Egyptian Journal Of

GENETICS AND CYTOLOGY

INTERNATIONAL JOURNAL DEVOTED TO GENETICAL AND CYTOLOGICAL SCIENCES

Published by

THE EGYPTIAN SOCIETY OF GENETICS

Volume 51 **January 2022** **No. 1**

MOLECULAR CLONING AND CHARACTERIZATION OF TERPENE SYNTHASE 4 (*SgTPS4*) GENE FROM *Salvia guaranitica* PLANT

ESRAA A. ELSHERBENY[†], MOHAMMED ALI [†*], F. A. EL-RAMAH[†] AND MANAL K. AHMED[†]

[†]*Genetic Resources Department, Desert Research Center (DRC), 1, Mathaf El-Matariya Street, El-Matariya B.O.P 11753 El-matariya, Cairo, Egypt.*

* To whom correspondence should be addressed:Mohammed Ali, Ph.D.

Department of Genetic Resources Desert Research Center (DRC) 1 Mathaf El-Mataria St, 11753 Matariya, Cairo, Egypt.Phone: +2 01553727631 . Fax: (+202) 26357858

Email: mohammedalidrc@gmail.com

Profile URL: https://www.researchgate.net/profile/Mohammed_Ali268

Google Scholar: https://scholar.google.com/citations?hl=en&user=uWcwSCgAAAAJ

ORCID: https://orcid.org/0000-0001-9232-1781

[†] ESRAA A. ELSHERBENY and MOHAMMED ALI contributed equally to this work.

Abbreviations: OE: Overexpression, EOs: Essential oils. TPS: Terpene synthase. SgTPS4: *S. guaranitica* Terpene synthase 4. Semi-RT-PCR: Semiquantitative RT-PCR

Terpenoid is considered the largest group of natural products and a class of secondary metabolites, which have been identified from different plant species and many other organisms with more than 40,000 different structures (Bohlmann *et al.*, 1998). Terpenoid derives its shape from odd backbone

molecule called isopentenyl diphosphate (IPP), which have five carbon atoms (C5) (Wang *et al*., 2019 and Volke *et al*., 2019). The origin name of these different structures comes from the terebinth tree (*Pistacia terebinthus*), so we give these different structures names of terpene (Degenhardt *et al*., 2009). The structure of these units was illustrated by Wallach then modified by Ruzicka (Wallach, 1887; Ruzicka, 1953; 1959; 1973 and Pott *et al*., 2019). The plant produces multiple terpenoid compounds with highly diverse structures. Some terpenes are related to the primary plant metabolism such as the carotenoid pigments, phytol side chain of chlorophyll, gibberellin plant hormones, and phytosterols of cellular membranes (Trapp and Croteau, 2001; Gershenzon, 1999; Gutensohn *et al*., 2013 and Luck *et al*., 2020) and are important for plant growth and development. However, large majority of terpenes that have been identified are categories as secondary metabolites and play essential roles in the interactions of plants with the environment (Christianson *et al*., 2006). Both non-volatile and volatile terpenes have roles in such processes as the predators of herbivores and protection against photo-oxidative stress, attraction of both pollinators and the direct defense against insects and microbes (Tholl *et al*., 2006; K ollner *et al*., 2008 and Korankye *et al*., 2017). Numerous studies are found for understand in-depth the mechanisms of terpene and terpenoid functions.

The genus *Salvia* (*Lamiaceae*) includes over than 1,000 species of woody aromatic shrubs, among which e.g., *S. epidermindis, S. japonica, S. fruticosa, S. tuxtlensis, S. miltiorrhiza, S. aureus, S. przewalskii, S. santolinifolia, S. hydrangea, S. tomentosa, S. isensis, S. lavandulifolia, S. chloroleuca, S. glabrescens, S. nipponica, S. allagospadonopsis, S. macrochlamys* and *S. recognita* are economically important and cultivated worldwide for its vast medicinal properties and the production of their essential oils (EOs). Most of wild and cultivated salvia species are distributed in Central America, South America, East Asia and West Asia, while the remaining species are spread around the world (Alziar, 1988-1993; Ali *et al*., 2017 and Ali *et al*., 2018). Recently, *Salvia* species EOs have become a valuable source for aromatic and pharmaceutical research for discovering and identifying biologically active compounds (Takano and Okada, 2011; Ali *et al*., 2017 and 2018). Essential oils of *Salvia* species exhibit significant bioactivities, antimicrobial activities, including antimicrobial, anticancer, choleretic, anti-inflammatory, antioxidant and antimutagenic,

The fragrant oil of the *Salvia* mainly contains monoterpenes, sesquiterpenes, diterpene and triterpene. The composition of the terpenes in the salvia genus depends on the species or cultivars and type of tissues (Ali *et al*., 2017; 2018 and Aminfar *et al*., 2019). This study aimed at clone and functionally characterizes Terpene synthase 4 (*SgTPS4*) cDNA from *Salvia*

guaranitica. Here, we report the expression and functional characterization of *SgTPS4* cDNA in *Nicotiana tabacum*. The recombinant *SgTPS4* catalyses (2E, 6E)-farnesyl diphosphate to product bicyclogermacrene as a sesquiterpene through the pathway of sesquiterpenoid and triterpenoid biosynthesis.

MATERIALS AND METHODS

Plant materials and tissue collection

Plantlets of *S. guaranitica* L. were sampled from the Wuhan Botanical Garden farm, China. For gene cloning, three biological replicates from leaves were sampled from four years- old *S. guaranitica* plants. The samples were immediately frozen in liquid nitrogen and then stored at −20ºC until RNA extraction.

In silico analysis of *SgTPS4* gene

The nucleotide sequence of *SgTPS4* gene was selected from our previous RNA-Seq (Ali *et al*., 2018). The physiochemical properties of the *SgTPS4* were determined using PROTPARAM software(http://web.expasy.org/protparam). The amino acid sequencing for *SgTPS4* protein was further analyzed for protein subcellular location prediction using bioinformatics tools, WoLF PSORT Prediction(https://www.genscript.com/wolf-psort.html). Comparative sequence analysis of *SgTPS4* was performed using NCBI blastx against the protein database (http://blast.ncbi.nlm.nih.gov/).Phylogenetic tree was built using PhyML server with the default parameters of the (http://www.phylogeny.fr/) (Dereeper *et al*. 2008). To assess the phylogeny of the *SgTPS4* protein sequence in relation to other orthologous plant *TPS* genes, the protein sequences of functionally characterized *TPS* genes were retrieved from the National Center for Biotechnology Information (NCBI) database.

RNA extraction and cDNA library preparation

Total RNAs from three biological leaf replicates were extracted for *SgTPS4* gene cloning. Moreover, total RNAs from three biological replicates of *N. tabacum* were extracted for semi-quantitative RT-PCR using the TransZol Reagent (Focus Bioscience, Australia) and treated with DNase I (Takara). RNA quality was examined on 1.2% agarose gels, and the purity was analyzed using a Nanodrop ND1000 (NanoDrop technologies, Wilmington, DE, USA). RNAs from three replications were mixed into one tube for prepare RNA pools that will used to syntheses cDNA libraries. Two micrograms of total RNA (900 ng approximately) per sample was used for the synthesis of total cDNA with TransScript® First-Strand cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China). Afterwards, PCR was performed for cDNA synthesis at 42ºC for 15 min followed by 85ºC for 5 second

(Ali *et al*., 2017 and 2018).

Full-length terpene synthase cDNA clone and vector

Full-length cDNAs for *SgTPS4* was obtained by PCR amplification using short and long gene-specific primers based on RNA-Seq sequence information from our transcriptome sequencing of *S. guaranitica* leaves (Ali *et al*., 2017 and 2018). Leaf cDNA was used as a template for the initial PCR amplification and performed using short primers, such as *SgTPS4*

Forward:
5′-ATGAAACACCAACACTCTTCTCTCT-3′
Reverse:
5-TTCAGTGTTCATCTGTGATTACAACGATT-3

with the TaKaRa Ex Taq® DNA Polymerase (TaKaRa, China) under the following PCR conditions: 4 min at 96ºC followed by 12 s at 98ºC; 30 s at 58ºC (Annealing temperatures), 2.20 min at 72ºC, and then 10 min at 72ºC. This process was repeated for 30 cycles. The first PCR products was used as a template for the PCR cloning using long primers, such as *SgTPS4*

Forward:
5′-GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAAACACCAACACT-3′
Reverse:
5′-GGGGACCACTTTGTACAAGAAAG CTGGGTTTCAGTGTTCATCTGT-3′

with the TaKaRa Ex Taq® DNA Polymerase for the Gateway pDONR221 vector. The amplified PCR bands were purified from agarose gel and binding to pDONR221 vector, then our target gene were transfer to pB2GW7 overexpression vector for *N. tabacum* plant transformation. The positive construct vectors that containing our target gene was confirmed by sequencing.

Semiquantitative RT-PCR analysis

Semiquantitative real-time PCR was performed on a Eppendorf PCR (Master cycler Nexus PCR Machine from Eppendorf, UK) system with a total reaction volume of 25 µl. A gene-specific primer for *NtEF-1α* forward: 5′-TGGTTGTGACTTTTGGTCCCA-3′ and reverse: 5′-ACAAACCCACGCTTGAGATCC-3′ was used as a reference gene with 155 bp, and *SgTPS4* forward: 5′-ATCTTCGTGCTTTGCTACTC -3′ and reverse: 5′-ATTATGCGACTCGTCTTCTTC-3′ with 155 bp length, gene involved in the biosynthesis of Terpene synthase 4 (*SgTPS4*), were designed using the primer designing tools of IDTdna (http://www.idtdna.com/scitools/Applications/RealTimePCR/). Semi-qRT-PCR was ran using the following program [95°C for 4 min, 95ºC for 30 s, 58ºC for 30 s, 72ºC for 1 min and 72ºC for 10 min] for 35 cycles. The PCR products were resolved on 1.6 % agarose gel, and the expression levels of *NtEF-1α* and *SgTPS4* genes were detected.

***Nicotiana* plant growth conditions and preparation of *Agrobacterium* cultures for infection**

Wild-type *N. tabacum* plant seeds were grown under standard greenhouse conditions for ten days at our lab. Our construct vector pB2GW7-*SoTPS4* was

inserted into *Agrobacterium* strain EHA105 using direct electroporation method. Recombinant *A. tumefaciens* was grown for two days at 28ºC in solid LB media supplemented with 50 µg/ml each of rifampicin and spectinomycin. An individual colony was inoculated into 1.0 ml of liquid medium and grown at 28ºC under 200 rpm agitation overnight with the same media composition. After one day, 1.0 ml from liquid medium sample was transferred to a 250-ml conical flask containing 50 ml of LB media supplemented with the same compositions; the sample was grown at 28°C in a shaker overnight until an optical density of 0.7-1.0 (OD 600) was reached. Overnight cell culture was harvested by centrifugation at 4,500 rpm for 12 min at 4ºC, and the pellet was re-suspended in the infection medium (50 ml of LB-free media + 50 µl of acetosyringone). *N. tabacum* plantlet leaves were sampled and sterilized using 70% ethanol for 30 s, then 0.1% HgCl for 6 min, after that washing three times for 3 min using sterilized cold water. Then, leaves without petiole and midrib were cut into small pieces and soaked into infection media for 10 min. The transformation procedure was performed as described previously (Sunjung, 2006 and Ali *et al*., 2017). More than 12 individual transgenic tobacco lines were generated and examined with PCR for positive transgenic lines. The positive transgenic tobacco plants were selected for isolation the terpenoid.

Phenotypic evaluation

Transformed plants were watered and fertilized regularly with Miracle Gro fertilizer (Scott's Company, USA) prepared according to manufacturer's instructions for phenotypic comparisons between *N. tabacum* plants transformed with *SgTPS4* and its counterpart wild-type plants. Plants were grown in growth chamber at a temperature of 22ºC day/20ºC night with humidity of 60-70%, and photoperiod at 16 hours day/8 hours night, with a light density of 100–150 µ moles m−2 s−1 using fluorescent bulbs for vegetative growth and for flowering, respectively. Plants were assessed about leaf morphology, growth and terpene metabolic.

Metabolite extraction from transgenic *N. tabacum* leaves

Terpenoid compounds from non-transgenic *N. tabacum* leaves (wild type) and transgenic *N. tabacum* leaves containing *SgTPS4* expression construct were extracted and isolated. For this, twelve leaves from each transgenic *N. tabacum* line (one leaf from each plant) and wild type were homogenized in liquid nitrogen with a mortar and pestle, then the powder was soaked in Amber storage bottles ((20 ml screw-top vials with silicone/PTFE septum lids) (http://www.sigmaaldrich.com)) containing n-hexane as a solvent. After that, Amber storage bottles were

incubated in shaking at 37ºC and 210 rpm for 70 h. Afterward, the supernatant solvent was collected by centrifuged at 5,000 rpm for 10 minutes at 4ºC, then pipette into glass vials and concentrated to 1.5 ml of concentrated oils under a stream of nitrogen gas with a nitrogen evaporator (Organomation; Toption-China-WD-12). The concentrated oils were transferred to a fresh 1.5 ml crimp vial amber glass, and placed on the auto-sampler of the gas chromatography mass spectrometer (GC-MS) system for GC-MS analysis as described previously by (Ali *et al*. (2017 and 2018).

GC-MS analysis of essential oil components

Shimadzu model GCMS-QP2010 Ultra (Tokyo, Japan) system was used for GC analysis. An approximately 1µl aliquot of each sample was injected (split ratios of 15:1) into a GC-MS equipped with an HP-5 fused silica capillary column (30 m x 0.25 mm ID, 0.25 µm film thicknesses), and Helium at 1.0 ml/min^{-1} as carrier gas.. The mass spectra were monitored between 50- 450 m/z. Temperature was initially under isothermal conditions at 60ºC for 10 minutes. Temperature was then increased at a rate of 4ºC/min^{-1} to 220ºC, held isothermal at 220ºC for 10 minutes, increased by 1ºC/ min^{-1} to 240ºC, held isothermal at 240°C for 2 min, and finally held isothermal for 10 minutes at 350ºC. The volatile constituents were identified based on the mass spectra stored in the NIST Library (2014 edition), Volatile Organic Compounds (VOC), Wiley GC/MS Library (10th Edition) (Wiley, New York, NY, USA), and the Analysis S/W software. The relative % amount of each component was calculated by comparing its average peak area to the total areas. The GC-mass experiments was repeated three times with the same conditions, with total GC running time was 80 minutes (Ali *et al*., 2017 and 2018).

RESULTS AND DISCUSSION

Isolation of full-length terpene synthase 4 (*SgTPS4*) genes and sequence characterization

The *SgTPS4* gene has an open reading frame of 2289 bp, which encodes a 763 amino acid protein with a calculated molecular mass of 82.54 kDa and a theoretical isoelectric point (pI) of 9.59. The WoLF PSORT Prediction tools used to analyzed the *SgTPS4* protein subcellular location prediction, suggests that *SgTPS4* is localized at different organelles (such as, Mitochondrial, Chloroplast, Peroxisomal, Nuclear, Golgi and Vacuolar) with different presence and identity level from 13.9723% to 11.1732% (https://www.genscript.com/tools/wolfpsort/detail?file=2021/10/02/htdocs/results/163320872328765.detailed1.html#163320872328765). Based on the blastx analysis (Table 1), the closest homologue to *SgTPS4* is the Bicyclogermacrene synthase-like from *Salvia splendens*, which it shares 97.07 % identity. Although the level of amino acid

sequence similarity between *SoAMYS* and the other homologues was relatively higher (≧84.89%). On the other hand, Phylogenetic analysis of the deduced amino acid sequence of *SgTPS4* showed that it belongs to the TPS-c subfamily of angiosperm sesquiterpene synthases which may be encodes sesquiterpene and diterpenes (Chen *et al*., 2011; Bohlmann *et al*., 1998; Külheim *et al*., 2015 and Danner *et al*., 2011) (Fig. 1). To date, seven TPS subfamilies have been detected and identified in various plant species genomes, including *Selaginella moellendorffii* (Li *et al*., 2012), *Camellia sinensis* (Zhou *et al*., 2020), *Eucalyptus globulus* (Külheim *et al*., 2015), *Daucus carota* (Keilwagen *et al*., 2014), *Arabidopsis thaliana* (Aubourg *et al*., 2002), *Solanum lycopersicum* (Falara *et al*., 2011), *Malus domestica* (Nieuwenhuizen *et al*., 2013), and *Vitis vinifera* (Martin *et al*., 2010).

Functional characterization of Terpene synthase 4 (*SgTPS4*) genes in transgenic *N. tabacum* leaves

The role and product specificity of *SgTPS4* was determined by generating transgenic *N. tabacum*. Overexpression of *SgTPS4* in *N. tabacum* was accomplished using *A. tumefaciens* strain EHA105 harboring the transformation vector pB2GW7-*SgTPS4*. Using the *Agrobacterium*-mediated transformation method, more than twelve transgenic *N. tabacum* plants were successfully generated. These plants have large green oval leaves (Fig. 2A). In contrast, the non-transformants plants showed small green oval leaves (Fig. 2A). The putative transformants were further verified using semiquantitative RT-PCR of the plant genomic cDNA. Fully mature leaves from twelve putative transgenic plants and three wild type plants were collected for RNA extraction and cDNA synthesis. All the putative transformants showed high expression of the *SgTPS4* gene by the amplification of a distinct band at 155 bp, which was absent in the wild type plants (Fig. 2B). This result confirmed the presence of the *SgTPS4* gene in the genomes of the transgenic plants. Two of the transgenic plants, designated as *OE- SgTPS4-1* and *OE- SgTPS4-2*, were selected for further analysis. Meanwhile, from the morphological analysis, wild type plants showed a little delayed in growth with a few number of leaf compared to the transgenic plants (Figs. 2A and B). In context, the obtained findings are in line with our previous works of Ali *et al*. (2017 and 2018) who reported that the overexpression of genes that involved in the terpenoid biosynthesis, such as *SoLINS, SoNEOD, SoTPS6, SoSABS, SoCINS, SgGPS, SgFPPS* and *SgLINS* from *S. officinalis* and *Salvia guaranitica* in *N. tabacum* and *A. thaliana*, also resulted in delayed growth and flowering formation in wild type plants compared to the transgenic plants.

Metabolite extraction from transgenic and non- transgenic *N. tabacum* leaves

Phytochemicals were extracted from transgenic and non- transgenic (wild type) *N. tabacum* leaves with hexane and

analyzed by GC-MS to identify the specific product produced by transformation with the *SgTPS4* gene. Various types and amounts of terpene compounds were observed, and the quantities of terpene were represented by the percentage of peak area (% peak area). Compounds were identified in transgenic *N. tabacum* and non- transgenic (wild type plants) as the control by comparing their mass spectra of the compounds with mass spectra libraries. The detected components were also confirmed by comparing them with the published references and extracts of wild-type *N. tabacum* which produce different types and amounts of terpenoids. Overexpression of *SgTPS4* genes in *N. tabacum* plants produced different amounts of sesqui-, di- and triterpenes. Moreover, from the results shown in Table (2) and Fig. (3), very clear differences were observed for the transgenic plants, as an additional peak was present at the retention time of 50.461. This peak was characterized as Bicyclogermacrene compound, based on the closest mass spectra with the data stored in the Wiley GC/MS Library (10th Edition) (Wiley, New York, NY, USA) https://www.chromservis.eu/p/wiley-10th-edition-library-in-nist-format, volatile organic compounds (VOC) http://www.physchem.uni-wuppertal.de/voc-database , analysis S/W software https://www.acronymfinder.com/Software-(S%2FW).html, and the NIST Library (2014 edition) https://webbook.nist.gov/cgi/cbook.cgi?Name=hopanoid&Units=SI.The production of Bicyclogermacrene by *SgTPS4* was in agreement with the findings from Ali *et al*., (2017 and 2018) and Su-Fang *et al*., (2014). These results also showed that the overexpression of terpene syntheses genes introduced by Ali *et al*., (2017 and 2018) and Su-Fang *et al*., (2014), does not affect the product specificity of *SgTPS4* in producing Bicyclogermacrene. Having obtained the similar terpene products in both *N. tabacum* and *A. thaliana*, we have showed that *SgTPS4* was responsible for the production of Bicyclogermacrene as a sesquiterpene through the pathway of sesquiterpenoid (Wang *et al*., 2016 and Ro *et al*., 2006).

In conclusions, the diversity of the sesquiterpenes found in *S. guaranitica* renders this plant a major resource for research related to sesquiterpene biosynthesis. In this study, we cloned and functionally characterized one of the scarcely expressed sesquiterpene synthase (*SgTPS4*), which is responsible for the production of Bicyclogermacrene in *S. guaranitica*. Also, transgenic technology was applied by overexpressing *SgTPS4* in *N. tabacum*. Positive growth acceleration was clearly observed in the transgenic lines *OE-SgTPS4-1* and *OE-SgTPS4-2*. These two plants showed a high expression of the *SgTPS4* gene, which resulted in the production of Bicyclogermacrene. The Bicyclogermacrene produced in these *N. tabacum* transgenic plants indicated the effectiveness of *N. tabacum* in synthesizing the same product as a sesquiterpene through the common pathway of sesquiterpenoid and

triterpenoid biosynthesis. *SgTPS4* protein exhibits a strong sequence similarity to other sesquiterpene synthases, and clustered under TPS-c group. This research strongly suggests the potential usage of the *N. tabacum* plant as a model system for studying the Bicyclogermacrene synthase gene from *S. guaranitica* for understanding of plant sesquiterpenoid biosynthesis and the potential for biotechnology application.

SUMMARY

Salvia guaranitica is a medicinal and aromatic plant with highly valued in traditional medicine for its abundance of terpenes, especially the monoterpenes (C10) and sesquiterpenes (C15). Various terpenes were believed to contribute to the many useful biological properties in plants. This study aimed at cloning and functionally characterizes a full length sesquiterpene synthase gene from *S. guaranitica*. Terpene synthase 4 (*SgTPS4*) has a complete open reading frame (ORF) of 2289 base pairs encoding a 763 amino acids protein. The phylogenetic tree demonstrates that *SgTPS4* protein was clustered into the subfamily TPS-c, which belongs to the angiosperm terpenoid synthase. To examine the function of *SgTPS4*, we expressed this gene in *N. tabacum*. Two transgenic lines, designated as *OE- SgTPS4 -1* and *OE-SgTPS4 -2* were further characterized, both molecularly and functionally. The wild type plants showed a little delayed growth compared to the transgenic plants. Gas chromatography-mass spectrometry analysis of the transgenic plants showed that *SgTPS4* was responsible for the production of Bicyclogermacrene. This is the first report of a gene involved in the Bicyclogermacrene as a sesquiterpene from *S. guaranitica* plant.

Ethics approval and consent to participate

No investigations were undertaken using humans/human samples in this study. No experimental animals were used to conduct any of the experiments reported in this manuscript. Our study did not involve endangered or protected species.

Competing interests

The authors declare that they have no competing interests.

Funding

This research does not receive external funding

Consent for publication

Not applicable.

Authors' contributions

MA conceived and designed the study; MA, EAE, FAE and MKA performed experiments, MA wrote the paper. All authors discussed the results and commented on the manuscript and participated in the analysis of the data. All authors participated in reading and approving the final manuscript.

ACKNOWLEDGEMENTS

This work is under the activity of the Egyptian Desert Gene Bank (EDGB), Genetic Resources Department, Desert Research Center (DRC), Egypt. Special thanks are given to Dr. Ahmed Ali at Department of Plant Agricultural, Faculty of Agriculture Science, Al-Azhar University, Assiut, Egypt for valuable support. We also owe thanks to Dr. Wael Moussa at Desert Research Center (DRC) for constructive comment and help.

REFERENCES

Ali M., Hussain R. M., Rehman N. U., She G., Li P., Wan X., Guo L. and Zhao J. (2018). *De novo* transcriptome sequencing and metabolite profiling analyses reveal the complex metabolic genes involved in the terpenoid biosynthesis in blue anise sage (*Salvia guaranitica* L.). DNA Res., 25(6): 597-617. doi.org/10.1093/dnares/dsy028

Ali M., Li P., She G., Chen D., Wan X. and Zhao J. (2017). Transcriptome and metabolite analyses reveal the complex metabolic genes involved in volatile terpenoid biosynthesis in garden sage (*Salvia officinalis*). Sci. Rep., 7(1): 16074. doi.org/10.1038/s41598-017-15478-3

Alziar G. (1988-1993). Catalogue synonymique des *Salvia* L. dumonde (*Lamiaceae*). I.–VI. Biocosme Mesoge´en., 5 (3–4): 87-136; 6 (1–2, 4): 79–115, 163-204; 7 (1–2): 59-109; 9 (2–3): 413-497; 10 (3–4): 33-117.

Aminfar Z., Rabiei B., Tohidfar M. and Mirjalili M. H. (2019). Identification of key genes involved in the biosynthesis of triterpenic acids in the mint family. Sci. Rep., 9 (1): 15826. doi.org/10.1038/s41598-019-52090-z

Aubourg S., Lecharny A., Bohlmann J. (2002) .Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. Mol. Genet. Genom., 267:730-745. doi: 10.1007/s00438-002-0709-y.

Bohlmann J., Meyer-Gauen G. and Croteau R. (1998). Plant terpenoid synthases: molecular biology and phylogenetic analysis," Proc. Natl. Acad. Sci., USA., 95, 8: 4126-4133.

Chen F., Tholl D., Bohlmann J. and Pichersky E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J., 66 (1): 212-229. doi:10.1111/j.1365-313X .2011 .04520.x.

Christianson D. W. (2006). Structural biology and chemistry of the terpenoid cyclase's, Chem Rev.,

106 (8): 3412-3442.

Danner H., Boeckler G. A., Irmisch S., Yuan J. S., Chen F., Gershenzon J., Unsicker S. B. and Köllner T. G. (2011). Four terpene synthases produce major compounds of the Gypsy moth feeding induced volatile blend of *Populus trichocarpa*. Phytochemistry, 72 (9): 897-908.

Degenhardt J., Köllner T. G. and Gershenzon J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry, 70: 1621-1637. https://doi.org/10.1016/j. phyto. Chem.2009.07.030.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J. F., Guindon S., Lefort V., Lescot M., Claverie J. M. and Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465-W469. doi.org/ 10.1093/nar/gkn180

Falara V., Akhtar T. A., Nguyen T. T. H., Spyropoulou E. A., Bleeker P. M., Schauvinhold I., Matsuba Y., Bonini M. E., Schilmiller A. L., Last R. L., Schuurink, R. C., and Pichersky E. (2011).The tomato terpene synthase gene family. Plant Physiol. 157:770–789. doi: 10.1104/pp.111.179648.

Tholl D., Boland W., Hansel A., Loreto F., Röse U. S. and J. P. Schnitzler (2006). Practical approaches to plant volatile analysis. Plant J. 45: 540-560.

Gershenzon J. and W. Kreish (1999). Biochemistry of terpenoids: monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins. In: Wink M., editor. Biochemistry of plant secondary metabolism. Florida: CRC Press, 222-299.

Gutensohn M., Orlova I., Nguyen T. T., Davidovich-Rikanati R., Ferruzzi M. G., Sitrit Y., Lewinsohn E., Pichersky E. and N. Dudareva (2013). Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J., 75 (3): 351-363.

Keilwagen J., Lehnert H., Berner T., Budahn H., Nothnagel T., Ulrich D., Dunemann F. (2017). The terpene synthase gene family of carrot (*Daucus carota* L.): Identification of QTLs and candidate genes associated with terpenoid volatile compounds. Front. Plant Sci. 8:1930. doi: 10.3389/fpls.2017.01930.

Köllner T. G., Held M., Lenk C., Hiltpold I., Turlings T. C., Gershenzon J. and Degenhardt J. (2008). A maize

(E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. Plant Cell, 20(2): 482-94. doi: 10.1105/tpc.107.051672. Epub, Feb 22. PMID: 18296628; PMCID: PMC2276456.

Korankye A. E., Lada R., Asiedu S. and Claude C. (2017). Plant senescence: the role of volatile terpenecompounds (VTCs). Am. J. Plant Sci., 8: 3120-3139.

Külheim C., Padovan A., Hefer C., Krause S. T., Köllner T. G., Myburg A. A., Degenhardt J. and W. J. Foley (2015). The Eucalyptus terpene synthase gene family. BMC genomics, 16 (1): 450. doi.org/10.1186/.

Li G., Köllner T. G., Yin Y., Jiang Y., Chen H., Xu Y., Gershenzon J., Pichersky E., Chen F. (2012). Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. Proc. Natl. Acad. Sci. USA. 2012;109:14711-14715. Doi: 10.1073/pnas.1204300109.

Martin D. M., Aubourg S., Schouwey M. B., Daviet L., Schalk M., Toub O., Lund S. T., Bohlmann J. (2010). Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol. 2010;10:226. doi: 10.1186/1471-2229-10-226.

Nieuwenhuizen N. J., Green S. A., Chen X., Bailleul E. J. D., Matich A. J., Wang M. Y., Atkinson R. G. (2013). Functional genomics reveals that a compact terpene synthase gene family can account for terpene volatile production in apple. Plant Physiol. 161:787-804. doi: 10.1104/pp.112.208249.

Luck K., Chen X., Norris A. M., Chen F., Gershenzon J. and Köllner T. G. (2020). The reconstruction and biochemical characterization of ancestral genes furnish insights into the evolution of terpene synthase function in the *Poaceae*. Plant. Mol. Biol. 104, 1-2: 203-215. doi:10.1007/s11103-020-01037-4

Pott D. M., Osorio S., and Vallarino J. G. (2019). From central to specialized metabolism: an overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. Front. Plant Sci. 10:835. doi: 10.3389/fpls.2019.00835

Ro D., Ehlting J., Keeling C., Lin R., Mattheus N. and J. Bohlmann (2006). Microarray expression profiling and functional characterization of *AtTPS* genes: Duplicated *Arabidopsis thaliana* sesquiterpene

synthase genes *At4g13280* and *At4g13300* encode root-specific and wound-inducible (Z)-Y-bisabolene synthases. Arch. Biochem. Biophys., 448, 1-2: 104-116. : doi.org/10.1016/j.abb.2005.09.019.

Ruzicka L. (1953). The isoprene rule and the biogenesis of terpenic compounds. Experientia, 9: 357-367.

Ruzicka L. (1959). Faraday Lecture (History of the isoprene rule), Proc. Chern. Soc. (Lond.): 341-360.

Ruzicka L. (1973). In the borderland between bioorganic chemistry and biochemistry, Annu. Rev. Biochem., 42: 1-20.

Su-Fang E., Zeti-Azura M., Roohaida O., Noor A. S., Ismanizan I. and Zamri Z. (2014). Functional Characterization of Sesquiterpene Synthase from *Polygonum minus*. Scientific World Journal. doi.org/10.1155/2014/840592.

Sunjung P. (2006). *Agrobacterium tumefaciens* –mediated transformation of tobacco (*Nicotiana tabacum* L.) leaf disks: evaluation of the co-cultivation conditions to increase β -Glucuronidase gene activity. (Master's dissertation). Retrieved from http://etd.lsu. edu /docs/available/etd-07052006-173930/unrestricted/Park_thesis.Pdf .

Takano A. and Okada H. (2011). Phylogenetic relationships among subgenera,species, and varieties of *Japanese Salvia* L. (*Lamiaceae*), J. Plant Res., 124: 245–52.

Trapp S. and Croteau R. (2001). Defensive resin biosynthesis in conifers. Ann. Rev. Plant Physiol., Plant Mol. Biol., 52: 689-724.

Volke D. C., Rohwer J., Fischer R. and Jennewein S. (2019). Investigation of the methylerythritol 4-phosphate pathway for microbial terpenoid production through metabolic control analysis. Microb Cell Fact. 18, 1,192. doi: 10.1186/s12934-019-1235-5

Wallach O. (1887). Zur KentniB der Terpene uDd dec atherischen Oele,Liebig's Ann. Chern., 239: 1-54.

Wang Q., Jia, M., Huh J. H., Muchlinski A., Peters R. J. and Tholl D. (2016). Identification of a dolabellane type diterpene synthase and other root-expressed diterpene synthases in *Arabidopsis*. Front Plant Sci., *7*: 1761. doi.org/10.3389/fpls.2016.01761

Wang Q., Quan S. and Xiao H. (2019). Towards efficient terpenoid biosynthesis: manipulating IPP and DMAPP supply. Bioresour. Bioprocess. 6**,** 6. doi.org/10.1186/s40643-019-0242-z

Zhou H. C., Shamala L. F., Yi X. K., Yan Z., Wei S. (2020). Analysis of terpene synthase family genes in Camellia sinensis with an emphasis on abiotic stress conditions. Sci. Rep. 10:933. doi.org/10.1038/s41598-020-57805-1

Table (1): BLASTX analysis *SgTPS4* was compared with the NCBI protein database for gene identification purposes.

NCBI Accession	[a]Descriptiona	Organism	*E* value	Identity (%)	Accession length
XP_041989963.1	Bicyclogermacrene synthase-like	*Salvia splendens*	0	97.07%	555
XP_041993267.1	Bicyclogermacrene synthase-like isoform	*Salvia splendens*	0	86.76%	557
XP_042006403.1	Bicyclogermacrene synthase-like	*Salvia splendens*	0	72.81%	558
XP_042005662.1	Bicyclogermacrene synthase-like	*Salvia splendens*	0	69.27%	559
XP_042008743.1	Bicyclogermacrene synthase-like	*Salvia splendens*	0	71.06%	559
XP_042006468.1	Bicyclogermacrene synthase-like	*Salvia splendens*	0	70.52%	557
XP_041993268.1	Bicyclogermacrene synthase-like isoform	*Salvia splendens*	0	84.89%	485

[a]Description—homology search using blastx.

Table (2): The major terpenoid compositions in transgenic *N. tabacum* leave over-expressing of *SgTPS4.*

N	Compound name	R.T (min.)	Formula	Molecular Mass (g mol-1)	Terpene of Type	% Peak area	
						NtW.T	*SgTPS4*
1	6-Amino-o-toluic acid; Benzoic acid, 2-amino-6-methyl-	6.882	C8H9NO2	151.1626		11.13	
2	Lupetidin	8.217	C7H15N	113.2007		0.61	
3	Piperidine, 2,6-dimethyl-;	14.27	C7H15N	113.2007		1.44	
4	4-Pipecoline	14.845	C6H13N	99.1741		1.38	
5	Dimethylsiloxane cyclic trimer	17.016	C6H18O3Si3	222.4618		4.44	
6	3,5-Dihydroxy-6-methyl-2,3-dihydro-4H-pyran-4-one	19.901	C6H8O4	144.1253		1.34	
7	Dimethylsiloxane cyclic trimer	26.286	C6H18O3Si3	222.4618		0.59	
8	**L-(-)-Nicotine**	27.823	C10H14N2	162.232		**50.96**	
9	α-Nicotine	28.319	C10H14N2	162.232		4.13	
10	Tetradecamethylcycloheptasiloxane	29.69	C14H42O7Si7	519.0776			0.02
11	trans-β-Ionone	30.644	C13H20O	192.2973			0.05
12	Topanol;Stavox	31.345	C15H24O	220.3505	Sesqui		0.07
13	Ethyl isopropylidene(cyano)acetate	32.112	C8H11NO2	153.1784		0.92	
14	Hexadecamethylcyclooctasiloxane	34.574	C16H48O8Si8	593.2315			0.06
15	Bisphenol C	34.748	C17H20O2	256.3395		0.61	
16	6-Aminouracil	35.302	C4H5N3O2	127.1014		0.55	
17	m-Cresyl N-methylcarbamate	36.029	C9H11NO2	165.1891			0.09
18	2(1H)-Pyrimidinone, tetrahydro-1,3-dimethyl-	37.487	C6H12N2O	128.1723		2.72	
19	Myristaldehyde	37.76	C14H28O	212.3715			0.14
20	(+)-Pyrethronyl (+)-trans-chrysanthemate;	38.579	C21H28O3	328.4452		1.28	
21	Octadeamethyl-cyclononasiloxane	38.761	C18H54O9Si9	667.3855			0.14
22	Methyl isohexadecanoate	39.636	C17H34O2	270.4507			0.39
23	2(1H)-Pyrimidinone, tetrahydro-1,3-dimethyl-	40.11	C6H12N2O	128.1723		3.37	
24	Hexadecane, 1,2-epoxy-; Hexadecylene oxide	40.777	C16H32O	240.4247			1.29
25	Cyclohexane, tert-pentyl-	42.053	C11H22	154.2924		0.53	

Table (2): Cont'

26	Alpha.-Linolenic acid, trimethylsilyl ester	42.458	C21H38O2Si	350.6107			0.8
27	1-(3-methylbutyryl)pyrrolidine	42.837	C9H17NO	155.237		0.76	
28	Palmitic acid, methyl ester	43.187	C17H34O2	270.4507			0.1
29	Linolenic acid, methyl ester	44.012	C19H32O2	292.4562			3.48
30	Bromocriptine	44.52	C32H40BrN5O5	654.594		1.84	
31	n-Hexadecanoic acid	44.686	C16H32O2	256.4241			
32	2,5-Piperazinedione, 3,6-bis(2-methylpropyl)-	45.151	C12H22N2O2	226.3153			11.74
33	Palmitic acid	45.848	C16H32O2	256.4241		2.94	
34	Hexadecamethylcyclooctasiloxane	45.865	C16H48O8Si8	593.2315		5.33	
35	4,8,13-Duvatriene-1,3-Diol	45.95	C20H34O2	306.4828			0.3
36	1,3-Distearin	46.238	C39H76O5	625.018			0.2
37	δ-Guaiene;	46.392	C15H24	204.3511	Sesqui		0.11
38	(+)-Ledol	46.691	C15H26O	222.3663	Sesqui		0.65
39	All-trans-Retinol acetate	46.926	C22H32O2	328.4883			0.63
40	Methyl cis,cis-9,12-octadecadienoate; Methyl linoleate	47.192	C19H34O2	294.4721			0.87
41	Linolenic acid, methyl ester	47.341	C19H32O2	292.4562			0.45
42	Phytol	47.603	C20H40O	296.531	Diter		0.74
43	4,8,13-Duvatriene-1,3-Diol	48.064	C20H34O2	306.4828			4.82
44	Cycloartanyl acetate	48.148	C32H54O2	470.77			1.22
45	Phytol, TMS derivative	48.303	C23H48OSi	368.7121	Diter	0.64	
46	Geranylgeraniol	48.418	C22H36O2	332.52			0.28
47	α-Linolenic acid;	48.907	C18H30O2	278.4296			23.73
48	Stearic acid	49.281	C18H36	284.477			2.86
49	cis-Bicyclogermacradiene	49.552	C15H24	204.3511	Sesqui		1.15
50	(Z)-9-Tetradecenal	50.013	C14H26O	210.36		2.49	
51	**Bicyclogermacrene**	50.461	C15H24	204.3511	Sesqui		**33.8**
52	d-Ledol	51.493	C15H26O	222.3663	Sesqui		0.07
53	Octadeamethyl-cyclononasiloxane	52.037	C18H54O9Si9	667.3855			0.65
54	6,9-Octadecadienoic acid, methyl ester	53.436	C19H34O2	294.4721			0.05

Table (2): Cont'

55	Squalene	54.046	C30H50	410.718	Triter		0.43
56	n-Heneicosane	54.964	C21H44	296.5741			0.05
57	Octadeamethyl-cyclononasiloxane	56.355	C18H54O9Si9	667.3855			0.79
58	Linolenic acid, methyl ester	57.393	C19H32O2	292.4562			0.09
59	n-Pentatriacontane	58.992	C35H72	492.9462			0.79
60	Phthalic acid dioctyl ester	60.497	C24H38O4	390.5561			0.15
61	Nopol	61.574	C11H18O	166.26			0.13
62	Octadeamethyl-cyclononasiloxane	62.682	C18H54O9Si9	667.3855			1.06
63	n-Tetracontane	64.117	C40H82	563.0791			0.14
64	Isovaleric acid, allyl ester	68.645	C8H14O2	142.1956			0.23
65	n-Tetracontane	69.7	C40H82	563.0791			1.61
66	Octadeamethyl-cyclononasiloxane	70.054	C18H54O9Si9	667.3855			1.15
67	Tetrapentacontane	71.42	C54H110	759.4512			0.03
68	n-Pentatriacotane	72.462	C35H72	492.9462			0.19
69	O-Benzyllinalool	73.038	C17H24O	244.37			0.15
70	3-Methyloctadecane	73.842	C19H40	268.5209			0.03
71	n-Nonacosane	75.547	C29H60	408.7867			0.38
72	Octadeamethyl-cyclononasiloxane	77.732	C18H54O9Si9	667.3855			1.17
73	n-Nonacosane	79.245	C29H60	408.7867			0.27
	Total % of sesquiterpene						35.85
	Total % of titerpene						0.43
	Total % of diterpene					0.64	0.74

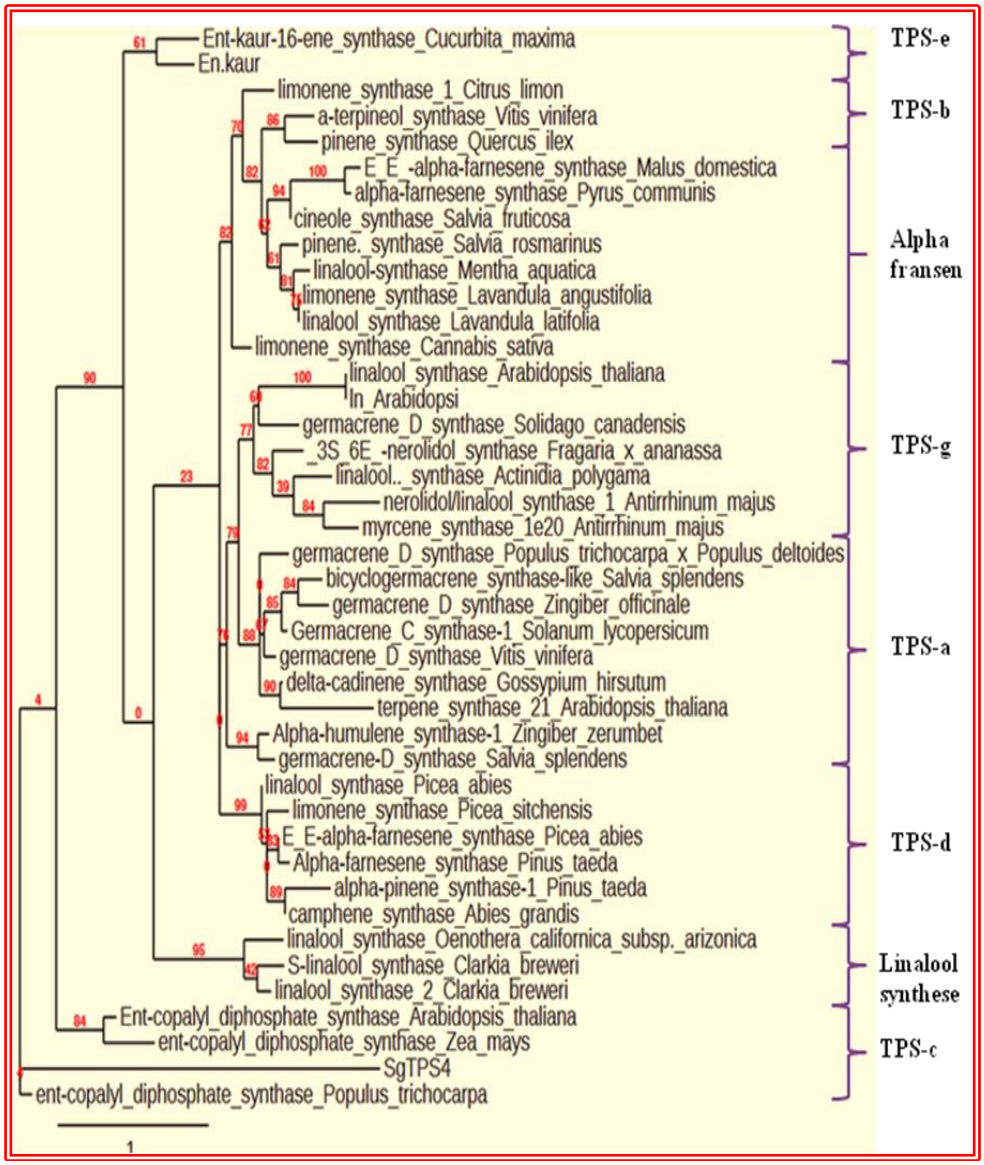

Fig. (1): Phylogenetic tree of *SgTPS4* with selected terpene synthases from other plants. Seven previously identified TPS subfamilies (Tps-a to Tps-g) were chosen based on Bohlmann *et al*., (1998) and Danner *et al*., (2011). The alignment was performed using the PhyML server. The numbers indicated are the actual bootstrap values of the branches.

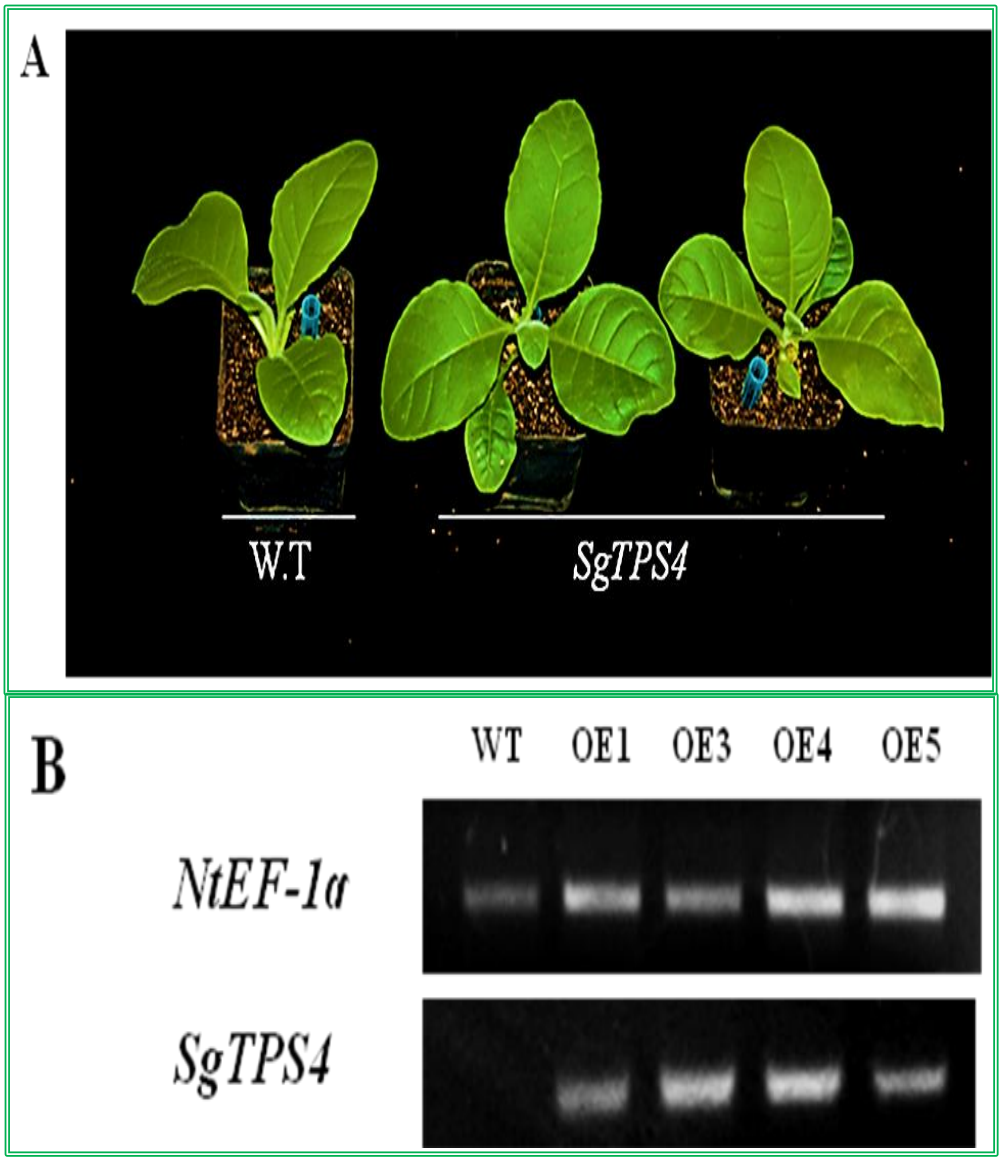

Fig. (2): Overexpression of *S. guaranitica* Terpene synthase 4 gene (*SgTPS4*) in transgenic tobacco. (A) Comparison of the phenotypes of the transgenic *N. tabacum* and wild type (W.T) *N. tabacum*. (B) Semiquantitative RT-PCR to confirm the expression of Terpene synthase 4 gene.

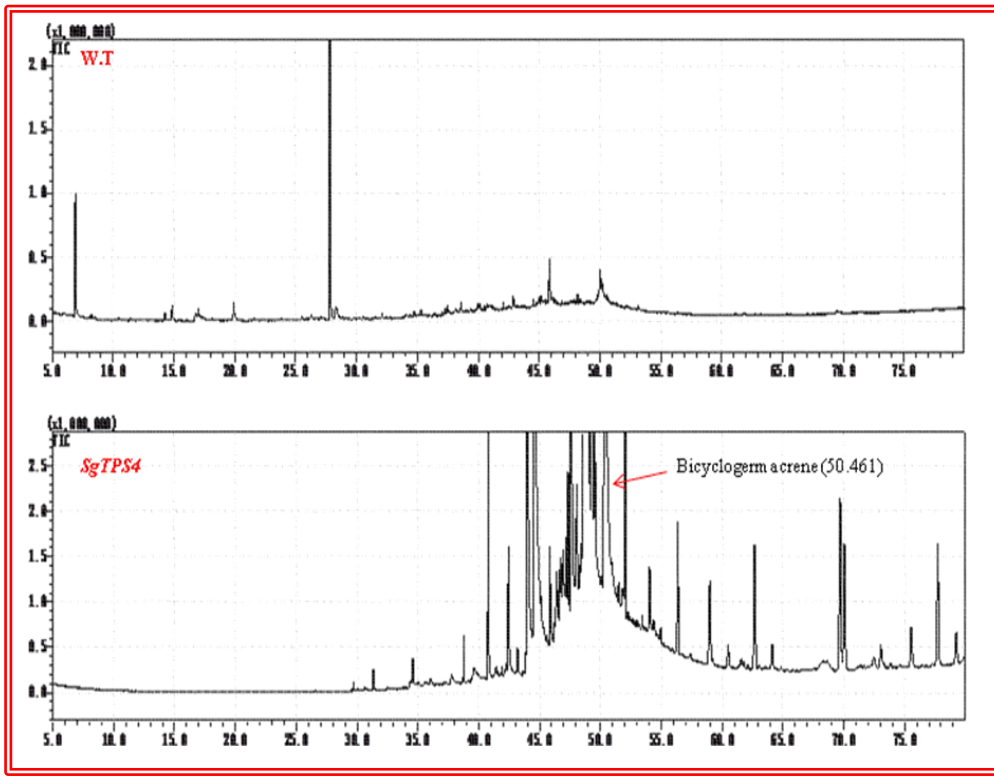

Fig. (3): Typical GC-MS mass spectrographs for terpenoids from leaf of *N. tabacum* plants.