Genome-wide detection of terpene synthase genes in holy basil (Ocimum sanctum L.)

Yogesh Kumar1, Feroz Khan1,*, Shubhra Rastogi2, Ajit Kumar Shasany2

1 Metabolic and Structural Biology Dept, CSIR-Central Institute of Medicinal and Aromatic Plants, Lucknow (U.P.), INDIA, 2 Biotechnology Division, CSIR-Central Institute of Medicinal and Aromatic Plants, Lucknow (U.P.), INDIA

* f.khan@cimap.res.in

Abstract

Holy basil (Ocimum sanctum L.) and sweet basil (Ocimum basilicum L.) are the most commonly grown basil species in India for essential oil production and biosynthesis of potentially volatile and non-volatile phytomolecules with commercial significance. The aroma, flavor, and pharmaceutical value of Ocimum species is a significance of its essential oil, which contains most of the monoterpenes and sesquiterpenes. A large number of plants have been studied for characterization and identification of terpene synthase genes, involved in terpenoids biosynthesis. The goal of this study is to discover and identify the putative functional terpene synthase genes in O. sanctum. HMMER search was performed by using a set of 13 well sequenced and annotated plant genomes including the newly sequenced genome of O. sanctum with Pfam-A database locally, using HMMER 3.0 hmmsearch for the two Pfam domains (PF01397 and PF03936). Using this search method 81 putative terpene synthases (OsaTPS) were identified in O. sanctum; the study further reveals 47 OsaTPS were putatively functional genes, 19 partial OsaTPS, and 15 OsaTPS as probably pseudogenes. All these identified OsaTPS genes were compared with other plant species, and phylogenetic analysis reveals the subfamily classification of OsaTPS in TPS-a, -b, -c, -e, -f and TPS-g subfamilies clusters. This genome-wide identification of OsaTPS genes, their phylogenetic analysis and secondary metabolite pathway mapping predictions together provide a comprehensive understanding of the TPS gene family in Ocimum sanctum and offer opportunities for the characterization and functional validation of numbers of terpene synthase genes.

Introduction

The terpenes are a large class of plant specialized secondary metabolites, which derived from 5 carbon (C5) isoprenoid unit; these precursors are produced by two biosynthetic pathways, the methylerythritol phosphate pathway (MEP) in the chloroplast and the classical mevalonate pathway (MVA). MVA is also known as the isoprenoid pathway or HMG-CoA reductase pathway present in the cytosol [1,2], the pathway produces five-carbon building blocks of terpenoids called isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) [3,4], they are used to make isoprenoids. IPP and DMAPP combined by prenyltransferases
supported by the grants from CSIR, New Delhi networking Project BSC0121 (GENESIS) to F.K. The CIMAP Communication No is CIMAP/PUB/2016/80. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** OsaTPS, (*Ocimum sanctum* terpene synthase); TPS, (terpene synthase).

and known as isoprenyl diphosphate synthases [5]. The condensation reaction by prenyltransferases produces geranyl diphosphate (GPP C_{10}), farnesyl diphosphate (FPP C_{15}) in the cytosol and geranylgeranyl pyrophosphate (GGPP C_{20}) in the chloroplast. These all prenyl pyrophosphates work as a substrate for the terpene synthases, which further performed the catalysis for the production of different terpenoids: hemiterpene (C_{5}) is produced directly by isopentenyl pyrophosphate, monoterpenes (C_{10}) from GPP, sesquiterpene (C_{15}) from FPP, diterpene (C_{20}) form GGPP, and other terpenes like sesterterpenes (C_{25}), triterpenes (C_{30}), sesquiterpenes (C_{35}), tetraterpenes (C_{40}) and polyterpene (> C_{40}) [6]. These terpenes are responsible for plant defense against biotic and abiotic stresses, used to attract insects for pollination, and used by human for their beneficial purposes as a natural flavor, for perfumery, and for cosmetics. These terpenoids are synthesized by specialized genes known as terpene synthases.

The terpene synthase genes are classified into different classes. According to their genomic structures (including intron and exon numbers), terpene synthase sequences (TPSs) can be classified into three classes Class I, Class II and Class III terpene synthases. The class I terpene synthase genes contain 12–14 introns, and 13–15 exons, class II genes possess nine introns and ten exons. Moreover, class III terpene synthase genes have six introns and seven exons [7–9].

Based on catalytic mechanism and product formed, TPSs genes are also classified into two classes: class I and class II. The class I and II TPSs have unique conserved amino acid motifs which are essential for catalysis [10,11]. Class I TPSs has a C-terminal domain (also referred to as α-domain or class I fold) which catalyzes the ionization of substrate (i.e., prenyl diphosphate) mediated by a divalent cation. This metal-dependent ionization of substrate can further lead to cyclizations, hydride shifts, and structural rearrangements to produce the end product. The α-domain present in this class TPSs adopt the α-helical protein fold and holds two metal binding motifs ‘DDXXD’ a highly conserved and less conserved ‘NSE/DTE’ positioned on opposing helices near the entrance of the active site. The class II terpene synthases comprise a functional N-terminal (β-domain), with a third “insertion” γ-domain forming a vestigial class II fold. The enzymes of this class contain conserved functional ‘DXDD’ motif which resides in a separate β-domain and accountable for the protonation-initiated cyclization of the substrate [11,12]. The γ-domain carries a highly acidic EDXXD like motif, which contributes to the activity of class II TPSs (Fig 1) [13]. Many terpene synthases also have a highly conserved RR (x)8W motif downstream of the N-terminal transit peptide while this motif does not require in monoterpen synthase activity.

According to recent subfamilies classification, TPS genes are classified into eight subfamilies: TPS-a to TPS-h based on sequence properties and functional characteristics. Class I TPSs family contains TPS-a, -b, -e/f, and TPS-g [10], where class II contains TPS-c only which comprises genes for copalyl diphosphate related diterpene synthases. The TPS-d family mostly consists of bifunctional enzymes that are capable of the protonation-initiated cyclization, but also catalyze class I catalysis and mostly occur in gymnosperms. Whereas TPS-h is specific to the spikemoss *Selaginella moellendorffii* [10].

The TPS subfamily clades have been studied and recognized through sequencing and functional studies in a wide range of plants *i.e.*, *Arabidopsis thaliana* containing (32 functional and 8 pseudo genes) [14], *Vitis vinifera* (69 putative functional and 63 pseudo genes) [15,16], *Solanum lycopersicum* (29 functional or potential functional genes and 15 mutated genes) [17], *Oryza sativa* (40 putative functional TPS genes) [18], *Selaginella moellendorffii* (18 TPS genes) [19], *Cajanus cajan* [20] and *Populus trichocarpa* (38 putative functional TPS genes) [21], *Eucalyptus grandis* (113 putative functional gene) *Eucalyptus globulus* (106 putative functional genes, 37 putative pseudogenes) [7]. Genome-wide analysis of terpene synthase genes family in different plants shows that all the TPS gene family is the mid-size with the number ranging
from approximately 20 to 150 in sequenced plant genomes, but with only one exception of the moss *Physcomitrella patens*, which possess single functional TPS genes shown in Table 1.

*Ocimum sanctum* (Holy basil) due to its medicinal properties has great importance in Indian traditional medicine and Ayurveda. It is a sweet-scented, pubescent herb nearly 3 to 100cm in height, growing abundantly in tropical, sub-tropical and warm temperate regions. *Ocimum sanctum* plays a significant role in herbal as well as modern medicines. Many studies have focused on the medicinal prospect of this plant in the form of crude drugs, essential oil or pure compounds. *O. sanctum* shows antidiabetic, antimicrobial, antioxidant, anti-inflammatory, antinociceptive, antifertility, anticancer, anthelmintic and cardioprotective properties.
Several Ocimum species produce essential oil with methyl eugenol being a major constituent and reported to have anti-cancer properties [24,25]. Additionally, Ocimum terpenoids also used as biodegradable herbicide due to their phytotoxic activity [26]. The Lamiaceae family is the sixth most extensive family of the flowering plants and plays a significant role in the economy of natural products and used in foods, pharmaceuticals, flavor and perfumery industries. This family shows a remarkably high degree of secondary metabolite diversity specially terpenoids [27]. Because of their importance a number of TPSs genes which were responsible for terpenoids biosynthesis have been explored and characterized in many species of Lamiaceae family, these species were closely related to Ocimum sanctum, including Rosmarinus officinalis, Mentha piperita [28–30], Mentha cablin [31,32], Mentha aquatic [33], Salvia miltiorrhiza [34–36], Isodon eriocalyx [37,38], Coleus forskohlii [39], Salvia fruticosa [40], Salvia pomifera [41], Salvia sclarea [42,43] and Salvia divinorum [44]. Although in prior studies at least 44 monoterpenes and 39 sesquiterpenes were reported in literature reports [27,45].

The volatile constituents of various Ocimum sanctum tissues such as leaves, flowers, and seeds have been previously examined, and constitute mostly monoterpenes, sesquiterpenes and diterpenes [46], but least number of Terpene synthases genes were characterized in Ocimum species. Some characterized TPSs genes of Ocimum are (Geraniol synthase [47], R-linalool synthase, (-)-Endo-fenchol synthase, Selinene synthase, Terpinolene synthase, β-myrcene synthase, α-Zingiberene synthase, γ-Cadinene synthase, and Germacrene-D synthase, all monoterpenes and sesquiterpene synthases) [48,49] and one triterpene oxidosqualene synthases [50]. Due to unavailability of whole genome sequencing data, an actual number of TPS genes in Ocimum sanctum is still unknown. In this study, we have identified 47 putative functional terpene synthase genes in O. sanctum and classified the TPSs into subfamily. The comparative sequence study with closely related species reveals and suggest the possible

| Plant species | Putative TPS genes identified by updated TPS HMM model\(^1\) | Putative full-length TPS genes\(^2\) | TPS genes included in OsaTPS phylogenetic tree | TPS Subfamily\(^3\) |
|---------------|--------------------------------------------------|---------------------------------|------------------------------------------|------------------|
|               |                                                  |                                 |                                          | a | b | c | e/f | g | h |
| A. thaliana   | 40                                               | 32\(^5\)                        | 33                                       | 23 | 6 | 1 | 2 | 1 | 0 |
| V. vinifera   | 91                                               | 69\(^5\)                        | 55                                       | 29 | 9 | 2 | 1 | 14 | 0 |
| O. sativa     | 32                                               | 32\(^5\)                        | 32                                       | 19 | 0 | 3 | 9 | 1 | 0 |
| P. patens     | 2                                                 | 2\(^5\)                         | 0                                        | 0  | 0 | 2 | 0 | 0 | 0 |
| E. grandis    | 120                                              | 113\(^5\)                      | 113                                      | 52 | 36| 2 | 10| 13| 0 |
| E. globulus   | 106                                              | 106\(^5\)                      | 69                                       | 32 | 30| 2 | 3 | 0 | 0 |
| S. lycopersicum | 41                                             | 29\(^5\)                        | 29                                       | 12 | 8 | 2 | 5 | 2 | 0 |
| S. bicolor    | 41                                               | 24\(^5\)                        | 24                                       | 15 | 2 | 1 | 3 | 0 | 0 |
| P. trichocarpa | 77                                              | 32\(^5\)                        | 31                                       | 13 | 11| 2 | 3 | 2 | 0 |
| N. tabaccum   | 96                                               | 14\(^5\)                        | 2                                        | 0  | 0 | 1 | 1 | 0 | 0 |
| C. cajan      | 52                                               | -                               | 0                                        | 29 | 9 | 2 | 10| 2 | 0 |
| S. moellendorffii | 40                                           | 14\(^5\)                        | 0                                        | 0  | 0 | 3 | 0 | 8 | 0 |
| O. basilicum  | -                                                | 2\(^5\)                         | 2                                        | 0  | 0 | 0 | 0 | 2 | 0 |
| O. sanctum    | 47                                               | -                               | 47                                       | 18 | 16| 5 | 3 | 5 | 0 |

\(^1\)Total number of terpene synthase genes identified with the TPS HMM sequence models.
\(^2\)Putative full-length genes retrieved from available data.
\(^3\)Putative TPS genes classified into subfamily by phylogenetic analysis.
\(^5\)Numbers of TPS genes retrieved from the information based on literature data.
\(^6\)TPS sequence data retrieved from the non-redundant protein sequence database from NCBI

https://doi.org/10.1371/journal.pone.0207097.t001

Table 1. Terpene synthase genes size in different plant species, including O. sanctum genome.
biological role of these predicted *O. sanctum* terpene synthases. These results may provide insight for further characterizing the putatively functional terpene synthase genes in the Holy basil.

**Materials and methods**

**Data retrieval and identification of TPSs in *O. sanctum***

The 13 plant's genome and proteome data were downloaded from (https://www.ncbi.nlm.nih.gov/genome/) and (https://www.ncbi.nlm.nih.gov/assembly/). The genome data of *Ocimum sanctum* was taken from the in-house sequenced genome repository. Two terpene synthase specific domains Pfam N-terminal (PF01397) (https://pfam.xfam.org/family/PF01397#tabview=tab6) and Pfam C-terminal domains (PF03936) (https://pfam.xfam.org/family/PF03936#tabview=tab6) were downloaded from the Pfam database [51]. Standalone tool, HMMER version 3.1b2 was downloaded [52] and used to search the *Ocimum sanctum* predicted proteome data including 13 downloaded proteomes using the PF03936 and PF01397 domains model data as a query for proteome; the significant (e-value $< 10^{-3}$) was set for identification of candidate TPS genes. The candidate genes were also inspected with a ScanProsite tool (https://prosite.expasy.org/scanprosite/) for terpene synthase motifs identification. FGENESH gene prediction tool (www.softberry.com) was also used for initial annotation of the predicted *Ocimum sanctum* terpene synthase (*OsaTPS*) genes. Predicted *OsaTPS* were then monitored for the full-length gene, to make them full-length the upstream and downstream areas of the conserved scaffold region were screened, and reverse BLASTx search was used to confirm the identity of the putative functional TPS gene. Out of 90 identified *OsaTPS* genes, 9 were excluded due to short sequence or stop codons in the translated gene, among the 81 selected *OsaTPS* genes, only 47 were taken for further analysis, rest 19 partial *OsaTPS* and 15 *OsaTPS* probably pseudogenes were retained separately for study shown in S1 and S2 Tables. Some of the genes, which do not cover the full-length gene sequence were also included in this study. The Gene structure intron-exon and motif representation of putative *OsaTPS* genes were determined by the Gene structure display server 2.0 version [53].

**Phylogenetic analysis of terpene synthase genes from *O. sanctum* and other plants**

The terpene synthase genes from different plants were retrieved, i.e., *Arabidopsis thaliana* (http://arabidopsis.org/), *Oryza sativa* (http://www.plantdb.org/OsGDB/) [54], *Vitis vinifera* (http://www.plantdb.org/VvGDB), *Solanum lycopersicum* Potato Genome Sequencing Consortium [55], *Populus trichocarpa* (http://www.plantdb.org/PtGDB) [56], *Sorghum bicolor* (http://www.plantdb.org/SbGDB/) [57], *Ocimum basilicum* (http://www.uniprot.org), *Nicotiana tabacum* (http://www.uniprot.org), *Eucalyptus grandis* [58] and *Eucalyptus globulus* [59] as shown in Table 1 and TPS sequences shown in S1 File, and further classified into TPS subfamily as depicted in S3 Table. Predicted 47 *OsaTPS* genes were included for protein alignment, which was prepared using MUSCLE [60] integrated into MEGA version 6 [61] with standard parameters. The multiple sequence alignment was then manually adjusted to concentrate on various diagnostic conserved motifs of terpene synthases, i.e., RDR, RR(X)$_{4}$W, DDXXD, DXDD and NSE/ DTE. The neighbor-joining method with 1,000 number of bootstrap replications was used for phylogenetic tree reconstruction; substitution model was checked for best phylogenetic tree construction, the p-distance model was considered with pairwise deletion gaps/missing data treatment. The phylogenetic tree was visualized by Figtree 1.4.2 [62].
**OsaTPS sequence annotation and secondary metabolite pathway genes prediction**

The Blast search of *OsaTPS* genes was performed against UniProt database, and GO (Gene Ontology) terms were assigned for each unigene based on the GO term annotated to its corresponding homolog in the UniProt database. The protein sequences of *Arabidopsis thaliana*, *Solanum lycopersicum*, *Ocimum basilicum*, *Salvia officinalis*, *Ricinus communis*, *Lavandula angustifolia*, *Pisum sativum*, *Nicotiana tabacum*, *Mentha x piperita*, *Antirrhinum majus*, *Fragaria x ananassa*, *Phyla dulcis*, *Solanum tuberosum*, *Pogostemon cablin*, *Origanum vulgare*, *Hyoscymus muticus*, and *Populus tremuloides* were used for sequence alignment as shown in S4 Table. The predicted *OsaTPS* genes were mapped for annotation and prediction of putative functional genes involved in the biosynthesis of secondary metabolites against KEGG metabolic pathways using BlastKOALA (KEGG Orthology And Links Annotation), which is freely available at (http://www.kegg.jp/blastkoala/) [51].

**Results**

**In-silico identification of putative terpene synthase genes in *O. sanctum***

In the study 81, putative TPS genes in *Ocimum sanctum* genome were identified using a high sequence similarity screening search for HMM PF01397 and PF03936 TPS domain models including 13 other plant species genomes. Out of 81 putative *OsaTPS* genes, 19 *OsaTPS* were predicted as partial genes and 15 probably pseudogenes. The *OsaTPS* genes which are partial and having multiple frameshifts or stop codons were not considered in this study. The 34 partial and pseudogenes separately analyzed for terpene synthase subfamily classification, where out of these, 15 partial genes falls in TPS-a, and 3 in TPS-b subfamily group and 1 partial gene falls in TPS-e subfamily, rest 8 pseudogenes genes fall in TPS-a subfamily, 3 in TPS-b, 2 in TPS-c, 1 in TPS-e and 1 gene in TPS-f subfamily group. All the partial and pseudogenes were retained for further study as shown in S2 Table. Maximum pseudogenes were found among the TPS-a subfamily, after all, manual curation and sequence filtration, only 47 putative functional *OsaTPS* genes were considered for further analysis Table 2.

**Intron-exon structure and organization of *OsaTPS* genes**

In this study, two forms of TPS gene classification was represented, one according to the presence of number of introns and exons in the TPS genes and another according to the presence of conserved motif. The intron-exon classification shows three classes, class I, class II and class III terpene synthases [59,63]. The *OsaTPS* genes of the subfamilies TPS-a, TPS-b, and TPS-g, contain 5 to 7 introns and 6 to 8 exons; most of the *OsaTPS* of TPS-a subfamily contains six, introns and seven exons in their gene structure and expected to fall in class III TPSs, except *OsaTPS*4 which contains only three introns and four exons and supposed to be an incomplete sequence. *OsaTPS* genes are containing seven introns and eight exons likely to have disrupted 3’exon by an extensive or probably insertion of the new intron. *OsaTPS* of TPS-b subfamily genes contain 5, 6 and 7 introns with 6, 7 and 8 exons, except *OsaTPS*21 which contains only four exons due to incomplete sequence, while *OsaTPS* of TPS-g subfamily genes contain 5, 6 and 7 introns and 6, 7 and 8 exons and fall in class III TPSs. Genes of *O. sanctum* from the remaining subfamilies; TPS-c, TPS-e, and TPS-f contain introns between 9 to 14 and exons from 10 to 15 which fall with in class I TPS gene family. Here TPS-d is not observed in the study, as it is specific to gymnosperm which falls in class II TPS gene family (Fig 2). According to the conserved motif based TPSs classification, in *O. sanctum* an 18 *OsaTPS* falls in TPS-a subfamily, 16 *OsaTPS* falls in TPS-b, two *OsaTPS* in TPS-e, one *OsaTPS* in TPS-
| S. No | Gene code | TPS subfamily | Protein length | Gene ID (gi) | Organism          |
|-------|-----------|---------------|----------------|-------------|-------------------|
| 1     | OsaTPS1   | TPSa          | 507            | 75251482    | Ocimum basilicum  |
| 2     | OsaTPS2   | TPSa          | 535            | 75251483    | Ocimum basilicum  |
| 3     | OsaTPS3   | TPSa          | 320            | 939319666   | Solanum lycopersicum |
| 4     | OsaTPS4   | TPSa          | 577            | 75267592    | Solanum tuberosum |
| 5     | OsaTPS5   | TPSa          | 566            | 75251484    | Ocimum basilicum  |
| 6     | OsaTPS6   | TPSa          | 383            | 403399735   | Origanum vulgare  |
| 7     | OsaTPS7   | TPSa          | 545            | 122219293   | Pogostemon cablin |
| 8     | OsaTPS8   | TPSa          | 550            | 122219293   | Pogostemon cablin |
| 9     | OsaTPS9   | TPSa          | 543            | 122219293   | Pogostemon cablin |
| 10    | OsaTPS10  | TPSa          | 540            | 476007192   | Phyla dulcis      |
| 11    | OsaTPS11  | TPSa          | 521            | 75251482    | Ocimum basilicum  |
| 12    | OsaTPS12  | TPSa          | 551            | 75251484    | Ocimum basilicum  |
| 13    | OsaTPS13  | TPSa          | 468            | 476007821   | Hyoscyamus muticus |
| 14    | OsaTPS14  | TPSa          | 383            | 75252096    | Mentha x piperita |
| 15    | OsaTPS15  | TPSa          | 274            | 122219293   | Pogostemon cablin |
| 16    | OsaTPS16  | TPSa          | 379            | 75252096    | Mentha x piperita |
| 17    | OsaTPS17  | TPSa          | 507            | 75251484    | Ocimum basilicum  |
| 18    | OsaTPS18  | TPSa          | 388            | 75251482    | Ocimum basilicum  |
| 19    | OsaTPS19  | TPSb          | 533            | 75232269    | Populus tremuloides |
| 20    | OsaTPS20  | TPSb          | 530            | 487524480   | Ricinus communis  |
| 21    | OsaTPS21  | TPSb          | 316            | 487524480   | Ricinus communis  |
| 22    | OsaTPS22  | TPSb          | 546            | 476007226   | Phyla dulcis      |
| 23    | OsaTPS23  | TPSb          | 512            | 75251481    | Ocimum basilicum  |
| 24    | OsaTPS24  | TPSb          | 558            | 122210944   | Lavandula angustifolia |
| 25    | OsaTPS25  | TPSb          | 356            | 62900763    | Salvia officinalis |
| 26    | OsaTPS26  | TPSb          | 524            | 122210944   | Lavandula angustifolia |
| 27    | OsaTPS27  | TPSb          | 589            | 62900763    | Salvia officinalis |
| 28    | OsaTPS28  | TPSb          | 575            | 62900763    | Salvia officinalis |
| 29    | OsaTPS29  | TPSb          | 622            | 75251478    | Ocimum basilicum  |
| 30    | OsaTPS30  | TPSb          | 650            | 75251478    | Ocimum basilicum  |
| 31    | OsaTPS31  | TPSb          | 350            | 75251477    | Ocimum basilicum  |
| 32    | OsaTPS32  | TPSb          | 512            | 75251477    | Ocimum basilicum  |
| 33    | OsaTPS33  | TPSb          | 601            | 75251479    | Ocimum basilicum  |
| 34    | OsaTPS34  | TPSb          | 556            | 75251479    | Ocimum basilicum  |
| 35    | OsaTPS35  | TPSc          | 826            | 62900382    | Pisum sativum     |
| 36    | OsaTPS36  | TPSc          | 783            | 62900382    | Pisum sativum     |
| 37    | OsaTPS37  | TPSc          | 824            | 544602148   | Nicotiana tabacum |
| 38    | OsaTPS38  | TPSc          | 826            | 544602148   | Nicotiana tabacum |
| 39    | OsaTPS39  | TPSc          | 821            | 544602148   | Nicotiana tabacum |
| 40    | OsaTPS40  | TPsf          | 812            | 62900385    | Cucurbita maxima  |
| 41    | OsaTPS41  | TPsf          | 642            | 544602165   | Nicotiana tabacum |
| 42    | OsaTPS42  | TPSf          | 829            | 75163668    | Arabidopsis thaliana |
| 43    | OsaTPS43  | TPsf          | 426            | 75242446    | Antirrhinum majus  |
| 44    | OsaTPS44  | TPsf          | 529            | 332319699   | Fragaria x ananassa |
| 45    | OsaTPS45  | TPsf          | 560            | 75224312    | Ocimum basilicum  |
| 46    | OsaTPS46  | TPsf          | 467            | 75224312    | Ocimum basilicum  |
| 47    | OsaTPS47  | TPsf          | 395            | 75251480    | Ocimum basilicum  |

Table 2. Sequence annotation data of predicted genes from Ocimum sanctum.

https://doi.org/10.1371/journal.pone.0207097.t002
Genome wide detection of *O. sanctum* terpene synthase genes
Fig 2. Representation of the gene structure of putative functional TPS genes from Ocimum sanctum. Genes were predicted by FGENESH tool and Augustus but also determined manually by mapping to previously TPS characterized genes. The black line indicates the introns; yellow rectangles indicate exons, red lines show conserved DDXXD motif and the green line shows DxDD motif. Arrow indicated in the figure shows the incomplete gene sequence or loss of sequence. Color boxes indicate the TPS subfamily TPS-a (from OsaTPS1 to OsaTPS19), TPS-b from (OsaTPS19 to OsaTPS34), TPS-c from (OsaTPS35 to OsaTPS39), TPS-e (OsaTPS40) and OsaTPS41, TPS-f (OsaTPS42) and TPS-g from (OsaTPS43 to OsaTPS47).

f, and five OsaTPS fall in TPS-g subfamily, these all subfamilies of terpene synthases are of class I TPSs. While five OsaTPS falls with in TPS-c subfamily which is of class II terpene synthases.

Phylogenetic analysis of terpene synthase gene in Ocimum sanctum

The phylogenetic analysis of putative functional 47 OsaTPS genes was performed to know the evolutionary relationship between the O. sanctum and other plant species terpene synthase sequences from, Arabidopsis thaliana, Oryza sativa, Vitis vinifera, Solanum lycopersicum, Populus trichocarpa, Sorghum bicolor, Ocimum basilicum, Nicotiana tabaccum, Eucalyptus grandis and Eucalyptus globulus shown in S1 Fig and S1 File. The previous phylogenetic tree of TPS protein sequences from gymnosperms and angiosperms was divided into seven major subfamily clades designated as TPS-a through TPS-g [9,64,65] (Fig 3). The phylogenetic analysis of OsaTPS gene formed only six clades as; (i) TPS-a (Fig 4) (ii) TPS-b and -g (Fig 5) (iii) TPS-c, -e and -f (Fig 6). TPS-d is missing from this analysis as it is gymnosperm specific and TPS-h is specific to the spikemoss Selaginella moellendorffii [10].

In TPS-a subfamily clade, O. sanctum TPS genes formed orthologous pairs with closely related species Sorghum bicolor and Solanum lycopersicum and speciation clusters with distinctly related species as depicted in (Fig 4 and S2 Fig). In the tree, it was observed that Arabidopsis thaliana and Sorghum bicolor have a longer branch length than Ocimum sanctum and Solanum lycopersicum, which suggest a prolonged period of gene differentiation without gene duplication events. TPS-a group consists of all the known angiosperm sesquiterpene and diterpene synthases, along with predicted OsaTPS genes, it is, therefore, most likely that 18 OsaTPS clustered with this group are also members of sesquiterpene and diterpene synthases.

TPS-b and -g phylogenetic tree clusters, shows 16 and 5 members of OsaTPSs, respectively, these clades consist of known angiosperm monoterpane or isoprenes synthase genes (Fig 5 and S3 Fig). The subfamily TPS-b and TPS-g were shown in a combined state with the representation of separate clusters; it was observed that all 5 OsaTPS-g subfamily genes formed an orthologous pair with S. lycopersicum and O. basilicum genes. TPS-b further divided into two subclades; TPS-b1 and TPS-b2, only Solly TPS3, 4, 5 and 7 lies in the TPS-b1 subclade, these TPS belongs to monoterpane synthase, while TPS-b2 sub-cluster contains multiple species, in which OsaTPS24, 25, 26, 27, 28, 29, 30, 31, 32, 33, and OsaTPS34 formed a separate cluster with longer branch length, where OsaTPS20, 21, 22, and OsaTPS23 form a cluster with S. lycopersicum genes.

The TPS-c, -e, and -f of O. sanctum shared five members of genes with TPS-c, two with TPS-e and one with TPS-f clades. These all subfamilies were shown in a single phylogenetic tree (Fig 6) and represented with the labeling of different TPS subfamily clades as shown in S4 Fig. All the OsaTPS were indicated in the red color branch, OsaTPS42 lies in the subfamily of TPS-f with A. thaliana, Poplar and Eucalyptus species, where TPS-e consist of OsaTPS40 and OsaTPS41 sharing the cluster with orthologous species S. lycopersicum and N. tabacum. TPS-c cluster shared five members of OsaTPS genes, and formed an orthologous pair with multiple species, OsaTPS37 and OsaTPS38 formed a pair with S. lycopersicum and N. tabacum, whereas OsaTPS35 and OsaTPS35 paired with multiple species.
Genome wide detection of *O. sanctum* terpene synthase genes
Fig 3. Discovered 47 putative functional TPS genes from *O. sanctum*: Neighbor-joining phylogenetic tree was constructed using 1,000 bootstrap value and classified into terpene synthase gene subfamily as TPS-a, -b, -c, -e, -f and TPS-g. Bootstrap values assigned in the tree, which was higher than 80%, values below 80% were not shown in the figure.

https://doi.org/10.1371/journal.pone.0207097.g003

Fig 4. Representation of phylogenetic tree of *O. sanctum* compared to other species of TPS-a subfamily. 18 OsaTPS represented in the tree were compared with the neighboring species. The Neighbor-joining method was used with 1,000 replicates for bootstrap values. The values = 80% and = 95% were designated in S2 Fig. OsaTPS33 and Pt0092000200, two TPS gene were used as an out-group.

https://doi.org/10.1371/journal.pone.0207097.g004
Conserved sequence motif analysis of *O. sanctum* TPSs

All the discovered OsaTPS were investigated for the presence of terpene synthase conserved motifs, i.e., RR(X)_8W, RXR, DXDD, DDXXD, and NSE/DTE; except these motifs, 20 more motifs has been explored. Monoterpane and diterpene synthases typically contain an N-terminal plastidial targeting peptide upstream of the conserved or modified RR(x)_8W motif, this conserved motif was predicted in 31 OsaTPS genes in exact or in modified form, which
resemble sesquiterpene, diterpene and monoterpene synthases. In TPS-b clade RR(x)_W motif was not found to be conserved in all OsaTPS genes (Fig 7). The most conserved DDXXD motif of the class I TPS gene family was observed in 39 members of OsaTPS, also represented in sequence motif logo (Fig 7). The DDXXD motif is present on the N-terminal region of positional 35 amino acid downstream to RXR/RDR motif; it plays a role in the complexation of the diphosphate group after ionization of substrate. At the c-terminal, a fully conserved RDR motif was found in most of the OsaTPS sequences, with some variations: R(D/H)R, some of

Fig 6. TPS-c, -e, and TPS-f, combined phylogenetic tree were represented from O. sanctum putatively functional TPS genes compared with other species. The phylogenetic tree was constructed with 1,000 replicates for bootstrap values using the neighbor-joining method. Here in TPS-c clade five OsaTPS were shown and compared, TPS-e and TPS-f show two OsaTPS and single OsaTPS42. Whereas OsaTPS1 was shown as outgroup.

https://doi.org/10.1371/journal.pone.0207097.g006
| OsaTPS gene name | Subfamily | RR (X) 0W | NSE/DTE | RXR | DDXXD/DXDD |
|----------------|-----------|-----------|---------|-----|------------|
| OsaTPS1       | a         | PPVSFVPCIN | RYIDDTGSHKRE | LFIRDV | GIMDTRDYNYA |
| OsaTPS2       | a         | RPSVTYNPNV | RLNDIAGYGSE | FARBOL | SVIDDITVYDG |
| OsaTPS3       | a         | PRSSNSFSPIN | ---------- | YAROIR | SVIDDTYSGH |
| OsaTPS4       | a         | RPSNFSPSIN | RVINDVATYE | YARDYM | SVIDDTYSGH |
| OsaTPS5       | a         | RPSKSHTSVPN | RLNIDDLAGHGKE | FARBOL | TIDDDITYVG |
| OsaTPS6       | a         | RRSKKFHSIN | RLNIDDLVGL-- | ---- | DORDYAVSL-- |
| OsaTPS7       | a         | RPVSNFSPSIN | RVIDDIATYEVE | YARDVR | SIVDTDDYAG |
| OsaTPS8       | a         | RPVSNFSPSIN | RVIDDIATYEVE | YARDVR | SIVDTDDYAG |
| OsaTPS9       | a         | RPVSNFSPSIN | RVIDDIATYEVE | YARDVR | SIVDTDDYAG |
| OsaTPS10      | a         | RPLANYHPMN | RLNDTQLYQHEFE | YARDYM | SIMDTRDYAYA |
| OsaTPS11      | a         | PPITIHPASIN | RYRDIQASHKRE | YIRDOL | GIMDTRDYNYA |
| OsaTPS12      | a         | RKSNRTYHPSP | RLNDLDTAYKKE | FPBDOR | TIVDDBYVYG |
| OsaTPS13      | a         | PPTIVAYKPSIN | --MEDLGSHERE | YARDRI | SIMDTRDYNYA |
| OsaTPS14      | a         | PPSIFKPSIN | ---------- | YARHRI | SIMDTRDYNYA |
| OsaTPS15      | a         | ---------- | RVIDDIATYEVE | ---- | SIMDTRDYAG |
| OsaTPS16      | a         | PPMISFKPSIN | ---------- | YARHRI | TIMDTRDYNYA |
| OsaTPS17      | a         | ---------- | RLNDLDTAGHK | FARBOL | TVIDDBYVYG |
| OsaTPS18      | a         | PPISTHAPSIN | --FRDKLHLRRE | YIRDOL | GIMDTRDYNYA |
| OsaTPS19      | b         | ---------- | RLNDLQLYE | FARBOR | TVIDDDYVYG |
| OsaTPS20      | b         | RBTANYPKIN | RLNDNGQTSGAE | FTRORR | LIDDDITYBG |
| OsaTPS21      | b         | RSANYPKPIN | ---------- | FSBORR | VITDDITYBG |
| OsaTPS22      | b         | RRSANYKPSIN | RLADDLGSTEE | FARBOL | TIDDDITYVG |
| OsaTPS23      | b         | RRSNVGASIN | RLADDLGTS-- | FARBOL | TIDDDITYVG |
| OsaTPS24      | b         | RRSNYKPTLN | RLADDLGTPHE | FARBOR | TVIDDDYVYG |
| OsaTPS25      | b         | RRSNYKPSHN | --MEDIQRTSPH | FVRBIR | --E |
| OsaTPS26      | b         | RRSANPKPPF | RLADLETSGNE | FARBOR | TIDDDITYVAG |
| OsaTPS27      | b         | RRSGGYHPTLN | RLADDMTGSLDE | FVRBIR | SADDDITYVAG |
| OsaTPS28      | b         | RRSGGYPSALN | RLADDMTGSLDE | FARBOL | TIDDDITYE |
| OsaTPS29      | b         | R-GYONPSAN | RLPDGLDTAPLE | FARBOR | TVIDDDYVYG |
| OsaTPS30      | b         | R-GYONPSAN | RLPDGLDTAPLE | FARBOR | TVIDDDYVYG |
| OsaTPS31      | b         | RRSGNYPESAN | RLPDGLTGASHE | -- | HLDGTFFELL |
| OsaTPS32      | b         | R-GYONPSAN | RLPDGTANIT | FARBOR | TVIDDDYVYG |
| OsaTPS33      | b         | R-GYONPSAN | RLADDLGTAPFE | FVRBIR | TIDDDITYVAG |
| OsaTPS34      | b         | R-VGNYPESTW | RLADDLGTOFE | FVRBIR | TVIDDDITYVAG |
| OsaTPS35      | c         | SLEG-MGALT | DLRMFRORKE-- | LSERNL | QVDDIDTAMG |
| OsaTPS36      | c         | SLEG-MQGALT | HGLRFLKRRKVR | LSRLS | NVDDIDTAMG |
| OsaTPS37      | c         | SLEG-LNLDLW | QGSLQHNEKE-- | IRSREL | EFOCIDDSNG |
| OsaTPS38      | c         | SLEG-LDLDLW | HLSQIPQOQ-- | IRSKLD | NFDLDTSNG |
| OsaTPS39      | c         | ILEGLNENIKW | N--GLNQOKM-- | ISKNEI | TFCQBDTSMA |
| OsaTPS40      | e         | IAEONGKLONG | RLLNBIRSYERE | FARSKS | LVDDIDFTYG |
| OsaTPS41      | e         | QORRIFPSVE | RLLNBVOTEFEKE | YGBQV | LVDDIDFTYG |
| OsaTPS42      | f         | YLETIPSTYEF | RLANDINSQYKE | FGBRTK | YVIDDDFTY |
| OsaTPS43      | g         | ---------- | RLDDGLTAQ-- | GARSOP | GIMDTRDYNYA |
| OsaTPS44      | g         | ---------- | RLDLDLSAQE | SRSOP | SVIYDDYVYG |
| OsaTPS45      | g         | INRNDMLKNT | RLADDGTAKMD | FGDRBP | SVIDDTYSG |
| OsaTPS46      | g         | FND--- | RLDDGLTAK-- | FGDRBP | SVIDDTYSG |
| OsaTPS47      | g         | ---------- | RLMMGLGTSQEE | FGBRDA | TIDDDITYVAG |
the sequences have variation of this motif as R(D/H)(R/K/D/V/Q) and tree sequences were shown missing this motif due to sequencing error. The additional metal cofactor binding motif NSE/DTE, is less conserved in TPS, evolved from a second aspartate-rich region in prenyl transferases to form a consensus sequence of (L,V,Y)(I,A,M,P,W,S,C)(N,D,G)D(L,I,V,M,K,Q)x(T,G,S)x(E,T) and it was observed in all the OsaTPS gene sequences except OsaTPS3, 14, 16, 21, and 25 (Fig 7). Motif DXDD was found only in OsaTPS-c subfamily (Fig 7), with sequence similarity range of 62% to 68% with TPS-c specific ent-copalyl diphosphate synthase, this motif was conserved in class II TPS genes, this class of TPS lack DDXXD motif and have non-active α-domain [66].

## Functional annotation and pathway mapping of Ocimum sanctum TPS genes

Discovered 47 OsaTPS of O. sanctum were annotated with the UniProt protein database to know their sequence similarity to the well-characterized terpene synthase genes available till date Table 2. The classified OsaTPS subfamily substantially mapped with the TPS genes of different species of Lamiaceae family; Ocimum basilicum, Origanum vulgare, Pogostemon cablin, Mentha x piperita, Salvia officinalis, Lavandula angustifolia, and with other species.

Gene ontology search was done to search the associated hits by sequence homology for their respective Gene Ontology (GO), Kyto Encyclopaedia of Genes and Genomes (KEGG) and Enzyme commission codes (EC) for each query sequence and the highest bit score hit were selected. Annotation against the GO database yielded significant annotation for 47 OsaTPS gene representing the best hits. All the OsaTPS genes were classified into three major components, the Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). In (BP) all 47 OsaTPS genes were representing their participation in biosynthetic process and lipid metabolism process, whereas 35 OsaTPS among 47 tend to participate in the catabolic process, 28 OsaTPS among 47 participate for another process. OsaTPS distribution in (CC) was represented in three components as 29 OsaTPS shown in the plastid, 6 OsaTPS in the cytosol and 5 OsaTPS in the cytoplasm. In Molecular function (MF) all 47 OsaTPS show lyase activity, where 41 among the 47 shows ion binding activity, and 12 OsaTPS genes show isomerase activity, along with this analysis (Fig 8 and S4 Table).

KEGG pathway mapping against the OsaTPS genes was performed to predict and hypothesize the secondary metabolite biosynthesis pathway. In monoterpenoid biosynthesis pathway, total 15 OsaTPS genes were found, which may be involved in the synthesis of the particular secondary metabolite. Similarly, in the diterpenoid pathway, 8 OsaTPS genes show their presence, whereas the maximum 23 OsaTPS were observed in the sesquiterpenoid biosynthesis pathway (Fig 9). OsaTPS19 shows homology with the isoprene synthase gene, which is used to synthesize the isoprene from dimethylallyl diphosphate.

https://doi.org/10.1371/journal.pone.0207097.g007
Discussion

*Ocimum* genus is a significant genus in the Lamiaceae family; it was considered as an important medicinal plant from centuries and at present also used in flavor and fragrance industries.
due to the occurrence of some aroma compounds like α-Pinene, camphene, eugenol, limonene, camphor in their essential oil. The Ocimum sanctum is rich in mono and sesquiterpenes, which are essential and interested in their contribution to industries. A lot of research is done in Ocimum species on phenylpropene class for characterization of the responsible genes for synthesis of eugenol, methyl chavicol, chavicol and their derivatives [25]. Being such an
essential plant for human life, this species has not been explored much in the context of molecular characterization of terpene synthases genes. In this study, we have used genome data of *Ocimum sanctum* [67] which is 386 Mb in size along with the plastid genome of 142,245 bp and known to be smallest in Lamiaceae family till date. In this study, the newly sequenced genome of *O. sanctum* was compared with *Salvia miltiorrhiza*. The comparative study reveals the two genomes to be most similar and also share identical diploid number of chromosome (2n = 16). It also reveals that *O. sanctum* genome is almost half the size of *S. miltiorrhiza* and appears to be relatively compact with quite fewer repeat sequences, while it falls in the identical phylogenetic clade with *S. miltiorrhiza*.

*S. miltiorrhiza* and *O. sanctum* both are rich in phenylpropanoids and their derivatives, which are well known for their therapeutic activities in Chinese and Indian traditional medicine systems. Although in *O. sanctum* the presence of a large number of terpene synthases homologs emphasizes to explore more specific and medicinally important terpene synthase genes in the plants. From this previous study, we have methodically identified the 47 putatively functional terpene synthases in *O. sanctum* which were responsible for the biosynthesis of volatile terpenes. These predicted *O. sanctum* TPSs were divided into seven TPS sub-families according to the standard TPSs subfamily classification methods [6,68]. We now extended the analysis by including newly identified TPS sequences from the sequenced genome of *Ocimum sanctum*. Out of 47 *OsaTPS* genes maximum number of genes is the member of TPS-a subfamily from *OsaTPS*1 to *OsaTPS*18. This clade is composed explicitly of angiosperm sesquiterpene and diterpene synthases and contains conserved D DXxD motif in the c-terminal domain of TPS which is supposed to combining metal ions, with less conserved RR(X)W motif presents at downstream of the N-terminal transit which confirms the presence of typical sesquiterpene and diterpene synthases. *OsaTPS* of TPS-a group that lack RR(X)W motif does not have the twin-arginine residue but have conserved tryptophan residues.

In TPS-a, subfamily, members of *O. sanctum* either formed separate cluster or share clusters with *S. lycopersicum* (Fig 4 and S2 Fig) and shows orthologous relationship with Class I specific sesquiterpene synthases of TPS-a subfamily genes i.e., cadinene synthase, Germacrene-D-synthase, Viridiflorene synthase, Selinene synthase, Bi-cyclogermacrene synthase and Cis-muuroladiene synthase. The previous subfamily classification suggests that TPS-b contains conserved RR(X)W belongs to monoterpenoid synthases and synthesizes angiosperm monoterpenoids, while in *O. sanctum* it was seen that *OsaTPS*20 to *OsaTPS*28 and *OsaTPS*31 genes contain highly conserved RR(X)W motif which confirms the predicted *OsaTPS* genes to be monoterpenoid synthases. Other *OsaTPS* of this class lack conservation of this motif. In TPS-b subfamily, *OsaTPS* shared common clusters with *S. lycopersicum*, *S. bicolor* and with *O. basilicum*, all *O. sanctum* TPS genes from *OsaTPS*19 to *OsaTPS*34 shows an orthologous relationship with a subfamily of TPS-b terpene synthase genes as represented in S3 Fig. Five *OsaTPS* genes were observed in TPS-c subfamily from *OsaTPS*35 to *OsaTPS*39, these TPS genes contain only DxD(D/V) motifs and lack DDXxD motif, and suggested to be monofunctional gene of class II specifically Copalyl diphosphate related diterpene synthases, these features and functional sequence annotation of *OsaTPS* confirm the sequences to me Copalyl diphosphate synthases.

The TPS-c, -e and -f subfamilies of *OsaTPS* were closely related to each other and shared a common cluster with *N. tabacum* and *S. lycopersicum*. TPS-ef contains all valid kaurene synthases proteins from angiosperms which encode and annotate *OsaTPS*40, the presence of DDXxD motif but not DxD motif and annotation with kaurene synthase TPS gene confirm it to be a class I TPS gene as mentioned in S4 Fig. The TPS-g is a clade presented closely related to the TPS-b family as shown in (Fig 5) and lack RRXW motif; it is defined and classified as
monoterpene synthases which produce cyclic floral aroma compounds like Ocimene and Myrcene which firstly characterized in snap-dragon [64].

Plant secondary metabolism pathway produces a large number of specific compounds. These compounds do not play a direct role in the plant growth and development but do help plants to survive in its environment. These metabolites were synthesized by specialized enzymes known as terpene synthases. In this study, we discovered a large number of terpene synthase genes in *Ocimum sanctum* genome although from a biological point of view it is not necessary that if the plant contains a large number of TPS genes it will synthesize the higher diversity of terpenoids because single terpene synthase gene can synthesize several terpenes from a single substrate. In a family Lamiaceae, some TPS genes have been characterized till date, so far only 11 TPS gene are reported in *Ocimum basilicum* of which, nine were characterized [25]. In *Lavandula angustifolia*, only 10 TPS genes were found among which only three were characterized. Likewise in *Mentha x piperita*, only six TPS genes were found, and four were functionally characterized so far, this comparison among closely related species of *O. sanctum* shows that *O. sanctum* possesses a large number of putative functional TPS genes.

The functional annotation and pathway mapping of *OsaTPS* genes reveals the putative function for production of secondary metabolites. In Fig 9, the predicted 47 *OsaTPS* genes mapped with secondary metabolite biosynthesis pathway, which represents the presence of biosynthetic genes in the species. *OsaTPS*46 and *OsaTPS*45 mapped with geraniol synthase which synthesizes geraniol, also reported in *Ocimum basilicum* species. Similarly we identified R-linalool synthase (*OsaTPS*47), (-)-Endo-fenchol synthase (*OsaTPS*33, and *OsaTPS*34), Selinene synthase (*OsaTPS*5, *OsaTPS*12 and *OsaTPS*17), Terpinolene synthase (*OsaTPS*31, *OsaTPS*32), β-myrcene synthase (*OsaTPS*29 and *OsaTPS*30), α-zingiberene synthase (*OsaTPS*23), γ-cadinene synthase (*OsaTPS*1, *OsaTPS*11 and *OsaTPS*18), and Germacrene-D synthase (*OsaTPS*2). Our study shows that a number of terpene synthases gene in *Ocimum* species (*Ocimum sanctum*) might be present and code other terpenoids of secondary metabolite pathway, especially in diterpene and sesquiterpene biosynthetic pathways. In diterpenoid biosynthesis *O. sanctum* predicted gene mapped with different TPS enzymes like Ent-copalyl diphosphate synthase (*OsaTPS*36), Syn-copalyl diphosphate synthase (*OsaTPS*35), Ent-kaurene synthase (*OsaTPS*40), Copal-8-ol diphosphate synthase (*OsaTPS*37, *OsaTPS*38, and *OsaTPS*39), and Cis-abienol synthase (*OsaTPS*41), these all diterpene synthase genes were still unexplored in this species. Similarly in sesquiterpenoid biosynthesis *OsaTPS* genes mapped with (+)-Epi-alpha bisabolol synthase (*OsaTPS*22), α-Farnesene synthase (*OsaTPS*20 and *OsaTPS*21), Germacrene-A synthase (*OsaTPS*07, *OsaTPS*08, *OsaTPS*09 and *OsaTPS*15), Vetispiradiene synthase (*OsaTPS*4, *OsaTPS*13), Viridiflorene synthase (*OsaTPS*3), Bicyclogermacrene synthase (*OsaTPS*6 and *OsaTPS*10), Cis-muuroladiene synthase (*OsaTPS*14 and *OsaTPS*16) and (3S, 6E)-Nerolidol synthase (*OsaTPS*44) (Fig 9).

The chemical composition of *Ocimum sanctum* essential oil also suggests the presence of a large number of terpenoids in the plant, which were synthesized by specific TPS genes, but few TPS gene studies have been done at molecular level. Therefore, there is a need for characterization of more TPS genes to know their actual involvement in the biosynthesis of *Ocimum sanctum* secondary metabolites.

**Conclusion**

In the present study, we have discovered and analyzed the terpene synthase genes families from *Ocimum sanctum* genome by using TPS gene HMM models, which is used to identify genome-wide terpene synthase genes. The study provides the first comprehensive prediction and annotation of the very large *OsaTPS* (*O. sanctum* TPS) gene family about genomic
Results showed that OsaTPS gene family is one of the largest gene family of specialized secondary metabolism in Ocimum species as there is no prior information about O. sanctum TPS genes. The predictions are exploring and unfolding the most significant gene family, which has been expanded across the genome through gene duplication and functional diversifications. Remarkably, a large number of functionally diverse sesquiterpene, diterpene and monoterpene synthase, which are identified in our study may contribute to enlighten the way for researchers. The phylogenetic analysis of OsaTPS showed similar results in many of the plant species studied so far and lies within all six gymnosperm TPS genes classified subfamilies, i.e., TPS-a, TPS-b, TPS-c, TPS-e, TPS-f, and TPS-g. This study may provide a deeper understanding of the O. sanctum putative functional genes phylogenetic classification and may help the researchers in the characterization of discovered OsaTPS genes to enhance the metabolic process of O. sanctum. This study represents the first report for the identification of O. sanctum TPS genes and their subfamily classification. These findings are initial steps for a better understanding of O. sanctum TPS genes. Further characterization is required to determine the proper function of enzymes in secondary metabolism.

Supporting information

S1 Fig. Multiple sequence alignment was performed and a phylogenetic tree was generated using neighbor joining method with 1,000 replicates for bootstrap values. Complete phylogenetic tree of 14 species terpene synthase genes, including newly sequenced O. sanctum and discovered putative functional 47 OsaTPS was generated. 437 TPS genes were included to generate the phylogenetic tree, with 47 OsaTPS gene, and categorized into the TPS-subfamily (TPS-a, -b, -c, -e, -f and TPS-g) as shown in the figure. All the newly discovered OsaTPS genes were designated in green color, where TPS subfamilies were assigned with different branch colors mentioned in figure legends. (TIF)

S2 Fig. Phylogenetic rooted tree of all TPS-a subfamily of different species was generated with the Neighbor-joining method, where 1,000 bootstraps replicated were considered and compared with O. sanctum TPS genes which were shown in pink color branches. The bootstrap values ≥80% were designated in red color triangles and values supporting ≥95% were designated in red circles above the branches, OsaTPS35 and Pt0092s00200 were shown as outgroup. (TIF)

S3 Fig. TPS-b and TPS-g subfamily phylogenetic tree were constructed using the Neighbor-joining method with 1,000 bootstrap replicates. O. sanctum TPS genes were shown in pink color branches. TPS-b was subdivided into TPS-b1 and TPS-b2, where TPS-b1 shows only S. lycopersicum species and TPS-b2 show multiple species with 16 TPS genes from O. sanctum. TPS-g subfamily shows 5 OsaTPS genes. Both the TPS subfamily were designated and colored in different colors, OsTPS1 was used as an outgroup. (TIF)

S4 Fig. The combined phylogenetic tree was generated using the Neighbor-joining method with 1,000 bootstrap replicates values. TPS-c, -e and TPS-f subfamily were joined together. OsaTPS1 is used as an outgroup. Terpene synthase genes of O. sanctum was shown in red branch color. (TIF)
S1 Table. Representation of *O. sanctum* OsaTPS genes with information of, genome localization, TPS subfamily clusters and gene length. (XLSX)

S2 Table. Representation of *O. sanctum* OsaTPS partial genes and pseudogenes with subfamily classification. (XLSX)

S3 Table. Cluster id of all TPS genes from different species included in phylogenetic tree analysis. (XLSX)

S4 Table. Representation of *O. sanctum* OsaTPS genes with annotation, Gene ontology information, sequence alignment distribution and KEGG enzyme mapping. (XLSX)

S1 File. All terpene synthase sequences retrieved from different plants species used for phylogenetic analysis. (DOCX)

Acknowledgments

We acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi for financial support through the projects, ChemBio (BSC0203 to A.K.S.), OLP-07 to F.K. at the CSIR-Central Institute of Medicinal and Aromatic Plants (CIMAP), Lucknow, India. Author Y.K. acknowledges the University Grant Commission (UGC), New Delhi for RGNF (SRF) fellowship support. The Bioinformatics research facility is supported by the grants from CSIR, New Delhi networking Project BSC0121 (GENESIS) to F.K. The CIMAP Communication No is CIMAP/PUB/2016/80. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: Feroz Khan.

Data curation: Yogesh Kumar.

Formal analysis: Yogesh Kumar, Feroz Khan, Shubhra Rastogi, Ajit Kumar Shasany.

Investigation: Feroz Khan.

Methodology: Yogesh Kumar.

Software: Yogesh Kumar.

Supervision: Feroz Khan.

Validation: Yogesh Kumar.

Visualization: Yogesh Kumar.

Writing – original draft: Yogesh Kumar.

Writing – review & editing: Feroz Khan, Ajit Kumar Shasany.

References

1. Holstein SA, Hohl RJ. Isoprenoids: remarkable diversity of form and function. Lipids. 2004; 39: 293–309. Available: [http://www.ncbi.nlm.nih.gov/pubmed/15357017](http://www.ncbi.nlm.nih.gov/pubmed/15357017) PMID: 15357017
2. Rohmer M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat Prod Rep. 1999; 16: 565–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/10584331 PMID: 10584331

3. Lichtenthaler HK. THE 1-DEOXY-D-XYLULOSE-5-PHOSPHATE PATHWAY OF ISOPRENOID BIO-SYNTHESIS IN PLANTS. Annu Rev Plant Physiol Plant Mol Biol. 1999; 50: 47–65. https://doi.org/10.1146/annurev.arplant.50.1.47 PMID: 1501203

4. Sapir-Mir M, Mett A., ... EB-P, 2008 undefined. Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is. Am Soc Plant Biol. Available: http://www.plantphysiol.org/content/148/3/1219.short

5. Kühlheim C, Yeoh SH, Wallis IR, Laffan S, Moran GF, Foley WJ. The molecular basis of quantitative variation in foliar secondary metabolites in Eucalyptus globulus. New Phytol. 2011; 191: 1041–53. https://doi.org/10.1111/j.1469-8137.2011.03769.x PMID: 21609332

6. Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci U S A. National Academy of Sciences; 1998; 95: 4126–33. https://doi.org/10.1073/pnas.95.8.4126 PMID: 9539701

7. Trapp SC, Croteau RB. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics. 2001; 158: 811–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/11404343 PMID: 11404343

8. Bohlmann J, Crock J., ... RJ-P of the, 1998 undefined. Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-a-bisabolene synthase from grand fir (Abies. Natl Acad Sci. Available: http://www.pnas.org/content/99/12/6756.short

9. Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci U S A. 1998; 95: 4126–4133. https://doi.org/10.1073/pnas.95.8.4126 PMID: 9539701

10. Chen F, Tholl D, Bohlmann J, Pichersky E. 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. 2011; 66: 212–229. https://doi.org/10.1111/j.1365-313X.2011.04520.x PMID: 2143633

11. Gao Y, Honzatko R, reports RP-N product, 2012 undefined. Terpenoid synthase structures: a so far incomplete view of complex catalysis. pubs.rsc.org. Available: http://pubs.rsc.org/en/content/articlehtml/2012/np/c2np20059g https://doi.org/10.1039/c2np20059g PMID: 22907771

12. Thoma R, Schulz-Gasch T, D’Arcy B, Benz J, Aebi J, Dehmlow H, et al. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature. 2004; 432: 118–122. https://doi.org/10.1038/nature02993 PMID: 15525992

13. Cao R, Zhang Y, Mann FM, Huang C, Mukkamala D, Haddock MP, et al. Diterpene cyclases and the nature of the isoprene fold. Protein Struct Funct Bioinforma. 2010; 78: 2417–2432. https://doi.org/10.1002/pro.22751 PMID: 20602361

14. Aubourg S, Lechamny A, Bohlmann J. Genomic analysis of the terpenoid synthase (ATTPS) gene family of Arabidopsis thaliana. Mol Genet Genomics. 2002; 267: 730–45. https://doi.org/10.1007/s00438-002-0709-y PMID: 12207221

15. Martin DM, Aubourg S, Schouwye MB, Daviet L, Schalk M, Toub O, et al. 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. https://doi.org/10.1186/1471-2229-10-226 PMID: 20964856

16. Martin DM, Aubourg S, Schouwye MB, Daviet L, Schalk M, Toub O, et al. 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. https://doi.org/10.1186/1471-2229-10-226 PMID: 20964856

17. Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinitdoehl I, et al. The Tomato Terpene Synthase Gene Family. PLANT Physiol. 2011; 157: 770–789. https://doi.org/10.1104/pp.111.179648 PMID: 21813695

18. Harris LJ, Saparno A, Johnston A, Prisic S, Xu M, Allard S, et al. The maize An2 gene is induced by Fusarium attack and encodes an ent-copalyl diphosphate synthase. Plant Mol Biol 2005; 59: 881–94. https://doi.org/10.1007/s11103-005-1674-8 PMID: 16307364

19. Li G, Kollner TG, Yin Y, Jiang Y, Chen H, Xu Y, et al. Nonseed plant Selaginella moellendorffii has both seed plant and microbial types of terpene synthases. Proc Natl Acad Sci. 2012; 109: 14711–14715. https://doi.org/10.1073/pnas.1204300109 PMID: 22908266

20. Varshney RK, Chen W, Li Y, Bharat AK, Saxena RK, Schlueter JA, et al. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat Biotechnol. 2012; 30: 83–89. https://doi.org/10.1038/nbt.2202 PMID: 22057054
21. Irmisch S, Jiang Y, Chen F, Gershenzon J, Köllner TG. Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (Populus trichocarpa). BMC Plant Biol. 2014; 14: 270. https://doi.org/10.1186/s12870-014-0270-y PMID: 25303804

22. Pattanayak P, Behera P, Das D, Panda S. Ocimum sanctum Linn. A reservoir plant for therapeutic applications: An overview. Pharmacogn Rev. 2010; 4: 95. https://doi.org/10.4103/0973-7847.65323 PMID: 22228948

23. Kothari SK, Bhattacharyya AK, Ramesh S. Essential oil yield and quality of methyl eugenol rich Ocimum tenuiflorum L.f. (syn. O. sanctum L.) grown in south India as influenced by method of harvest. J Chromatogr A. 2004; 1054: 67–72. Available: http://www.ncbi.nlm.nih.gov/pubmed/15553132 PMID: 15553132

24. Omidvarborna H, Kumar A, Chemical DK-J of E, 2014 undefined. Characterization of particulate matter emitted from transit buses fueled with B20 in idle modes. Elsevier. Available: https://www.sciencedirect.com/science/article/pii/S2213343714002140

25. Gang DR. An Investigation of the Storage and Biosynthesis of Phenylpropanes in Sweet Basil. PLANT Physiol. 2001; 125: 539–555. https://doi.org/10.1104/pp.125.2.539 PMID: 11161012

26. Islam AKMM, Kato-Noguchi H. Phytotoxic activity of Ocimum tenuiflorum extracts on germination and seedling growth of different plant species. Sci World J. 2014; 2014: 1–8. https://doi.org/10.1155/2014/676242 PMID: 25032234

27. Lange BM. The Evolution of Plant Secretory Structures and Emergence of Terpenoid Chemical Diversity. Annu Rev Plant Biol. Annual Reviews; 2015; 66: 139–159. https://doi.org/10.1146/annurev-plant-040314-114639 PMID: 25621517

28. Ringer KL, Davis EM, Croteau R. Monoterpane Metabolism. Cloning, Expression, and Characterization Deguerry F, Pastore L, Wu S, Clark A, Chappell J, Schalk M. The diverse sesquiterpene profile of patchouli, Pogostemon cablin, is correlated with a limited number of sesquiterpene synthases. Arch Biochem Biophys. 2005; 405: 112–21. Available: http://www.ncbi.nlm.nih.gov/pubmed/3038029 PMID: 3038029

29. Cui G, Duan L, Jin B, Qian J, Xue Z, et al. Identification of a sesquiterpene synthase gene from peppermint (Mentha x piperita, L.) that produces the aphid alarm pheromone (E)-beta-farnesene. Phytochemistry. 2006; 67: 1564–1571. https://doi.org/10.1016/j.phytochem.2005.06.012 PMID: 16083926

30. Cui G, Duan L, Jin B, Qian J, Xue Z, et al. IeCPS2 is potentially involved in the biosynthesis of shisanol biosynthesis provides stereochemical insights. ACS Publ. Available: https://pubs.acs.org/doi/10.1021/ol902051v

31. Su P, Tong Y, Cheng Q, Hu Y, Zhang M, reports JY-S, et al. Functional divergences of diterpene synthases in the medicinal plant Salvia miltiorrhiza Bunge. Am Soc Plant Biol. Available: http://www.plantphysiol.org/content/early/2015/06/15/pp.15.00695.short

32. Crowell AL, Williams DC, Davis EM, Wildung MR, Croteau R. Molecular cloning and characterization of a new linalool synthase. Arch Biochem Biophys. 2006; 454: 123–136. https://doi.org/10.1016/j.abb.2006.08.006 PMID: 16970904

33. Cui G, Duan L, Jin B, Qian J, Xue Z, et al. Functional divergence of diterpene synthases in the medicinal plant Salvia miltiorrhiza Bunge. Am Soc Plant Biol. Available: http://www.plantphysiol.org/content/early/2015/06/15/pp.15.00695.short

34. Su P, Tong Y, Cheng Q, Hu Y, Zhang M, reports JY-S, et al. Functional characterization of ent-copalyl diphosphate synthase, kaurene synthase and kaurene oxidase in the Salvia miltiorrhiza giberellin biosynthetic pathway. nature.com. Available: https://www.nature.com/articles/sep23057 PMID: 26971881

35. Su P, Tong Y, Cheng Q, Hu Y, Zhang M, reports JY-S, et al. Functional characterization of ent-copalyl diphosphate synthase, kaurene synthase and kaurene oxidase in the Salvia miltiorrhiza giberellin biosynthetic pathway. nature.com. Available: https://www.nature.com/articles/sep23057 PMID: 26971881

36. Li J, Chen Q, Jin Q, Gao J, Zhao P, Lu S, et al. laCPS2 is potentially involved in the biosynthesis of pharmacologically active Isodon diterpenoids rather than gibberellin. Elsevier. Available: https://www.sciencedirect.com/science/article/pii/S003194221100611X

37. Liu M, Cui G, Guo J, Tang J, Duan L, et al. Functional diversification of kaurene synthase-like genes. Am Soc Plant Biol. Available: http://www.plantphysiol.org/content/early/2017/04/05/pp.17.00202.abstract

38. Jia X, Pateraki I, Andersen-Ranberg J, Hamberger B, Chesney AM, Martens HJ, Zerbe P, et al. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolin, Is Synthesized in Specialized Root Cork Cells in Coleus

39. Jia X, Pateraki I, Andersen-Ranberg J, Hamberger B, Chesney AM, Martens HJ, Zerbe P, et al. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolin, Is Synthesized in Specialized Root Cork Cells in Coleus
forskohlii. PLANT Physiol. 2014; 164: 1222–1236. https://doi.org/10.1104/pp.113.228429 PMID: 24481136

40. Božič D, Papaefthimiu D, Brückner K, de Vos RCH, Tsoulices CA, Katsarou D, et al. Towards Elucidating Carnosic Acid Biosynthesis in Lamiaceae: Functional Characterization of the Three First Steps of the Pathway in Salvia fruticosa and Rosmarinus officinalis. Hamberger B, editor. PLoS One. Public Library of Science; 2015; 10: e0124106. https://doi.org/10.1371/journal.pone.0124106 PMID: 26020634

41. Trikka FA, Nikolaidis A, Ignea C, Tsaballa A, Tziveleka L-A, Ioannou E, et al. Combined metabolome and transcriptome profiling provides new insights into diterpene biosynthesis in S. pomifera glandular trichomes. BMC Genomics. 2015; 16: 935. https://doi.org/10.1186/s12864-015-2147-3 PMID: 26572682

42. Caniard A, Zerbe P, Legrand S, Cohade A, Valot N, Magnard J-L, et al. Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in Salvia sclarea (L.) and their relevance for perfume manufacture. BMC Plant Biol. 2012; 12: 119. https://doi.org/10.1186/1471-2229-12-119 PMID: 22834731

43. Schalk M, Pastore L, Mirata MA, Khim S, Schouwey M, Deguerry F, et al. Toward a Biosynthetic Route to Sclareol and Amber Odorants. J Am Chem Soc. 2012; 134: 18900–18903. https://doi.org/10.1021/ja307404u PMID: 23113661

44. Pelot K, Mitchell R, Kwon M, . . . DH-TP, 2017 undefined. Biosynthesis of the psychotropic plant diterpene salvinorin A: Discovery and characterization of the Salvia divinorum clorodienyl diphosphate synthase. Wiley Online Lib. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpt.13427

45. Crocoll C, Asbach J, Novak J, Gershenzon J, Degenhardt J. Terpene synthases of oregano (Origanum vulgare L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Mol Biol. Springer Netherlands; 2010; 73: 567–603. https://doi.org/10.1007/s11103-010-9636-1 PMID: 20419468

46. Yahyaa M, Matsuba Y, Brandt W, Doron-Faigenboim A, Bar E, McClain A, et al. Identification, Functional Characterization, and Evolution of Terpene Synthases from a Basal Dicot. Plant Physiol. 2015; https://doi.org/10.1104/pp.15.00930 PMID: 26157114

47. Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E. Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil. PLANT Physiol. 2004; 137: 370–379. https://doi.org/10.1104/pp.103.032496 PMID: 14657409

48. Iijima Y, Davidovich-Rikanati R, Fridman E, Gang DR, Bar E, Lewinsohn E, et al. The Biochemical and Molecular Basis for the Divergent Patterns in the Biosynthesis of Terpenes and Phenylpropanes in the Peltate Glands of Three Cultivars of Basil. PLANT Physiol. 2004; 136: 3724–3736. https://doi.org/10.1104/pp.104.051318 PMID: 15516500

49. Davidovich-Rikanati R, Lewinsohn E, Bar E, Iijima Y, Pichersky E, Sitrit Y. Overexpression of the lemon basil α-zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit. Plant J. 2008; 56: 228–238. https://doi.org/10.1111/j.1365-313X.2008.03599.x PMID: 18643974

50. Misra RC, Maiti P, Chanotiya CS, Shanker K, Ghosh S. Methyl jasmonate-elicited transcriptional responses and pentacyclic triterpene biosynthesis in sweet basil. Plant Physiol. American Society of Plant Biologists; 2014; 164: 1028–44. https://doi.org/10.1104/pp.113.232884 PMID: 24367017

51. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. Oxford University Press; 2016; 44: D279–D285. https://doi.org/10.1093/nar/gkv1344 PMID: 26673716

52. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. Oxford University Press; 2011; 39: W29–W37. https://doi.org/10.1093/nar/gkr367 PMID: 21593126

53. Hu P, Wang D, Cassidy MJ, Stanier SA. Predicting the resistance profile of a spudcan penetrating sand overlying clay. Can Geotech J. NRC Research Press; 2014; 51: 1151–1164. https://doi.org/10.1139/cgj-2013-0374

54. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. American Association for the Advancement of Science; 2002; 296: 92–100. https://doi.org/10.1126/science.1068275 PMID: 11935018

55. Characterization TFPC for GG, Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. Nature Publishing Group; 2007; 449: 463–467. https://doi.org/10.1038/nature06148 PMID: 17721507

56. Tuskan GA, Diffazio F, Jansson S, Bohmlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science. American Association for the Advancement of Science; 2006; 313: 1596–604. https://doi.org/10.1126/science.1128691 PMID: 16973872
57. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. Nature Publishing Group; 2009; 457: 551–556. https://doi.org/10.1038/nature07723 PMID: 19189423

58. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of Eucalyptus grandis. Nature. Nature Publishing Group; 2014; 510: 356–362. https://doi.org/10.1038/nature13308 PMID: 24919147

59. Kühlheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg AA, et al. The Eucalyptus terpene synthase gene family. BMC Genomics. 2015; 16: 450. https://doi.org/10.1186/s12864-015-1598-x PMID: 26062733

60. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. Oxford University Press; 2004; 32: 1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

61. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. Oxford University Press; 2013; 30: 2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

62. Morariu VI, Srinivasan BV, Raykar VC, Duraiswami R, Davis LS. Automatic online tuning for fast Gaussian summation. Transform. 2008; 1–8. Available: http://papers.nips.cc/paper/3420-automatic-online-tuning-for-fast-gaussian-summation

63. American Society for Horticultural Science. Y, Xia D, Louzada ES. Journal of the American Society for Horticultural Science. [Internet]. Journal of the American Society for Horticultural Science. [American Society for Horticultural Science]; 2005. Available: http://journal.ashspublications.org/content/130/3/454.short

64. Dudareva N, Martin D, Kish C, . . . NK-TP, 2003 undefined. (E)-β-Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene. Am Soc Plant Biol. Available: http://www.plantcell.org/content/15/5/1227.short

65. Martin D, Phytochemistry JB-, 2004 undefined. Identification of Vitis vinifera (−)−α-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. Elsevier. Available: https://www.sciencedirect.com/science/article/pii/S0031942204001190

66. Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry. Pergamon; 2009; 70: 1621–1637. https://doi.org/10.1016/J.PHYTOCHEM.2009.07.030 PMID: 19793600

67. Rastogi S, Kalra A, Gupta V, Khan F, Lal RK, Tripathi AK, et al. Unravelling the genome of Holy basil: an &quot;incomparable&quot; &quot;elixir of life&quot; of traditional Indian medicine. BMC Genomics. 2015; 16: 413. https://doi.org/10.1186/s12864-015-1640-z PMID: 26017011

68. Chen F, Tholl D, Bohm J, Journal EP-TP, 2011 undefined. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Wiley Online Libr. Available: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-313X.2011.04920.x