GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

# Cloning and expression analysis of terpenoid metabolism-related gene *NsyCMS* in *Nicotiana sylvestris*

**J. Lv, G.S. Liu, W. Li, Z. Hussian and Y.H. Sun**

Key Laboratory for Tobacco Gene Resources,
Tobacco Research Institute of Chinese Academy of Agricultural Sciences,
Ministry of Agriculture, Qingdao, China

Corresponding author: Y.H. Sun
E-mail: grace.andrew@163.com

Genet. Mol. Res. 14 (2): 3300-3308 (2015)
Received May 9, 2014
Accepted October 23, 2014
Published April 13, 2015
DOI http://dx.doi.org/10.4238/2015.April.13.9

**ABSTRACT.** Terpenoids constitute the main class of secondary metabolites produced in plants with industrial, pharmacological, and agricultural interests. *Nicotiana sylvestris* has been widely adopted as a diploid model system in plant biology for studies of terpenoid biosynthesis. In this paper, we report the isolation and analysis of the 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*CMS*) gene of the MEP (methylerythritol 4-phosphate) pathway from *N. sylvestris*. We used homologous-based cloning with a RACE method to obtain the full-length coding sequence of the *NsCMS*. Then, the physical and chemical properties, function, and three-dimensional structure of the NsyCMS protein were predicted. Fluorogenic quantitative PCR was used to conduct an expression analysis at different developmental stages of various tissues of the *NsyCMS*. The sequence of the *NsyCMS* consists of a 954-bp open reading frame and encodes a predicted protein of 317 amino acids, with a molecular weight of approximately 49.6 kDa and pi of 6.92. The *in vivo* localization of the encoded protein was cytoplasmic

 

with no signal peptide, whereas 2 transmembrane regions were found in NsyCMS. The conserved domains of typical 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, aminotransferase, and pyridoxal phosphate-dependent transferase were found in NsyCMS. Differential expression patterns of the *NsyCMS* were observed throughout the different developmental stages and tissues. NsyCMS messenger RNA was expressed in all tissues, with the highest level of expression in the seedling leaves. NsyMK was expressed at a higher level in the resettling roots. The results from our study set the foundation for exploring the terpenoid biosynthetic pathways in *N. sylvestris*.

**Key words:** *Nicotiana sylvestris*; Fluorogenic quantitative PCR; 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; Terpene biosynthesis

## INTRODUCTION

Plants produce an enormous variety of naturally occurring compounds called terpenoids, which comprise more than 40,000 compounds. The wealth of these naturally occurring compounds contributes in many different ways to the various characteristics of plants, such as flavor, taste, pigmentation, and aroma (Vezzaro et al., 2012), and participates in various plant growth and developmental processes (Bouvier et al., 2005). All isoprenoid compounds are derived from 2 basic five-carbon precursors, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), that further undergo consecutive condensations to produce prenyl diphosphates. The different prenyl diphosphates intermediate precursors (GPP, FPP, and GGP) are subsequently converted by terpene synthase into a variety of terpene carbon skeletons. Among all living organisms, the 2 five-carbon isoprene precursors (i.e., IPP and DMAPP) to these terpenoids are synthesized by 2 different pathways [i.e., methyl-D-erythritol-4-phosphate (MEP) and mevalonic acid (MVA) pathways] (Vranová et al., 2013). The MVA pathway utilizes mevalonic acid as a precursor for the synthesis of IPP and DMAPP, whereas MEP utilizes pyruvate and glyceraldehyde-3-phosphate as precursors for the synthesis of IPP and DMAPP. Remarkable advances have been made in the last 2 decades after the discovery of the MEP pathway as an alternative for the biosynthesis of IPP and DMAPP.

Recently, the MEP pathway represents one of the most promising targets for developing new herbicides and antiparasitic drugs; it also aims to improve the nutritional value of crop plants (Cordoba et al., 2009). The genes and their respective enzymes involved in the MEP pathway are attractive targets for the production of commercially important products (e.g., flavors, fragrances, pigments, and polymers) (Phillips et al., 2008) and the development of new antibacterial and antiparasitic drugs and herbicides (Kuntz et al., 2005).

The MEP pathway operates through the participation of 8 consecutive enzymes to produce the IPP and DMAPP universal basic blocks of isoprenoid compounds from the precursors pyruvate and D-glyceraldehyde 3-phosphate (Cordoba et al., 2009). All steps involved in the MEP biosynthetic pathway have been identified, and the respective gene involved in coding the specific enzyme has been cloned from different plant species. In this regard, *DXR* was the first gene identified and is considered as the rate-limiting step in the MEP pathway. Overexpression of *DXR* genes in various plant species (Kuntz et al., 2005), including *Arabidopsis*

 

(Estévez et al., 2001), tomato (Lois et al., 2000), potato (Morris et al., 2006), and *Ginkgo biloba* (Gong et al., 2006), resulted in a 2-7 fold increase in the level of various isoprenoid final products, including chlorophyll, carotenoids, tocopherols, and ABA. An interesting alternative to explore in the future will be the genetic manipulation of this enzyme in conjunction with other downstream enzymes of the MEP pathway (Cordoba et al., 2009). In this context, conversion of 2C-methyl-4-phosphate-D-erythritol to 2C-methyl-erythritol-4-phosphate is considered the next committed step in the MEP pathway. The gene responsible for synthesis of the enzyme catalyzing this step has been cloned from plant species such as *Arabidopsis* and rice. A major concern in the *CMS* today is the continued improvement of plant terpenoid biosynthesis.

*N. sylvestris* is grown as an ornamental plant and has been widely adopted as a diploid model system in plant biology for studies of terpenoid biosynthesis, plastid engineering, herbicide resistance, and plant virus resistance (Maliga and Svab, 2011; Sierro et al., 2013). The cloning and identification of genes involved in terpenoid biosynthesis have been carried out in different plant species, but little attention has been given to *N. sylvestris*. Isolation of the genes involved in catalyzing the early step of the terpene biosynthetic pathway will open new paradigms to explore the synthesis of terpenoids in the diploid model species of *N. sylvestris*. Genetic manipulation of the genes involved in isoprenoid biosynthesis in *N. sylvestris* will offer new corridors for the biosynthesis of nutritionally and commercially important metabolites.

Keeping in mind the importance of terpenoid synthesis in *N. sylvestris*, the main theme of the current paper was the exploration and experimentation of the *CMS* in *N. sylvestris*. First, the *CMS* involved in the synthesis of the five-carbon backbone for the synthesis of downstream terpenoids was cloned; then, the expression profile was identified using analyses of *in vivo* localization, the physical and chemical properties, and three-dimensional structure. Lastly, future directions for research are considered.

## MATERIAL AND METHODS

### Plant sample preparation

The seeds of *N. sylvestris* were grown under controlled conditions in a glass house. The roots, stems, and leaves were harvested from the plants and stored at -70°C.

### Preparation of total RNA

Total RNA was extracted using the GeneJET™ Plant RNA Purification Mini Kit (MBI Fermentas, USA), and the integrity of those samples was assessed by agarose gel (1%) electrophoresis. The concentration and purity of the RNA samples were detected using Nano Drop 2000. The first-strand complementary DNA (cDNA) was synthesized using the Revert Aid™ First Strand cDNA Synthesis Kit (MBI Fermentas).

### Cloning of cDNA

The *CMS* of *Salvia miltiorrhiza* (GenBank: AEZ55667.1) was used in our TBLASTN search against the EST database. One sequence (GB: EB430262.1, 864 bp) retrieved after a BLAST search was found to be highly homologous to the *S. miltiorrhiza CMS*. The primer pairs Mid-F and Mid-R, based on the conserved region, were designed to amplify the middle

fragment of the *N. sylvestris CMS*. The polymerase chain reaction (PCR) program used to amplify the middle fragment was as follows: 10 min at 94°C for 1 cycle; 30 s at 94°C, 30 s at 50°C, and 1 min at 72°C for 35 cycles; and 10 min at 72°C. Rapid amplification of cDNA ends (RACE) was used to amplify the 5' and 3' ends of the *CMS*. RACE was performed with the SMART™ RACE cDNA Amplification Kit (Clontech) according to manufacturer instructions. Gene-specific primers (GSP) based on the middle sequence of *N. sylvestris* were designed (Table 1). GSP1 combined with the UPM primer was used to amplify the 5'-RACE fragment with the following PCR conditions: 10 min at 94°C; 30 s touch-down from 70 to 50°C in 0.5°C increments; followed by 20 cycles of 30 s at 94°C, 30 s at 50°C, and 90 s at 72°C; and, finally, 10 min at 72°C. GSP2 combined with the UPM primer was used to amplify the 3'-RACE fragment with the following PCR conditions: 10 min at 94°C; 30 s touch-down from 60 to 50°C in 0.5°C steps; followed by 20 cycles of 30 s at 94°C, 30 s at 50°C, and 90 s at 72°C; and finally, 10 min at 72°C. According to the assembled sequences, we designed the full-length gene primers ORFF and ORFR (Table 1) to isolate the full-length sequence of the *CMS*. The PCR conditions used to isolate the full-length sequence were as follows: 10 min at 94°C; and 35 cycles of 30 s at 94°C, 30 s at 55°C, and 90 s at 72°C. The PCR products were separated on 1% agarose gel via electrophoresis; then, the respective bands were recovered (TaKaRa Agarose Gel DNA Purification Kit Ver. 2.0) and subcloned into the PMD18-T vector (TaKaRa, China) for sequencing (BGI-Beijing).

**Table 1.** Primers used in cloning and expression analysis of the NsyCMS.

| Primer | Primer sequence |
|---|---|
| Mid-F | ATGGCTATGGCGTCTTCACTATTC |
| Mid-R | AGCTTCTTCCATGAATACCGAAGAAGGC |
| UPM | 5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3';<br>5'-CTAATACGACTCACTATAGGGC-3' |
| GSP1 | 5'-GCTTCTGATGAAATTTTCATC-3' |
| GSP2 | 5'-ATGGCTATGGCGTCTTCACTATTC-3' |
| ORFF | GATCTTGAGCCTTGACTTATACTGC |
| ORFR | ACACCAAACTTCGACCACCTTTA |
| NsyCMS-YF | TACAAAGACCAAAAGTGAGCCCGC |
| NsyCMS-YR | GAAGAACAACAGTATGAGCAGCA |
| NsyMK-YF | ATGGAGGTGAGAGCAAGAGCCCC |
| NsyMK-YR | TGTTTCGTCATTATCAGCAGGAG |
| Actin-F | CAAGGAAATCACCGCTTTGG |
| Actin-R | AAGGGATGCGAGGATGGA |

## Bioinformatic analysis of the *CMS* sequence

The open reading frame (ORF) of the *NsyCMS* sequence was deduced by the ORF finder software at NCBI. DNAMAN, InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/), WOLF PSORT (http://psort.hgc.jp/form.html), and DAS (http://www.sbc.su.se/~miklos/DAS/) were used to predict the physicochemical property, structure, localization, and transmembrane helices of the encoded protein. ClustalW (http://www.ebi.ac.uk/clustalw/) and MEGA 4.0 were used for alignment and phylogenetic analysis (bootstrap = 100). A 3-D structural model was constructed with SWISS-MODEL (http://swissmodel.expasy.org/) using Modeller 8.0 (Fiser and Šali, 2003). The secretory protein was also analyzed by Signal P4.0 Server (http://www.cbs.dtu.dk/services/SignalP/).

## Gene expression analysis

The gene-specific primer pairs NsyCMS-YF, NsyCMS-YR, NsyMK-YF, and NsyMK-YR were used for fluorogenic quantitative PCR (Table 1), and the actin gene (GenBank: AB181991) served as the endogenous control. The best annealing temperature was finally determined by gradient PCR. Fluorogenic quantitative PCR was performed with the SYBR Premix EX Taq (Takara) on the ABI 7500 Real-Time PCR System (Applied Biosystems, USA). PCR was carried out with 2 µL cDNA as the template, 0.5 µL of each primer, 7.5 µL 2X SYBR Green PCR Master Mix, and 4.2 µL deionized water in a 15-µL reaction mixture. The PCR program was as follows: one cycle of 5 min at 95°C; and 40 cycles of 30 s at 95°C, 5 s at 95°C, and 34 s at 60°C. Each sample was quantified in triplicate for each biological replicate, and 3 biological replicates were conducted. The fluorescence signal was collected and the dissociation curve plotted. For standard curve drawing, a template concentration of 243 ng/µL was used; for species-specific gene amplification, the template was 3-fold serially diluted from 243 ng/µL to 3 pg/µL. Based on the CT values of all dilution points in a series, a standard curve was generated.

Expression levels for each sample were recorded as CT. CT values were transformed to quantities relative to the sample. Using $2^{\Delta\Delta CT}$ as the PCR amplification efficiency:

$$\text{Fold Change}=2^{-\Delta\Delta CT}=[(C_T{}^{(1)}\ \text{gene of sample}-C_T\ \text{internal control})\text{sample A} -(C_T\ \text{gene of sample}-C_T\ \text{internal control})\text{sample B}]$$

(1) CT: the threshold cycle

# RESULTS

## Isolation and sequencing of the ORF

A BLAST search of the *CMS* against the tobacco genome database revealed one conserved gene sequence in *N. sylvestris*. The middle gene sequence of 462 bp from *N. sylvestris* was obtained by PCR performed with the Mid-F and Mid-R primer pair (Figure 1A). Based on the middle sequence of the *CMS*, sense and antisense primers for 3'-RACE and 5'-RACE were designed, and both the 5' and 3' end fragments of the respective genes were amplified (Figure 1B, C). The 5' and 3' end fragments were assembled, and the full-length putative coding sequence was obtained. The complete nucleotide sequence contained a 954-bp ORF (Figure 1D). The sequence was renamed as NsyCMS and registered in GenBank (accession No. KC961733).

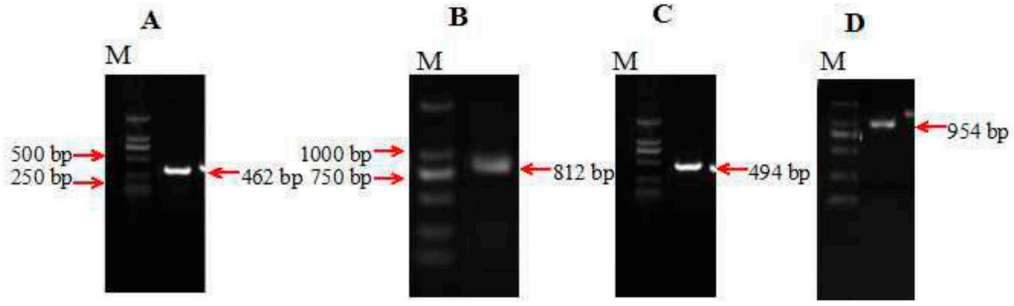


**Figure 1.** Agarose gel electrophoresis of PCR products of the *NsyCMS*.

## Bioinformatic analysis of the *NsyCMS* sequence

NsyCMS encodes a deduced protein of 317 amino acids, with a mass of 49.6 kDa and a pI of 6.92. Potential modification site prediction with NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) revealed 11 serine phosphorylation sites, 4 threonine phosphorylation sites, and one tyrosine kinase phosphorylation site (Figure 2), which indicates that the NsyCMS protein might undergo active post-translational modification to adjust its expression level or enzymatic activity. The secondary protein structure was 10.3% alpha-helix, 40.0% beta-sheet, and 49.7% random coil and appeared as a disordered structure, in general (Figure 3). Three-dimensional structure modeling utilized *Arabidopsis thaliana* 2c-methyl-D-erythritol 2,4-cyclo-diphosphate synthase (Figure 4) with a sequence identity of 91.57%.

No signal peptide was found in the NsyCMS encoded protein that revealed cytoplasmic localization of the protein rather than that of the chloroplast (chloroplast 9.0, mitochondria 3.0). It is inferred that NsyCMS may directly transfer phosphate groups to MVA in the chloroplast after its initial synthesis in the cytoplasm and protein trafficking. One transmembrane helix between 100-150 aa was tested; the grand average of hydropathicity of NsyCMS was 0.19. Both hydrophilic and hydrophobic amino acids were evenly distributed, indicating that NsyCMS is a stable, water soluble protein in addition to the GHMP_kinases_N and GHMP_kinases_C conserved domains.

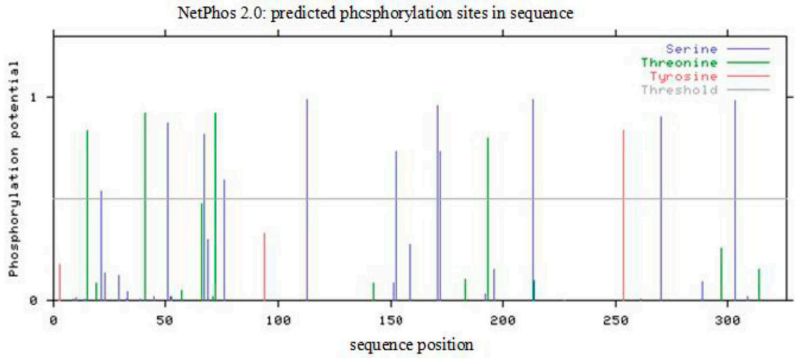


**Figure 2.** Predicted phosphorylation sites of the *NsyCMS*.

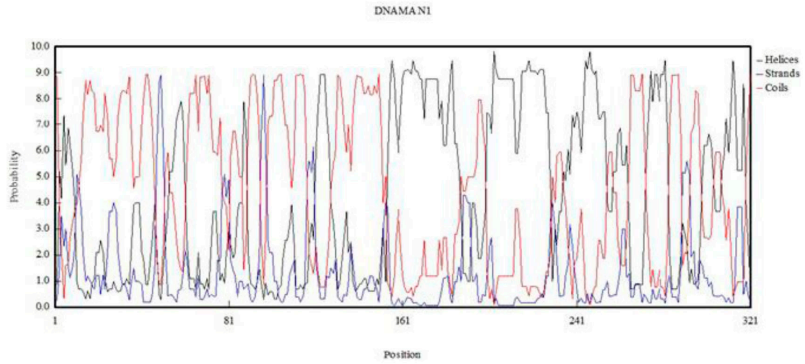


**Figure 3.** Secondary structure prediction of the *NsyCMS*.

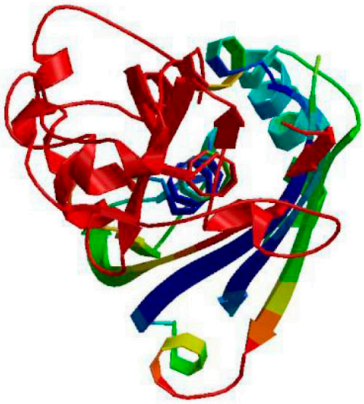

**Figure 4.** Three-dimensional structure modeling of the *NsyCMS*.

## Sequence alignment and cluster analysis

A phylogenetic tree was constructed with the *CMS* sequence from other plants based on BLASTP via NCBI (Figure 5). In this tree, NsyCMS was located at the closest branch to *Picrorhiza* rhizome*, Astragalus membranaceus*, and *Olea europaea*, indicating a similar evolutionary history.

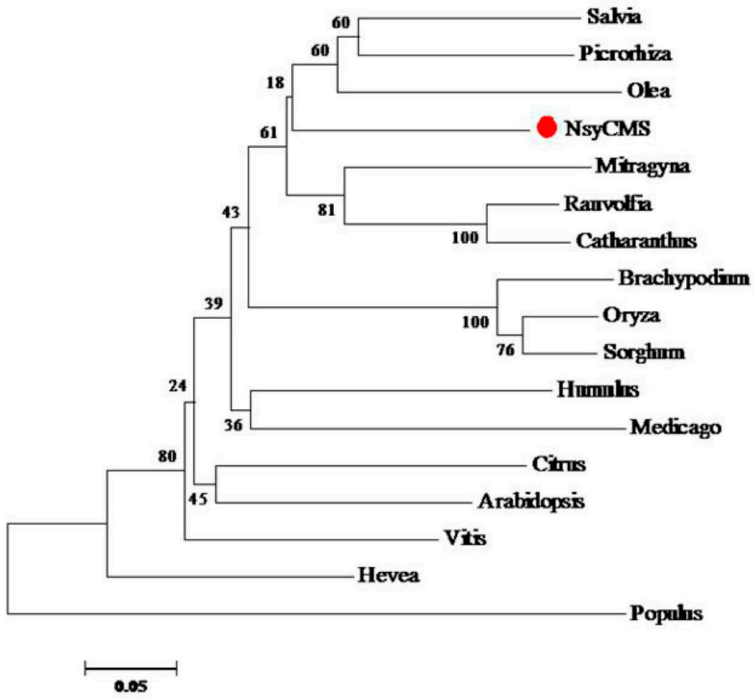


**Figure 5.** Phylogenetic relationships of *CMS* from various species.

## Expression patterns of *NsyCMS* and *NsyMK* at different developmental stages

The expression patterns of the *NsyCMS* and *NsyMK* at 2 different developmental stages were tested by fluorogenic quantitative PCR. The RNA transcription levels of the roots, stems, and leaves were measured across 2 developmental periods, including the seedling and resettling stages. The dissociation curve represented unimodal distribution, which demonstrates the exclusion of contamination and non-specific products (results not given). The repetition of each sample was relatively good. The expression level of NsyMK was high in the resettling roots, whereas extremely lower expression levels were observed in other plant parts. Moreover, the *NsyCMS* exhibited higher levels of expression in the seedling leaves, seedling roots, and resettling roots (Figure 6).

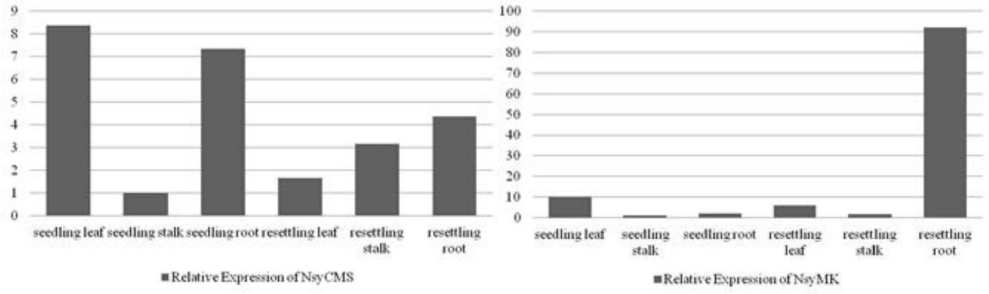


**Figure 6.** Expression levels of the *NsyMK* and the *NsyCMS*.

## DISCUSSION

The expression of NsyCMS may affect the content of terpenoid substances, particularly because MVA generates a precursor substance of isoprene in higher plants. The results suggest that NsyCMS belongs to the chloroplast enzyme, whereas NsyMK is a cytoplasm enzyme. In addition, NsyMK was expressed at a higher level in the resettling roots. However, NsyCMS exhibited a relatively strong expression in the seedling leaves and roots, which reflect various expression patterns during isoprenoid metabolic pathways. The results do not imply the effect of lack or overexpression of the *NsyCMS* function in *N. sylvestris*, which will be added to our follow-up study, including protein interactions and functional identification among different cultivars.

## REFERENCES

Bouvier F, Rahier A and Camara B (2005). Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* 44: 357-429.

Cordoba E, Salmi M and León P (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60: 2933-2943.

Estévez JM, Cantero A, Reindl A, Reichler S, et al. (2001). 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* 276: 22901-22909.

Fiser A and Šali A (2003). Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374: 461-491.

Gong YF, Liao ZH, Guo BH, Sun XF, et al. (2006). Molecular cloning and expression profile analysis of *Ginkgo biloba*

*DXS* gene encoding 1-deoxy-D-xylulose 5-phosphate synthase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Planta Med.* 72: 329-335.

Kuntz L, Tritsch D, Grosdemange-Billiard C, Hemmerlin A, et al. (2005). Isoprenoid biosynthesis as a target for antibacterial and antiparasitic drugs: phosphonohydroxamic acids as inhibitors of deoxyxylulose phosphate reducto-isomerase. *Biochem. J.* 386: 127-135.

Lois LM, Rodríguez-Concepción M, Gallego F, Campos N, et al. (2000). Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-d-xylulose 5-phosphate synthase. *Plant J.* 22: 503-513.

Maliga P and Svab Z (2011). Engineering the plastid genome of *Nicotiana sylvestris*, a diploid model species for plastid genetics. In: Plant chromosome engineering (James A) Springer, 37-50.

Morris WL, Ducreux LJ, Hedden P, Millam S, et al. (2006). Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life cycle. *J. Exp. Bot.* 57: 3007-3018.

Phillips MA, León P, Boronat A and Rodríguez-Concepción M (2008). The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* 13: 619-623.

Sierro N, Battey JN, Ouadi S, Bovet L, et al. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* 14: R60.

Vezzaro A, Krause ST, Nonis A, Ramina A, et al. (2012) Isolation and characterization of terpene synthases potentially involved in flavor development of ripening olive (*Olea europaea*) fruits. *J. Plant Physiol.* 169: 908-914.

Vranová E, Coman D and Gruissem W (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64: 665-700.