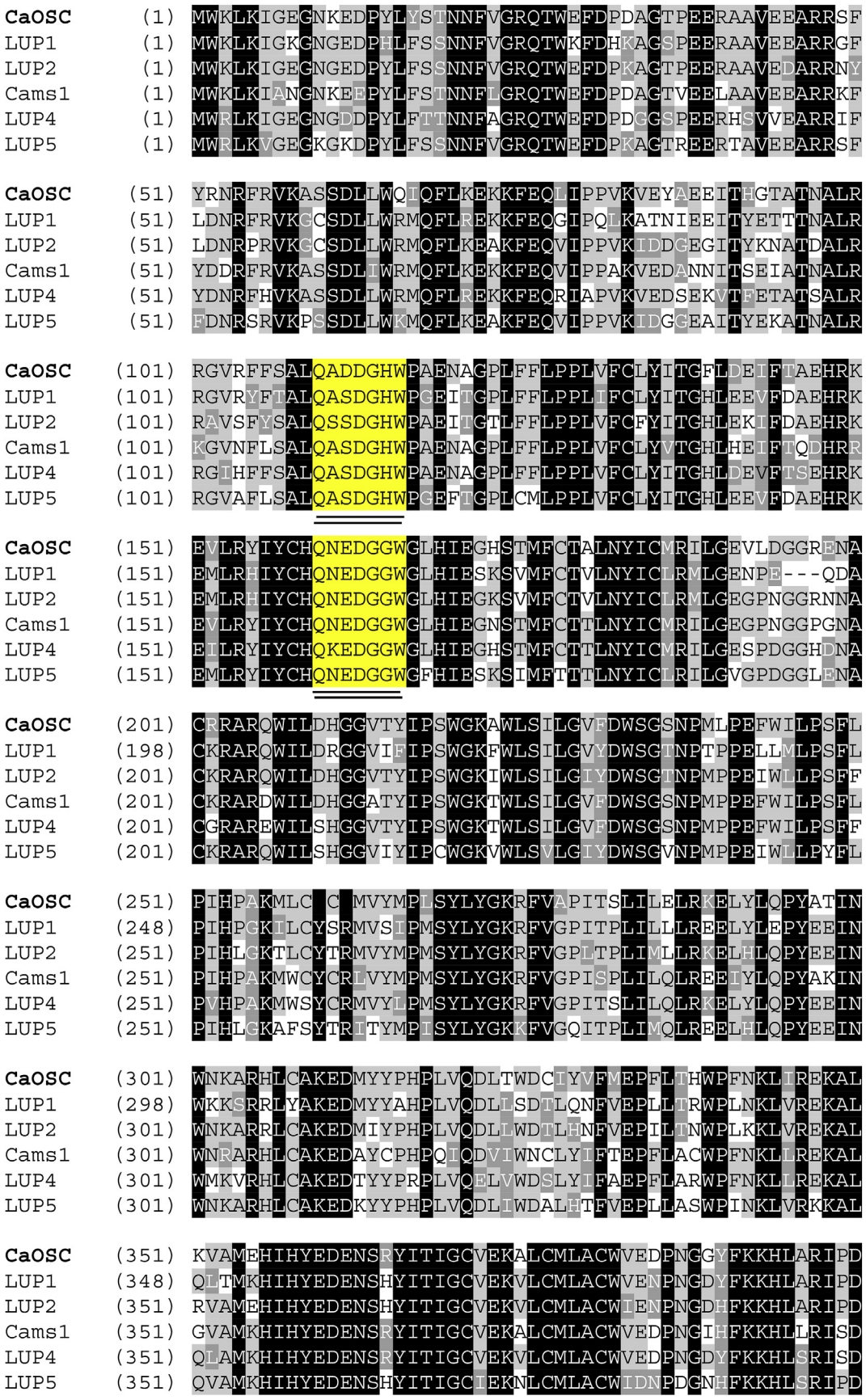

Fig. 3. Alignment of deduced amino acid sequences of *C. arabica* (CaOSC) with the sequences of *A. thaliana* camelliol C synthase Cams1 (at1g78955), *A. thaliana* lupeol synthase, LUP1 (At1g78970), LUP2 (At1g78960), LUP4 (At1g78950), and LUP5 (At1g66960). The DCTAE motif is underlined with black line, and the four QW motifs are marked using double black lines.

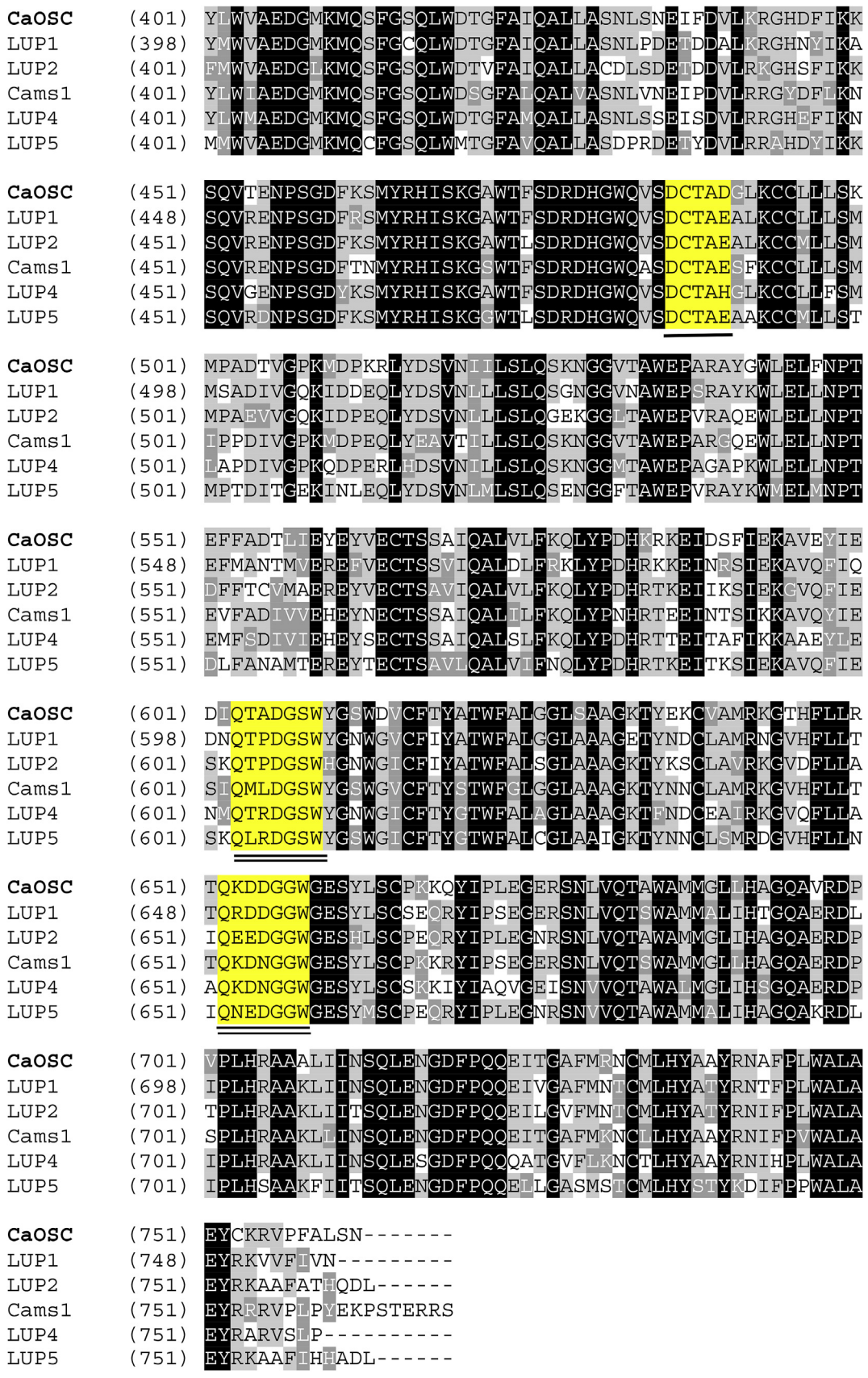

Fig. 3. (*continued*)

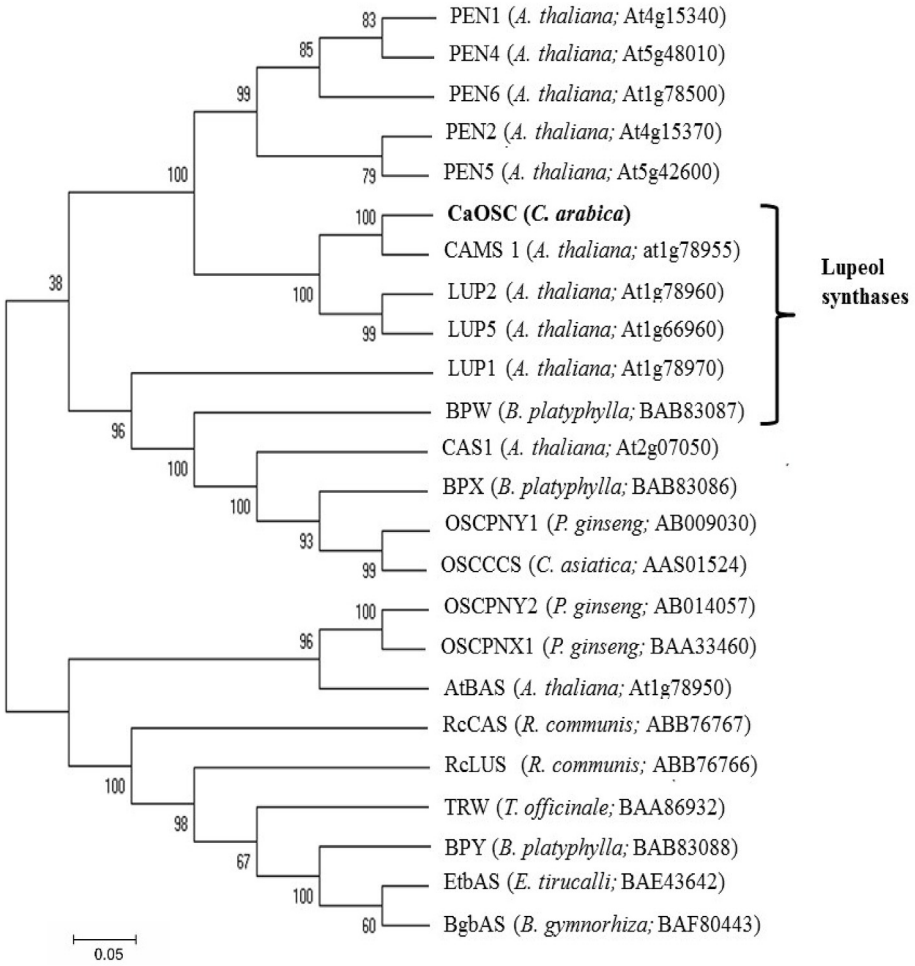

Fig. 4. Phylogenetic analysis comparing the CaOSC cDNA cloned from *Cleome arabica* with previously known OSCs from other plant species. The tree was constructed by MEGA 7 software using the neighbor-joining method, with an evolution method of bootstrap for 1000 times.

control and had identical retention time with authentic lupeol standard (Figs. 5 and 6). To elucidate the structure, MS characteristics were compared with the lupeol standard, revealed fraction peaks at *m/z* 189 and 498, which were also identical to the lupeol standard (Fig. 6). Consequently, we conclude that CaOSC is a monofunctional lupeol synthase.

The production of lupeol in yeast mutant Gill77 increased rapidly after induction, which was correlated with the growth of cell and time (Fig. 7). The production was exponentially increased to 46.14 ng/mL after 96 h of induction, which was then stabilized and reached the maximum of 51.29 ng/mL at the stationary phase of the yeast growth (Fig. 7A). To achieve greater volumetric productivity, a high density of cells was observed in fermentation process, and was developed in 5 L of SCE-ura, then the induction was occurred within 24 h after the glucose was exhausted. The maximum production of 2.33 μg/mL reached at 240 h and held constant until 288 h, then reduced with yeast growth (Fig. 7B). In conclusion, the productivity of lupeol in fermenter was higher than the flask with 45-fold.

3.4. Expression analysis of CaOSC transcripts in different tissue of C. arabica (qRT-PCR)

In order to investigate the expression levels of lupeol gene in different tissues of *C. arabica*, total RNA was extracted from roots, stems, leaves, siliques at different stages. In fact, the spatial expression pattern of CaOSC was expressed in all the tissues, with the strongest signal in leaves followed by siliques, stems and roots (Fig. 8). These data were also supported by semi-quantitative PCR. The temporal expression profile of CaOSC during roots, stems, siliques and leaves development was determined by qRT-PCR. The expression levels were mainly showed in leaves during the development stages. In contrast to these observations, the lupeol gene was slightly expressed at immature stage of different tissues and was not detected in early stage of roots. This level of expression increased during development stage and reached the maximum at 18 weeks.

4. Discussion

There is a considerable interest in chemical composition variation in Cleome species due to their numerous beneficial medicinal properties. The Cleome genus was distributed in all over the world, but *C. arabica* is the unique species that is presented in Tunisia. However, very few studies are conducted on the chemical and/or biological aspects of this plant. Among the phytochemical investigation, only the terpenoids were mainly described in aerial part of *Cleome arabica* (Ladhari et al., 2013, 2018), that could be associated with its biological activities. To our knowledge, our study is the first detailed study of triterpenoids in different tissues of *C. arabica* at different development stages. Here we show that the leaves and siliques have the important components of triterpenes, and the chemical composition was varied between the organs. According to our study, lupeol was identified for the first time in *C. arabica* and the maximum accumulation was recorded in leaves at the mature stage up to 320 μg/g DW followed by the siliques, stems and roots. Similar significant accumulation was observed by Guhling et al.

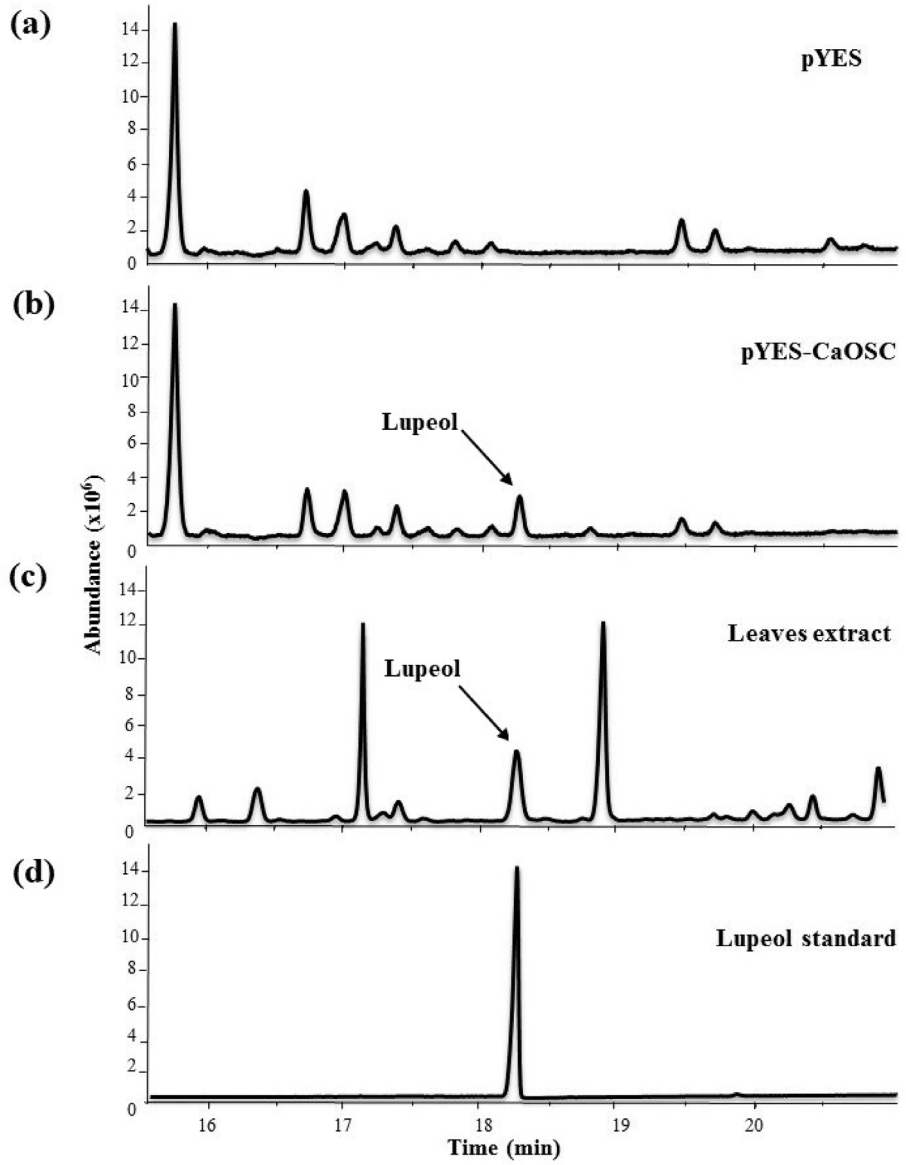

Fig. 5. Chromatographic analysis of GIL77 yeast extracts transformed with the empty expression vector pYES as negative control (a) and pYES-containing CaOSC cDNA (b). GC analysis of leaves petroleum extract (c) and authentic lupeol standard (d) (all samples TMS derivatized).

(2006) who reported that the cuticular lupeol was accumulated already in very early stages of the stem development of glaucous *R. communis*. In addition, several studies revealed that many compounds are accumulated in a particular compartment, and some are even translocated from source cells to sink organs via long distance transport (Gillissen et al., 2000; Dhaubhadel et al., 2003; Kim et al., 2014; Park et al., 2016). In fact, lupeol accumulation in *C. arabica* was considered in a scarce amount, and their biosynthetic machinery are inadequately understood in entirety. In the literature, lupeol was isolated from numerous plant species and its antitumor (Gauthier et al., 2006; Petronelli et al., 2009) and anti-inflammatory activities (Marquez-Martin et al., 2006; Soares et al., 2010) have been already demonstrated. Therefore, this pentacyclic triterpene is a promising compound for the development of new bioactive agents.

Detailed understanding of biosynthetic pathway of lupeol and its expression level across tissues of *C. arabica* are necessary to improve the terpenoids production. In terpenoids biosynthesis, cyclization of 2,3-oxidosqualene is the key step that determines the nature of triterpene scaffold. Numerous OSC genes encoding 2,3-oxidosqualene synthases that synthesize or are predicted to synthesize simple triterpenes have been identified from a wide range of different plant species (Haralampidis et al., 2001; Phillips et al., 2006). With this viewpoint, we have successfully cloned and characterized one member of the OSC superfamily belonging to protosterl cation group from *C. arabica*.

Indeed, the present study revealed that CaOSC is closely related to *A. thaliana* camelliol C synthase sharing four QW motifs (Poralla et al., 1994) and DCTAE motif (Abe and Prestwich, 1995), which are the characteristics of OSC superfamily including all lupeol synthases. As functional characterization of CaOSC in the yeast strain GIL77 yielded lupeol as the major cyclization product, we conclude that CaOSC encodes a *C. arabica* lupeol synthase. A similar result of metabolic conversion of lupeol by the yeast strain was reported on heterologous expression of lupeol synthase cDNAs from other plant species (Ebizuka et al., 2003). In our study, the lupeol production was increased by 45-fold in fermenter experiment compared to the flask experiment. The difference in production of lupeol could be explained by the controlled condition in fermenter compared to the flasks. The fermentation is a controlled process for producing active compound with a more reliable production method. The significant production during time of lupeol is reported here as a promising step towards development of a reliable

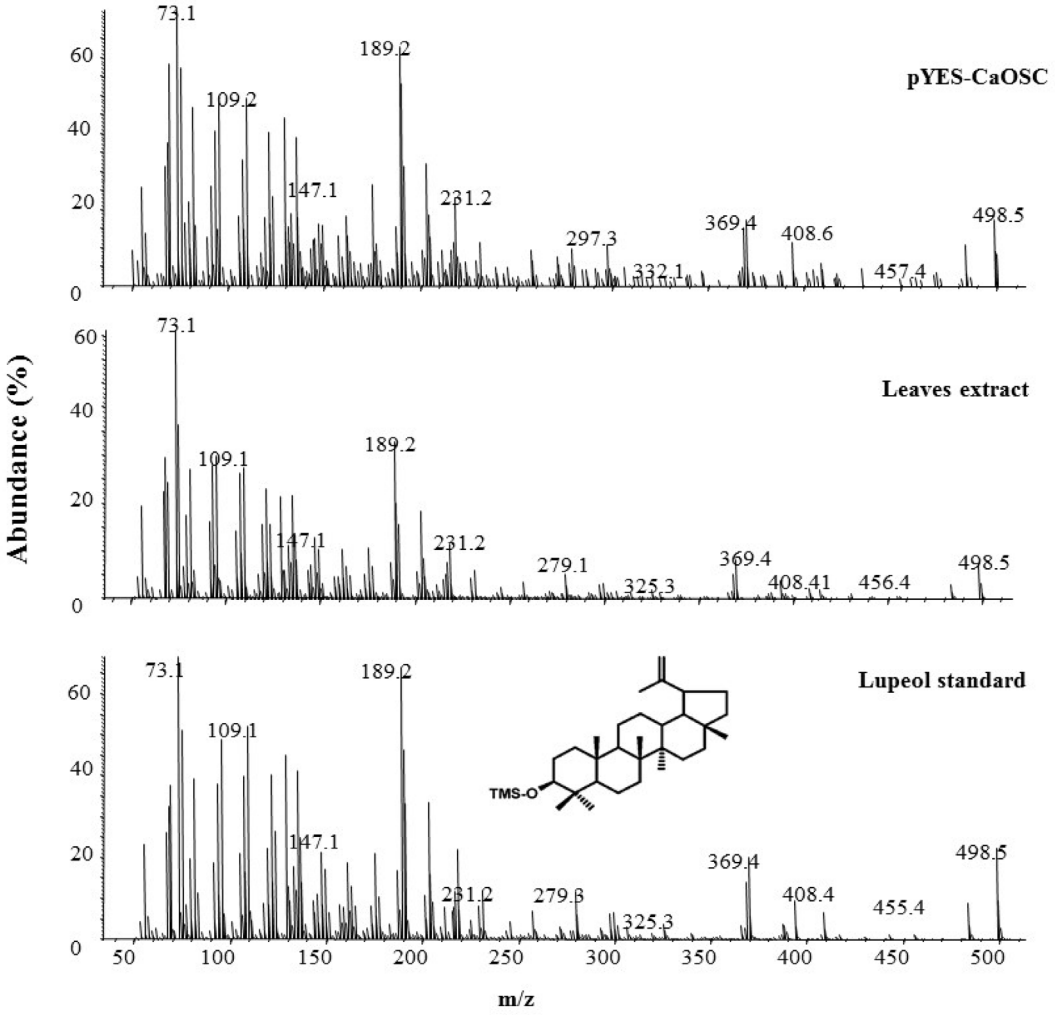

Fig. 6. Mass spectra of peak * from cell extract of transformed GIL77 and of authentic TMS-derivatized lupeol standard.

second source for economical production. In fact, improving production of terpenoids by different strategy was reported by Zhou et al. (2016), which they improve the betulinc acid production in yeast using multiple strategies, by including a higher BA-producing enzyme, assessment of two different yeast strains, and increasing of the BA pathway gene transcriptions. In addition, Bröker et al. (2018) revealed that the transformed yeast cells, that are with a newly isolated sequence encoding lupeol synthase from the Russian dandelion (*Taraxacum koksaghyz*), increased the yield of pentacyclic triterpenes by 127-fold. Besides, Henke et al. (2018) exhibited that the sesquiterpene alcohol (patchoulol) production from *Corynebacterium glutamicum* was increased by 30-fold in a bioreactor compared to the shake flasks. The strong improvement is likely due to improved oxygen input and saturation in the bioreactor compared to shake flasks and/or the consistency of pH in the controlled reactor. Liu et al. (2016) expressed the dammarenediol-II in engineered *Escherichia coli* and induced the production by changing the temperature and IPTG concentration with an important production at the lower temperature of 25 °C compared to the temperature of 37 °C. This production of lupeol precursor by fermentation will be an inexpensive chemical conversion and attractive supplement to current drug production methods.

The substantial increase of CaOSC expression in the leaves of *C. arabica* led us to examine the localization of the gene transcripts in order to investigate the possible physiological role of lupeol biosynthesis in the plant. The roles of *CaOSC* in plant development were examined by characterizing its spatial and temporal expression patterns. Its expression levels were varied with the organs during *C. arabica* development stages. The highest levels were clearly detected in leaves at the mature stage, demonstrating that lupeol may play an important role in *C. arabica* growth and development. The lupeol synthase expression pattern in *C. arabica* has been similarly reported for OSCs expression in other plants species. In addition, the lupeol accumulation in *C. arabica* leaves coincides with the higher lupeol expression level at the mature stage compared to the other tissues. This result was confirmed by Guhling et al. (2006), who reported that both the organ-specific expression of RcLUS and the expression pattern during development exactly matched the accumulation of cuticular lupeol in castor bean. The accumulation of pentacyclic triterpenoid (asiaiticosides) in *Centella* was higher in leaves and was corresponding with *OSC* transcript accumulation (Kim et al., 2005). In other reports, the accumulation of OSCs transcripts was highly expressed in leaves of *Gentiana straminea* (Liu et al., 2009) and *Catharanthus roseus* (Huang et al., 2012). Although, the highest expression level was recorded in flower and roots in the case of *Nigella sativa* (Scholz et al., 2009) and in hypocotyls of *Glycyrrhiza glabra* (Hayashi et al., 2001).

5. Conclusion

The phytochemical investigation of the different tissues of *C. arabica* revealed high levels of variability, that are associated with the stage of plant development, in the content of the investigated terpenes and phytosterols compounds. It appears that the lupeol was significantly

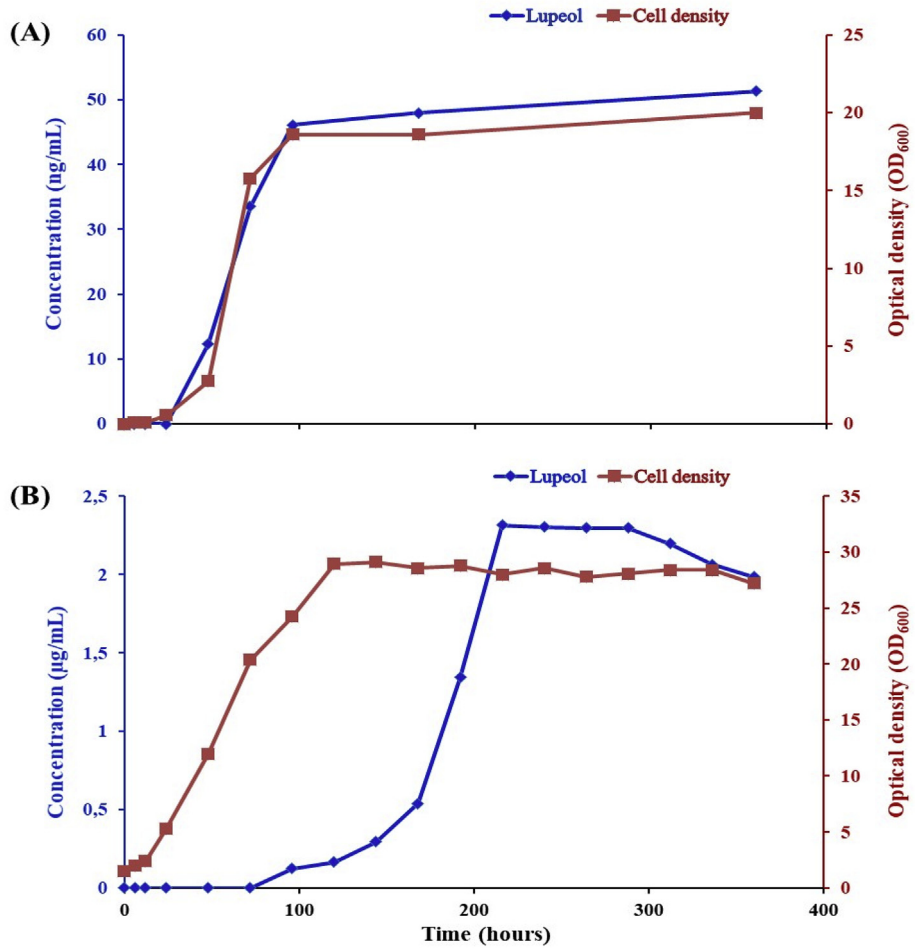

Fig. 7. Production of lupeol in Gill77 induced after 24 h in shake flask (A) and in fermenter (B) cultures for 15 days.

accumulated in leaves at maturity stage rather than in other tissue types. In addition, the lupeol biosynthesis was investigated through isolation and cloning of a full-length oxidosqualene cyclase cDNA (CaOSC) from the mature leaves. The successful cloning of new CaOSC cDNA has provided a sight of the molecular evolution of triterpene synthase in *C. arabica*. The lupeol production was enhanced in *Saccharomyces cerevisiae* by 45-fold in fermenter. The metabolic pathways for this pentacyclic triterpene should be explained to obtain a better understanding of the terpenoids production in different tissues of *C. arabica*. Further studies are required regarding genetic modifications on yeast strain as well as suitable bioprocessing strategies.

Author contribution

Ladhari A. performed the experiments, analysed the data and wrote the paper under the supervision of Prof. Chappell. J.

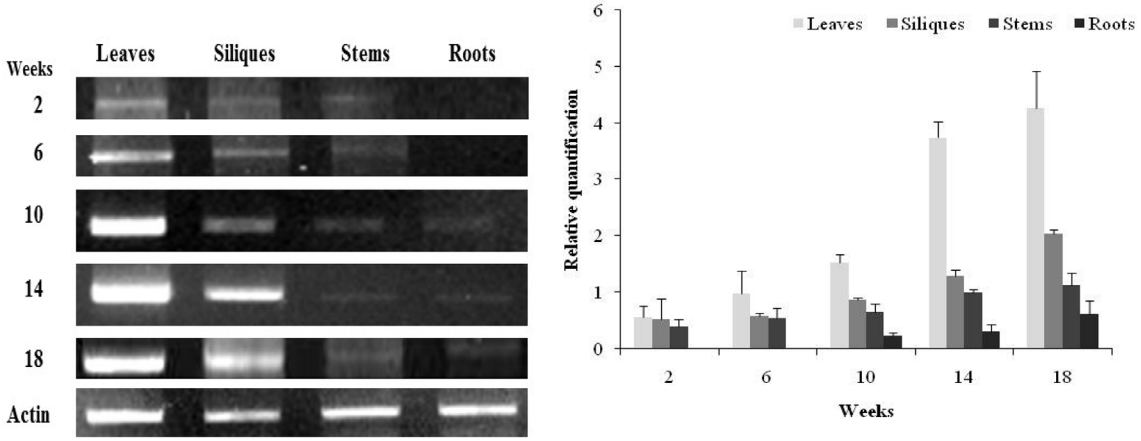

Fig. 8. Quantitative real time-PCR (qRT-PCR) analysis of the expression levels of lupeol synthase in different tissues of *C. arabica* during development stage. The data are presented as mean ± SE from three experimental analyses.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

This work was financially supported by Fulbright grant (CIES-ECA-68150488) from the Council for International Exchange of Scholars, On behalf of the U.S. Department of State's Bureau of Educational and Cultural Affairs (ECA).

References

Abe, I., Prestwich, G.D., 1995. Identification of the active site of vertebrate oxidosqualene cyclase. Lipids 30, 231–234.
Ahmed, Z.K., Kazmi, S.N., Malik, A.A., 1990. A new pentacyclic triterpene from *Abutilon pakistanicum*. J. Nat. Prod. 33, 1342–1344.
Basyuni, M., Oku, H., Inafuku, M., Baba, S., Iwasaki, H., Oshiro, K., Okabe, T., Shibuya, M., Ebizuka, Y., 2006. Molecular cloning and functional expression of a multi-functional triterpene synthase cDNA from a mangrove species *Kandelia candel* (L.) Druce. Phytochemistry 67, 2517–2524.
Basyuni, M., Oku, H., Tsujimoto, E., Kinjo, K., Baba, S., Takara, K., 2010. Triterpene synthases from the Okinawan mangrove tribe, Rhizophoraceae. FEBS J. 274, 5028–5042.
Bohlmann, J., Keeling, C.I., 2008. Terpenoid biomaterials. Plant J. 54, 656–669.
Brendolise, C., Yauk, Y.K., Eberhard, E.D., Wang, M., Chagne, D., Andre, C., Greenwood, D.R., Beuning, L.L., 2011. An unusual plant triterpene synthase with predominant α-amyrin-producing activity identified by characterizing oxidosqualene cyclases from Malus × domestica. FEBS J. 278, 2485–2499.
Bröker, J.N., Müller, B., Deenen, N.V., Prüfer, D., Gronover, C.S., 2018. Upregulating the mevalonate pathway and repressing sterol synthesis in Saccharomyces cerevisiae enhances the production of triterpenes. Appl. Microbiol. Biotechnol. 102, 6923–6934.
Calegario, G., Pollier, J., Arendt, P., Oliveira, L.S.D., Thompson, C., Soares, A.R., Pereira, R.C., Goossens, A., Thompson, F.L., 2016. Cloning and functional characterization of cycloartenol synthase from the red seaweed *Laurencia dendroidea*. PLoS One 11 e0165954.
Chappell, J., Wolf, F., Proulx, J., Cuellar, R., Saunders, C., 1995. Is the reaction catalyzed by 3-hydroxy-3-methylglutaryl Coenzyme-A reductase a rate-limiting step for isoprenoid biosynthesis in plants? Plant Physiol. 109, 1337–1343.
Dhaubhadel, S., McGarvey, B.D., Williams, R., Gijzen, M., 2003. Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol. Biol. 53, 733–743.
Djeridane, A., Yousfi, M., Brunel, J.M., Stocker, P., 2010. Isolation and characterisation of new steroid derivative as a powerful antioxidant from *Cleome arabica* in screening the in vitro antioxidant capacity of 18 Algerian medicinal plants. Food Chem. Toxicol. 48, 2599–2606.
Ebizuka, Y., Katsube, Y., Tsutsumi, T., Kushiro, T., Shibuya, M., 2003. Functional genomics approach to the study of triterpene biosynthesis. Pure Appl. Chem. 75, 369–374.
Gauthier, C., Legault, J., Lebrun, M., Dufour, P., Pichette, A., 2006. Glycosidation of lupine-type triterpenoids as potent *in vitro* cytotoxic agents. Bioorg. Med. Chem. 14, 6713–6725.
Gillissen, B., Bürkle, L., André, B., Kühn, C., Rentsch, D., Brandl, B., Frommer, W.B., 2000. A new family of high-affinity transporters for adenine, cytosine, and purine derivatives in arabidopsis. Plant Cell 12, 291–300.
Guhling, O., Hobl, B., Yeats, T., Jetter, R., 2006. Cloning and characterization of a lupeol synthase involved in the synthesis of epicuticular wax crystals on stem and hypocotyl surfaces of *Ricinus communis*. Arch. Biochem. Biophys. 448, 60–72.
Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., Osbourn, A., 2001. A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. PNA S. 98, 13431–13436.
Hayashi, H., Huang, P., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Shibuya, M., Ebizuka, Y., 2001. Cloning and characterization of a cDNA encoding β-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. Biol. Pharm. Bull. 24, 912–916.
Henke, N.A., Wichmann, J., Baier, T., Frohwitter, J., Lauersen, K.J., Risse, J.M., Peters-Wendisch, P., Kruse, O., Wendisch, V.F., 2018. Patchoulol production with metabolically engineered *Corynebacterium glutamicum*. Genes 9, 219–233.
Huang, L., Li, J., Ye, H., Li, C., Wang, H., Liu, B., Zhang, Y., 2012. Note added in proof to: molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. Planta 236, 1571–1581.
Kawai, S., Hashimoto, W., Murata, K., 2010. Transformation of Saccharomyces cerevisiae and other fungi methods and possible underlying mechanism. Bioeng. Bugs 6, 395–403.
Kempinski, C., Jiang, Z., Zinck, G., Sato, S.J., Ge, Z., Clemente, T.E., Chappell, J., 2019. Engineering linear, branched-chain triterpene metabolism in monocots. Plant Biotechnol. J. 17, 373–385.
Kim, O.T., Kim, M.Y., Huh, S.M., Bai, D.G., Ahn, J.C., Hwang, B., 2005. Cloning of a cDNA probably encoding oxidosqualene cyclase associated with asiaticoside biosynthesis from *Centella asiatica* (L.) Urban. Plant Cell Rep. 24, 304–311.
Kim, Y.B., Thwe, A.A., Li, X., Tuan, P.A., Lee, S., Lee, J.W., Arasu, M.V., Al-Dhabi, N.A., Park, S.U., 2014. Accumulation of astragalosides and related gene expression in different organs of astragalus membranaceus BGE. Var mongholicus (BGE.). Molecules 19, 10922–10935.
Kindermans, J.-M., Pilloy, J., Olliaro, P., Gomes, M., 2007. Ensuring sustained ACT production and reliable artemisinin supply. Malar. J. 6, 125–130.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Kushiro, T., Shibuya, M., Ebizuka, Y., 1998. In: Ageta, H., Aimi, N., Ebizuka, Y., Fujita, T., Honda, G. (Eds.), In towards Natural Medicine Research in the 21st Century. Elsevier Science, Amsterdam, pp. 421–427.
Ladhari, A., Haouala, R., DellaGreca, M., 2014. New dammarane triterpene from *Cleome arabica*. Chem. Nat. Compd. 50, 684–686.
Ladhari, A., Omezzine, F., DellaGreca, M., Zarrelli, A., Zuppolini, S., Haouala, R., 2013. Phytotoxic activity of *Cleome arabica* L. and its principal discovered active compounds. South Afr. J. Bot. 88, 341–351.
Ladhari, A., Tufano, I., DellaGreca, M., 2018. Influence of new effective allelochemicals on the distribution of *Cleome arabica* L. community in nature. Nat. Prod. Res. 1–9. https://doi.org/10.1080/14786419.2018.1501688.
Liu, Y., Zhao, Z., Xue, Z., Wang, L., Cai, Y., Wang, P., Wei, T., Gong, J., Liu, Z., Li, J., 2016. An Intronless β-amyrin synthase gene is more efficient in oleanolic acid accumulation than its paralog in *Gentiana straminea*. Sci. Rep. 6, 33364.
Liu, Y.L., Cai, Y.F., Zhao, Z.J., Wang, J.F., Li, J., Xin, W., Xia, G.M., Xiang, F.N., 2009. Cloning and functional analysis of a β-amyrin synthase gene associated with oleanolic acid biosynthesis in *Gentiana straminea* Maxim. Biol. Pharm. Bull. 32, 818–824.
Marquez-Martin, A., De La Puerta, R., Fernandez-Arche, A., Ruiz-Gutierrez, V., Yaqoob, P., 2006. Modulation of cytokine secretion by pentacyclic triterpenes from olive pomace oil in human mononuclear cells. Cytokine 36, 211–217.
Nagaya, H., Tobita, Y., Nagae, T., Itokawa, H., Takaya, K., Halim, A.F., Abdel-Halim, O.B., 1997. Cytotoxic triterpenes from *Cleome africana*. Phytochemistry 44, 1115–1119.
Park, Y.J., Valan, A.M., Al-Dhabi, N.A., Lim, S.S., Kim, Y.B., Lee, S.W., Park, S.U., 2016. Expression of terpenoid biosynthetic genes and accumulation of chemical constituents in *Valeriana fauriei*. Molecules 21, 691–706.
Petronelli, A., Pannitteri, G., Testa, U., 2009. Triterpenoids as new promising anticancer drugs. Anti Canccer Drugs 20, 880–892.
Phillips, D.R., Rasbery, J.M., Bartel, B., Matsuda, S.P.T., 2006. Biosynthetic diversity in plant triterpene cyclization. Curr. Opin. Plant Biol. 9, 305–314.
Poralla, K., Hewelt, A., Prestwich, G.D., Abe, I., Reipen, I., Sprenger, G., 1994. A specific amino acid repeat in squalene and oxidosqualene cyclases. Trends Biochem. Sci. 19, 157–158.
Scholz, M., Lipinski, M., Leupold, M., Luftmann, H., Harig, L., Ofir, R., Fischer, R., Prüfer, D., Müller, K.J., 2009. Methyl jasmonate induced accumulation of kalopanaxsaponin I in *Nigella sativa*. Phytochemistry 70, 517–522.
Selloum, L., Sebihi, L., Mekhalfia, A., Mahdadi, R., Senator, A., 1997. Antioxidant activity of *Cleome arabica* leaves extract. Biochem. Soc. Trans. 25, 608.
Soares, D.C., Andrade, A.L., Delorenzi, J.C., Silva, J.R., Freire-de-Lima, L., Falcão, C.A., Pinto, A.C., Rossi-Bergman, B., Saraiva, E.M., 2010. Leishmanicidal activity of *Himatanthus sucuuba* latex against *Leishmania amazonensis*. Parasitol. Int. 59, 173–177.
Sun, R., Liu, S., Tang, Z.Z., Zheng, T.R., Wang, T., Chen, H., Li, C.L., Wu, Q., 2017. β-amyrin synthase from *Conyza blini* expressed in *Saccharomyces cerevisiae*. FEBS Open Bio 7, 1575–1585.
Takhi, D., Ouinten, M., Yousfi, M., 2011. Study of antimicrobial activity of secondary metabolites extracted from spontaneous plants from the area of Laghouat, Algeria. Adv. Environ. Biol. 5, 469–476.
Tsichritzis, F., Abdel-Mogip, M., Jakupovic, J., 1993. Dammarane triterpenes from *Cleome africana*. Phytochemistry 33, 424.
Tu, Y.Y., 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat. Med. 17, 1217–1220.
Van Beilen, J.B., Poirier, Y., 2007. Establishment of new crops for the production of natural rubber. Trends Biotechnol. 25, 522–529.
Wu, S., Schalk, M., Clark, A., Miles, R.B., Coates, R., Chappell, J., 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat. Biotechnol. 24, 1441–1447.
Xu, R., Fazio, G.C., Matsuda, S.P., 2004. On the origins of triterpenoid skeletal diversity. Phytochemistry 65, 261–291.
Zhou, C., Li, J., Li, C., Zhang, Y., 2016. The improvement of betulinic acid biosynthesis in yeast employing multiple strategies. BMC Biotechnol. 16, 59.
Zhuang, X., Chappell, J., 2015. Building terpene production platforms in yeast. Biotechnol. Bioeng. 112, 1854–1864.