

RESEARCH ARTICLE

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

Yogesh Kumar¹, Feroz Khan^{1*}, 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 phytochemicals 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 hmsearch for the two Pfam domains (PF01397 and PF03936). Using this search method 81 putative terpene synthase genes (*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.

OPEN ACCESS

Citation: Kumar Y, Khan F, Rastogi S, Shasany AK (2018) Genome-wide detection of terpene synthase genes in holy basil (*Ocimum sanctum* L.). PLoS ONE 13(11): e0207097. <https://doi.org/10.1371/journal.pone.0207097>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: March 22, 2018

Accepted: October 24, 2018

Published: November 16, 2018

Copyright: © 2018 Kumar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the required and supporting data for this manuscript are attached in supplementary files.

Funding: 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

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 isopentyl pyrophosphate, monoterpene (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 monoterpene 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

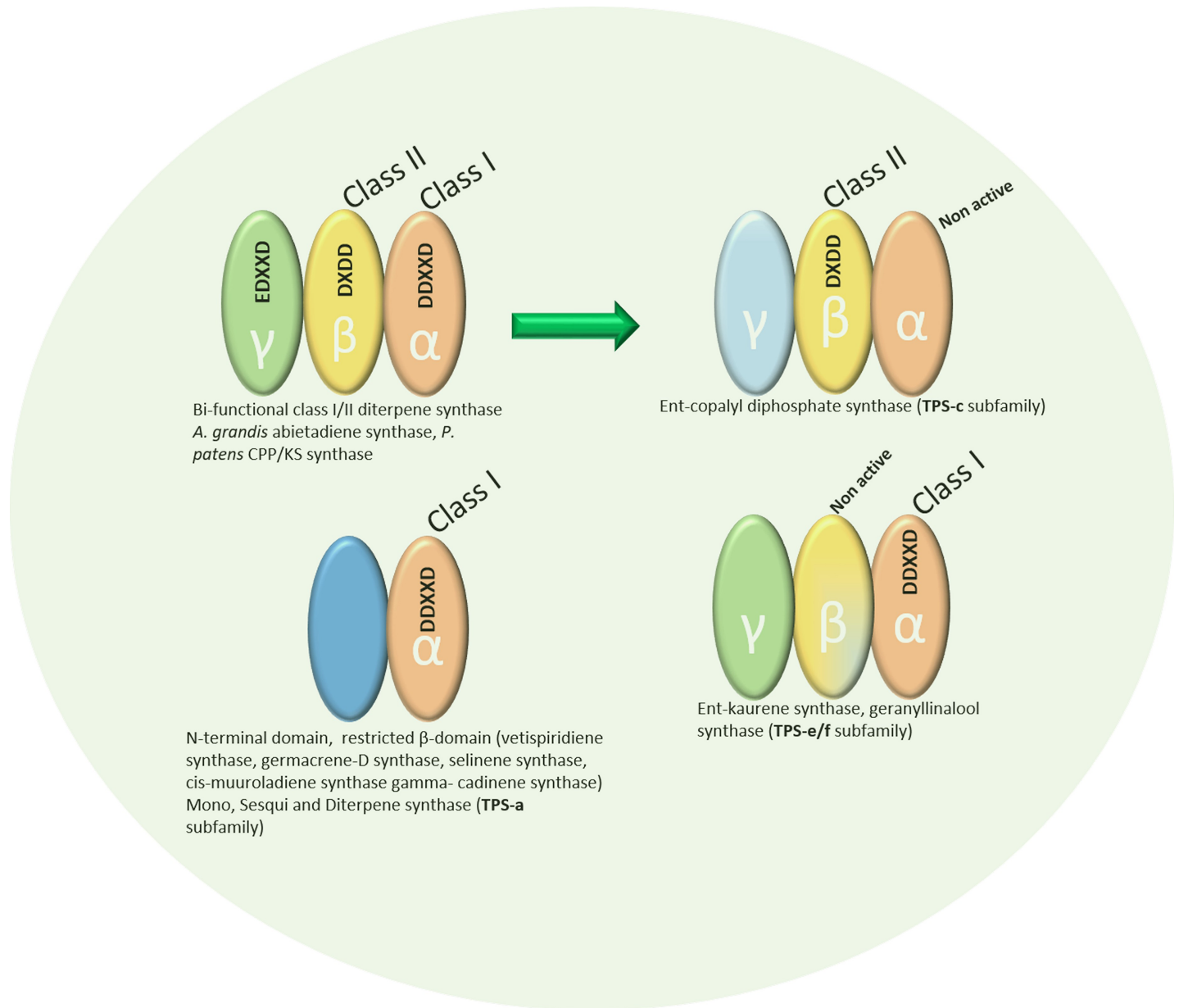


Fig 1. Representation of class I and class II domain within terpene synthase protein structure; class I consists of conserved DDXXD motif in its α -domain, whereas class II contains DXDD conserved motif.

<https://doi.org/10.1371/journal.pone.0207097.g001>

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

Table 1. Terpene synthase genes size in different plant species, including *O. sanctum* genome.

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

¹Total number of terpene synthase genes identified with the TPS HMM sequence models.

²Putative full-length genes retrieved from available data.

³Putative TPS genes classified into subfamily by phylogenetic analysis.

^SNumbers of TPS genes retrieved from the information based on literature data.

[@]TPS sequence data retrieved from the non-redundant protein sequence database from NCBI

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

[22,23]. 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 fruticose* [40], *Salvia pomifera* [41], *Salvia sclarea* [42,43] and *Salvia divimorum* [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 monoterpene 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

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.plantgdb.org/OsGDB/>) [54], *Vitis vinifera* (<http://www.plantgdb.org/VvGDB/>), *Solanum lycopersicum* Potato Genome Sequencing Consortium [55], *Populus trichocarpa* (<http://www.plantgdb.org/PtGDB/>) [56], *Sorghum bicolor* (<http://www.plantgdb.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 sub-family 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)₈W, 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*, *Hyoscyamus 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 *OsaTPS4* 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 *OsaTPS21* 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-

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

S. No	Gene code	TPS subfamily	Protein length	Gene ID (gi)	Organism
1	<i>OsaTPS1</i>	TPSa	507	75251482	<i>Ocimum basilicum</i>
2	<i>OsaTPS2</i>	TPSa	535	75251483	<i>Ocimum basilicum</i>
3	<i>OsaTPS3</i>	TPSa	320	939319666	<i>Solanum lycopersicum</i>
4	<i>OsaTPS4</i>	TPSa	577	75267592	<i>Solanum tuberosum</i>
5	<i>OsaTPS5</i>	TPSa	566	75251484	<i>Ocimum basilicum</i>
6	<i>OsaTPS6</i>	TPSa	383	403399735	<i>Origanum vulgare</i>
7	<i>OsaTPS7</i>	TPSa	545	122219293	<i>Pogostemon cablin</i>
8	<i>OsaTPS8</i>	TPSa	550	122219293	<i>Pogostemon cablin</i>
9	<i>OsaTPS9</i>	TPSa	543	122219293	<i>Pogostemon cablin</i>
10	<i>OsaTPS10</i>	TPSa	540	476007192	<i>Phyla dulcis</i>
11	<i>OsaTPS11</i>	TPSa	521	75251482	<i>Ocimum basilicum</i>
12	<i>OsaTPS12</i>	TPSa	551	75251484	<i>Ocimum basilicum</i>
13	<i>OsaTPS13</i>	TPSa	468	476007821	<i>Hyoscyamus muticus</i>
14	<i>OsaTPS14</i>	TPSa	383	75252096	<i>Mentha x piperita</i>
15	<i>OsaTPS15</i>	TPSa	274	122219293	<i>Pogostemon cablin</i>
16	<i>OsaTPS16</i>	TPSa	379	75252096	<i>Mentha x piperita</i>
17	<i>OsaTPS17</i>	TPSa	507	75251484	<i>Ocimum basilicum</i>
18	<i>OsaTPS18</i>	TPSa	388	75251482	<i>Ocimum basilicum</i>
19	<i>OsaTPS19</i>	TPSb	533	75232269	<i>Populus tremuloides</i>
20	<i>OsaTPS20</i>	TPSb	530	487524480	<i>Ricinus communis</i>
21	<i>OsaTPS21</i>	TPSb	316	487524480	<i>Ricinus communis</i>
22	<i>OsaTPS22</i>	TPSb	546	476007226	<i>Phyla dulcis</i>
23	<i>OsaTPS23</i>	TPSb	512	75251481	<i>Ocimum basilicum</i>
24	<i>OsaTPS24</i>	TPSb	558	122210944	<i>Lavandula angustifolia</i>
25	<i>OsaTPS25</i>	TPSb	356	62900763	<i>Salvia officinalis</i>
26	<i>OsaTPS26</i>	TPSb	524	122210944	<i>Lavandula angustifolia</i>
27	<i>OsaTPS27</i>	TPSb	589	62900763	<i>Salvia officinalis</i>
28	<i>OsaTPS28</i>	TPSb	575	62900763	<i>Salvia officinalis</i>
29	<i>OsaTPS29</i>	TPSb	622	75251478	<i>Ocimum basilicum</i>
30	<i>OsaTPS30</i>	TPSb	650	75251478	<i>Ocimum basilicum</i>
31	<i>OsaTPS31</i>	TPSb	350	75251477	<i>Ocimum basilicum</i>
32	<i>OsaTPS32</i>	TPSb	512	75251477	<i>Ocimum basilicum</i>
33	<i>OsaTPS33</i>	TPSb	601	75251479	<i>Ocimum basilicum</i>
34	<i>OsaTPS34</i>	TPSb	556	75251479	<i>Ocimum basilicum</i>
35	<i>OsaTPS35</i>	TPSc	826	62900382	<i>Pisum sativum</i>
36	<i>OsaTPS36</i>	TPSc	783	62900382	<i>Pisum sativum</i>
37	<i>OsaTPS37</i>	TPSc	824	544602148	<i>Nicotiana tabacum</i>
38	<i>OsaTPS38</i>	TPSc	826	544602148	<i>Nicotiana tabacum</i>
39	<i>OsaTPS39</i>	TPSc	821	544602148	<i>Nicotiana tabacum</i>
40	<i>OsaTPS40</i>	TPSe	812	62900385	<i>Cucurbita maxima</i>
41	<i>OsaTPS41</i>	TPSe	642	544602165	<i>Nicotiana tabacum</i>
42	<i>OsaTPS42</i>	TPSf	829	75163668	<i>Arabidopsis thaliana</i>
43	<i>OsaTPS43</i>	TPSg	426	75242446	<i>Antirrhinum majus</i>
44	<i>OsaTPS44</i>	TPSg	529	332319699	<i>Fragaria x ananassa</i>
45	<i>OsaTPS45</i>	TPSg	560	75224312	<i>Ocimum basilicum</i>
46	<i>OsaTPS46</i>	TPSg	467	75224312	<i>Ocimum basilicum</i>
47	<i>OsaTPS47</i>	TPSg	395	75251480	<i>Ocimum basilicum</i>

<https://doi.org/10.1371/journal.pone.0207097.t002>

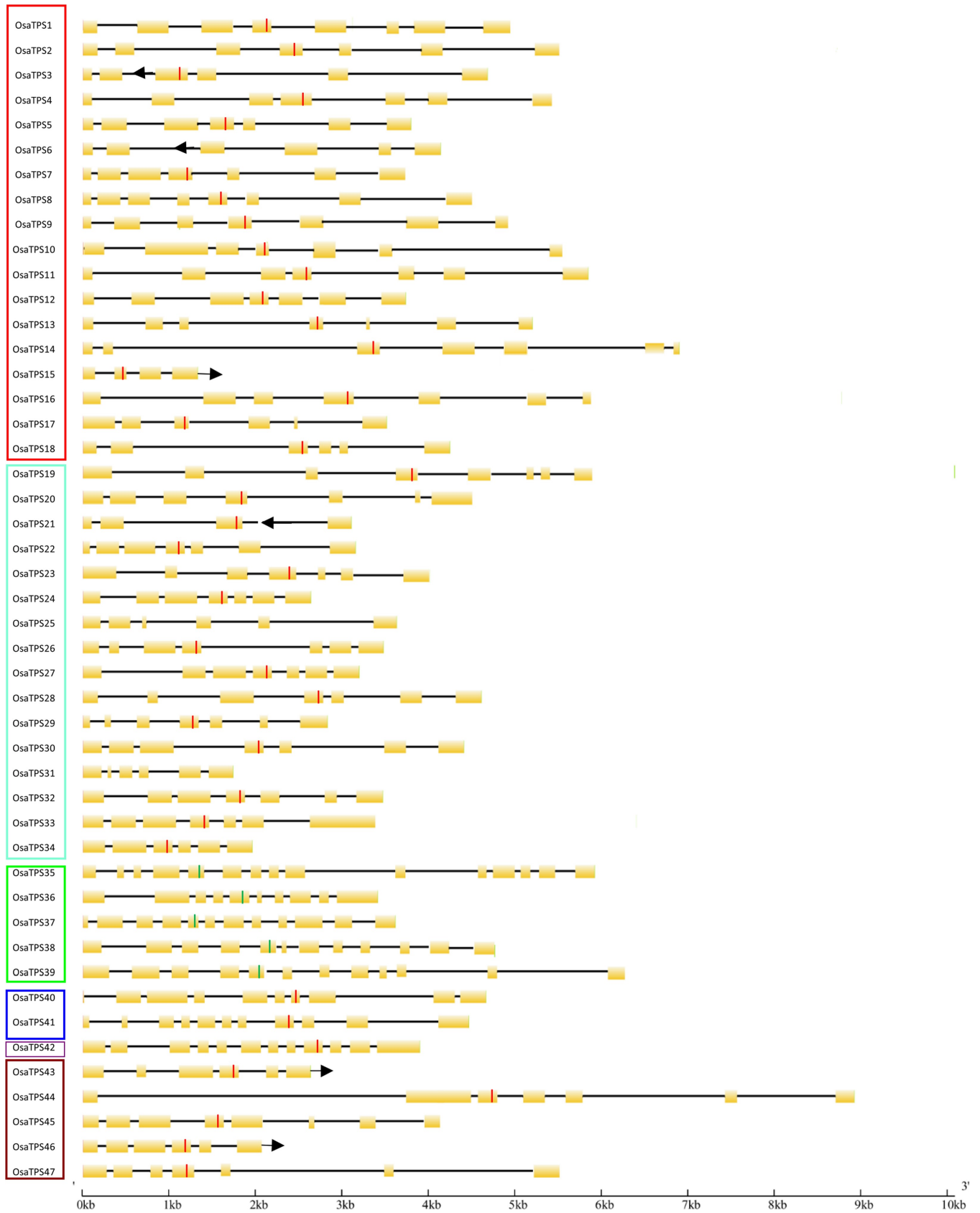


Fig 2. Representation of the gene structure of putative functional TPS genes from *Ocimum sanctum*. Genes were predicted by FGENESH tool and Augustus but also determined manually by mapping to previously TPS characterized genes. The black line indicates the introns; yellow rectangles indicate exons, 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*).

<https://doi.org/10.1371/journal.pone.0207097.g002>

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 tabacum*, *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 distantly 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 synthase proteins, 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 monoterpene 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 monoterpene 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 *OsaTP35* paired with multiple species.

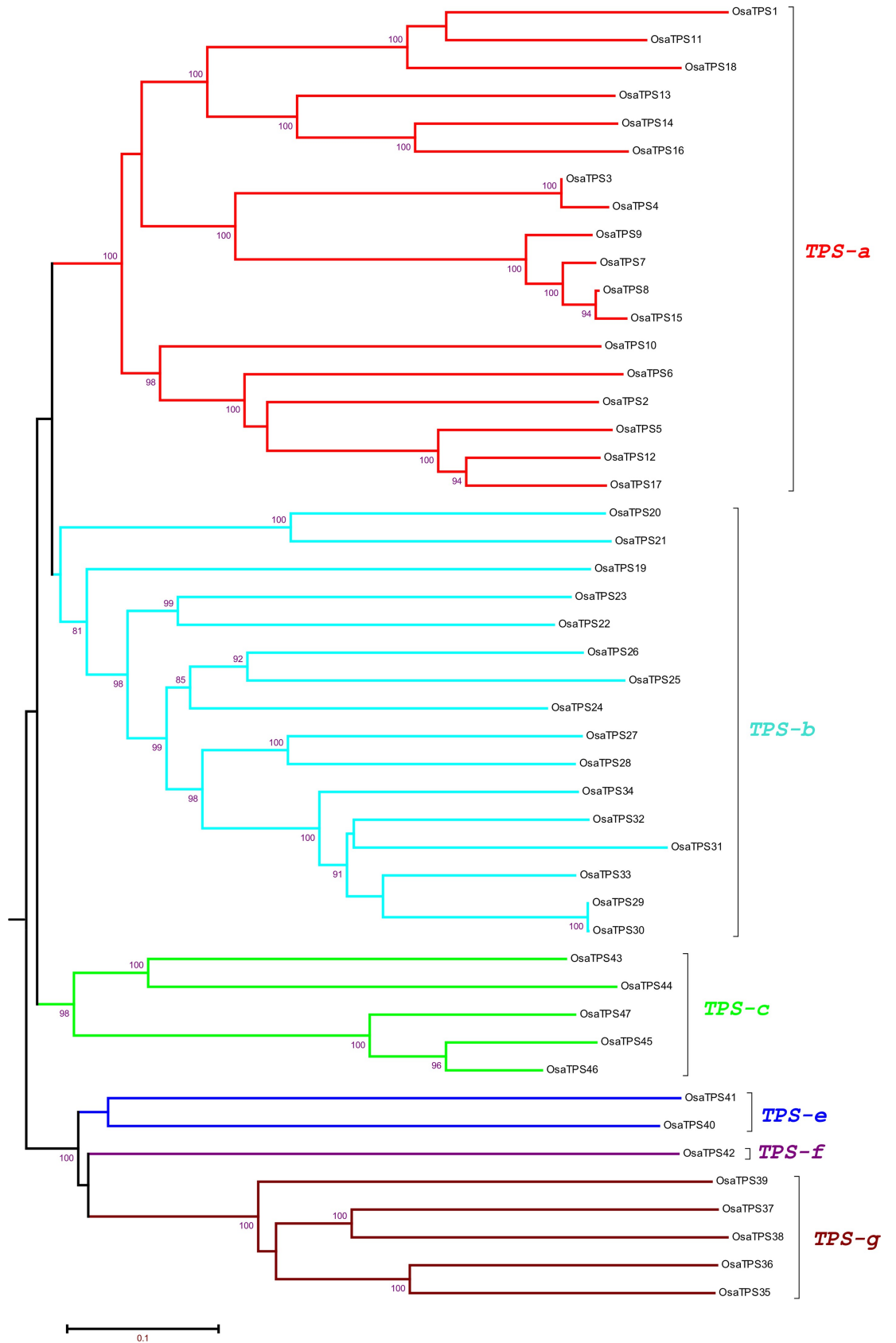


Fig 3. Discovered 47 putative functional TPS genes from *O. sanctum*: Neighbor-joining phylogenetic tree was constructed using 1,000 bootstrap value and classified into terpene synthase gene subfamily as TPS-a, -b, -c, -e, -f and TPS-g. Bootstrap values assigned in the tree, which was higher than 80%, values below 80% were not shown in the figure.

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

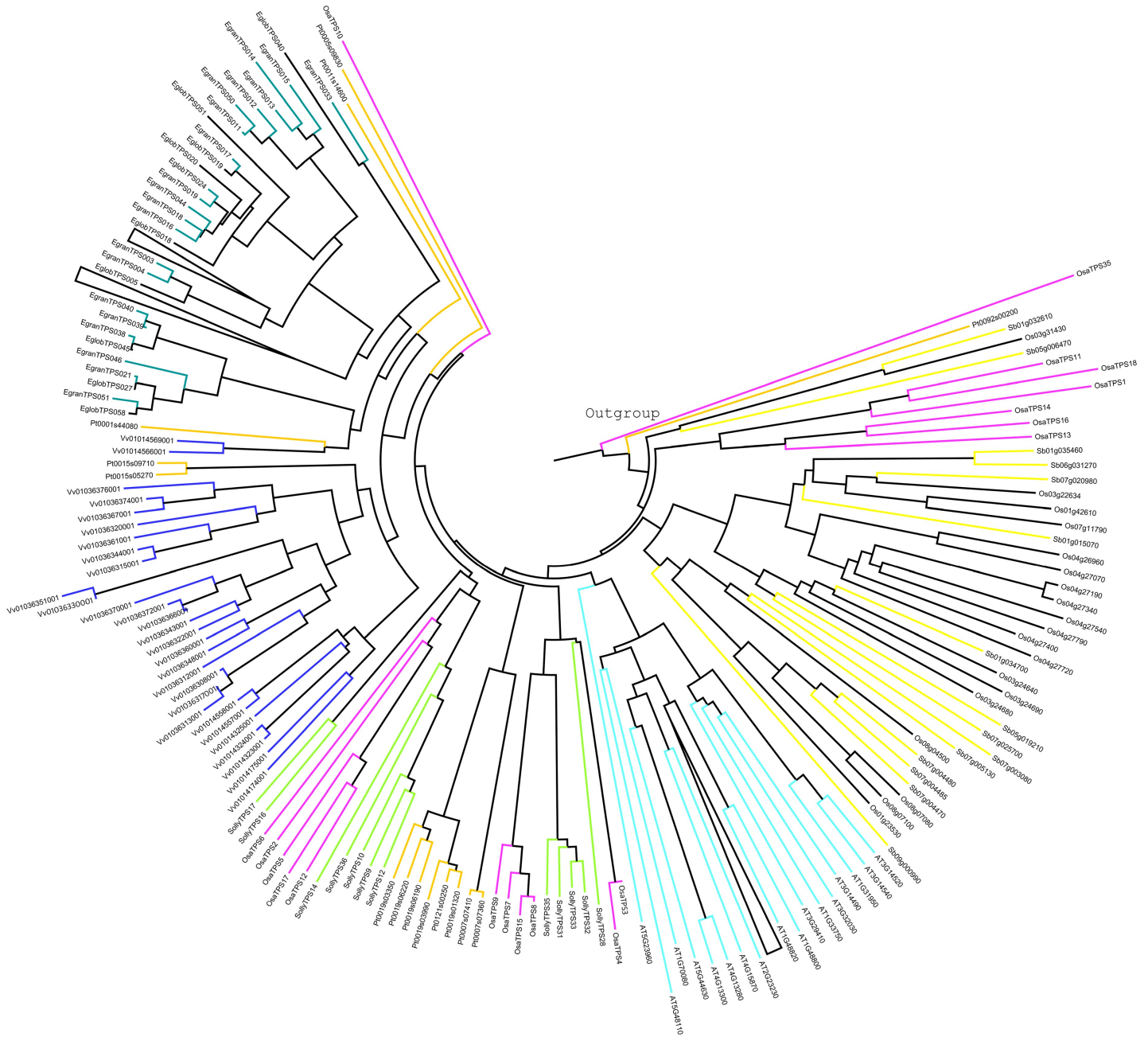


Fig 4. Representation of phylogenetic tree of *O. sanctum* compared to other species of TPS-a subfamily. 18 *OsaTPS* represented in the tree were compared with the neighboring species. The Neighbor-joining method was used with 1,000 replicates for bootstrap values. The values = 80% and = 95% were designated in S2 Fig. *OsaTPS35* and *Pt0092s00200*, two TPS gene were used as an out-group.

<https://doi.org/10.1371/journal.pone.0207097.g004>

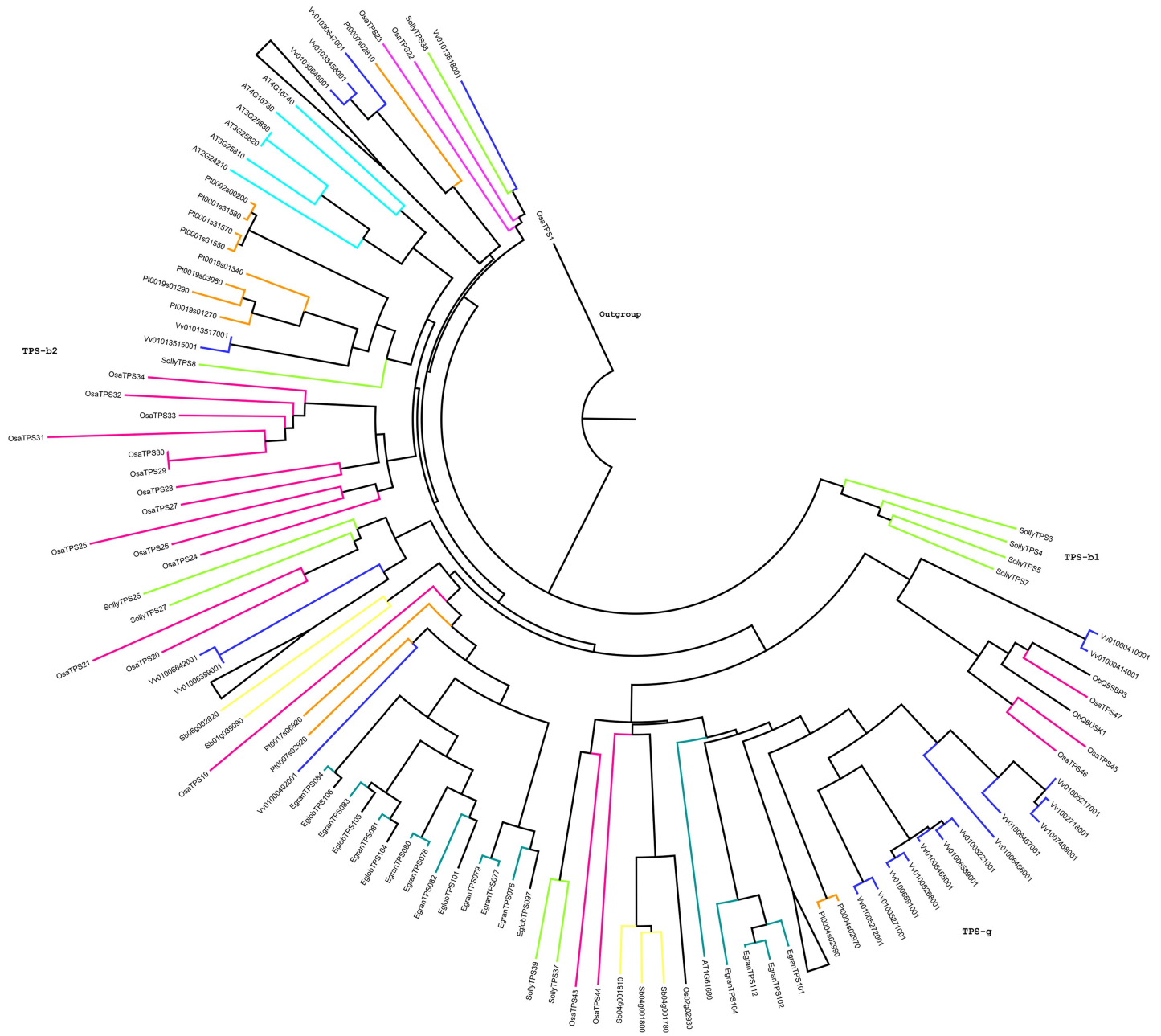


Fig 5. Phylogenetic tree of putative functional TPS genes from *O. sanctum* were represented, here TPS-b and TPS-g were grouped in a single tree. TPS-b was sub-classified into TPS-b1 and TPS-b2, where TPS-b1 shows only *S. lycopersicum* species TPS genes and TPS-b2 shows mixed species in which 16 *Osa*TPS were shown. In TPS-g subfamily 5 *Osa*TPS were represented. The phylogenetic tree was constructed with 1,000 replicates for bootstrap values using the neighbor-joining method.

<https://doi.org/10.1371/journal.pone.0207097.g005>

Conserved sequence motif analysis of *O. sanctum* TPSs

All the discovered *Osa*TPS 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. Monoterpene 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 *Osa*TPS genes in exact or in modified form, which

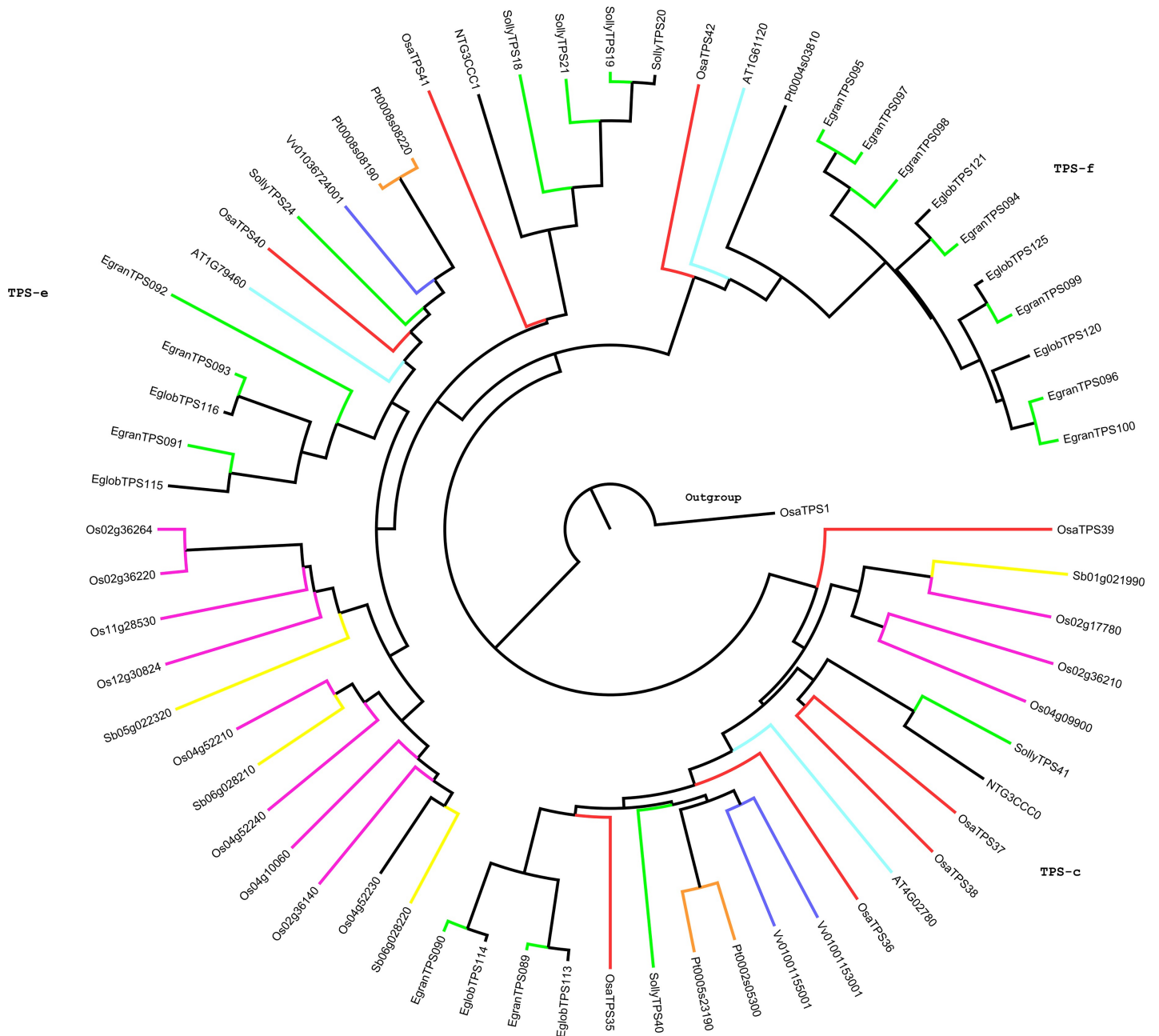


Fig 6. TPS-c, -e, and TPS-f, combined phylogenetic tree were represented from *O. sanctum* putatively functional TPS genes compared with other species. The phylogenetic tree was constructed with 1,000 replicates for bootstrap values using the neighbor-joining method. Here in TPS-c clade five *OsaTPS*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

resemble sesquiterpene, diterpene and monoterpene synthases. In TPS-b clade $RR(x)_sW$ 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

OsaTPS gene name	Subfamily	RR (X) 8W	NSE/DTE	RXR	DDXD/DXDD
OsaTPS1	a	PPVSSFVPCIW	RYWDDTGSHKRE	LFRDRV	GIMDDTYDNYA
OsaTPS2	a	RPSVTHPNVW	RLMDDIAGYGSE	FARDRL	SVLDDIYDVYG
OsaTPS3	a	RPSSNFSPSLW	-----	YARDRM	SVIDDTYDSHG
OsaTPS4	a	RPSSNFSPSLW	RVINDVATYEIE	YARDRM	SVIDDTYDSHG
OsaTPS5	a	RKSKTYHSSVW	RLMDDLGHGKE	FARDRL	TIIDDIYDVYG
OsaTPS6	a	RRSAKFHPSIW	RLMDDLVDG--E	-----	DRDVAVSL--
OsaTPS7	a	RPVANFSPSLW	RVIDDIATYEVE	YARDRV	SIIDDTYDAYG
OsaTPS8	a	RPVANFSPSLW	RVIDDIATYEVE	YARDRV	SIIDDTYDAYG
OsaTPS9	a	RPVANFSPSLW	RVIDDIATYEVE	YARDRV	SIIDDTYDAYG
OsaTPS10	a	RPLANYHPNIW	RLTDDLVGHEFE	YARDRM	SIMDDTYDAYA
OsaTPS11	a	PPITIHAPSIW	RYRGDIASHKRE	YIRDRL	GIMDDTYDNYA
OsaTPS12	a	RKSVTYHPSVW	RLMDDLAGYGKE	FPRDRL	TIVDDIYDVYG
OsaTPS13	a	PPTVAYKPSMW	--MEDLGHERE	YARDRI	SIMDDTYDNYA
OsaTPS14	a	PPSFIFKPSIW	-----	YARHRM	SIMDDTYDNYA
OsaTPS15	a	-----	RVIDDIATYEVE	-----	SIIDDTYDAYG
OsaTPS16	a	PPMISFKPSMW	-----	YARHRL	TIMDDTYDNYA
OsaTPS17	a	-----	RLMDDLGHGKE	FARDRL	TVIDDIYDVYG
OsaTPS18	a	PPVTSHAPSIW	-FRDNKHLRRE	YIRDRL	GIMDDTYDNYA
OsaTPS19	b	-----	RLANDLSLYEEE	FARDRV	TVLDDIFDVYG
OsaTPS20	b	RRTANYKPNIW	RLCNDQGTSAE	FTRDRV	LIIDDDYDIYG
OsaTPS21	b	RRSANYQPNIW	-----	FSRDRV	VIIDDDYDIYG
OsaTPS22	b	RRSANYKPSIW	RLADDLGTSTEE	FARDRL	TIIDDIYDVYG
OsaTPS23	b	RRSANLGASIW	RLADDLGT---	FARHRL	TILDDIYDVYG
OsaTPS24	b	RRSGMYKPTLW	RLADDLGTSPHE	FARDRV	TVIDDDYDVYG
OsaTPS25	b	RRSGNYKPSHW	-----E	FVRDRT	-----E
OsaTPS26	b	RRSANYKPPFW	RLANDLATSQNE	FARDRV	TIIDDFDVYA
OsaTPS27	b	RRSGGYHPTLW	RLADDMGTSLDE	FVRDRI	SAIDDDYDVYG
OsaTPS28	b	RRSGGYQPALW	RLADDMGTSLDE	FARDRI	TTIDDDYDIYG
OsaTPS29	b	R-SGNYQPSAW	RLPDDLGTAPLE	FARDRI	TVVDDYDVYG
OsaTPS30	b	R-SGNYQPSAW	RLPDDLGTAPLE	FARDRI	TVVDDYDVYG
OsaTPS31	b	RRSGNYEPSAW	RLPDDLGTASHE	-----	HLGEDTFELLD
OsaTPS32	b	R-SGNYQPSAW	RLPDDIGTAWIT	FARDRI	TVIDDFDVYG
OsaTPS33	b	R-SGNYQPSAW	RLADDLGTAPFE	FVRDRV	TTIDDDYDVYG
OsaTPS34	b	R-VGNYQPSTW	RLADDLGTQFE	FVRDRI	TVIDDDYDVYG
OsaTPS35	c	SLEG-MSGLTW	DKLRMFQRRKE-	LSERNL	QVKDIDDTAMG
OsaTPS36	c	SLEG-MQGLVW	HQLRLFKNRKVR	LSERSL	NVQDIDDTAMG
OsaTPS37	c	SLEG-LENLDW	QQLSQIHNEKE-	ISRKEL	EFCDIDDTSMG
OsaTPS38	c	SLEG-LDDLW	HHLSQIPNQKV-	ISRKDL	NFSDLDDTSMG
OsaTPS39	c	ILEGLMENIKW	N---QLRNQKM-	ISKNEI	TFCDVDTSMG
OsaTPS40	e	IAEGMGKLNW	RLLNDIRSYERE	FARSKS	LVIDDFDTYG
OsaTPS41	e	QCRRRIFPVVE	RLLNDVQTFEKE	YGRDVQ	LVIDDFDTYG
OsaTPS42	f	YLETLPSTYEF	RLANDMQSYQKE	FGREKT	YVIDDFDTYG
OsaTPS43	g	-----	RLSDDLGTAAQ--	GARSQP	GIMDDTYDNYA
OsaTPS44	g	-----	RLLDDLGSQAQDE	SSRNQP	SVLDDIYDVYG
OsaTPS45	g	INRDNLLKNT	RLWDDLGTAKMD	FGRDRP	SVIDDTYDSHG
OsaTPS46	g	FNRD-----	RLWDDLGTAK--	FGRDRP	SVIDDTYDSHG
OsaTPS47	g	-----	RLWDDLGTSEQEE	FGRDRA	TIIDDIYDVYG

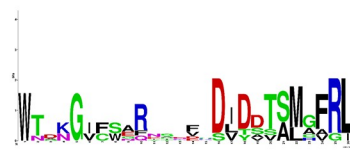
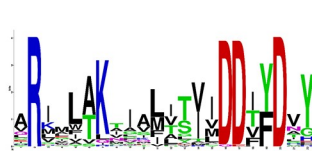


Fig 7. Conserved motifs of the TPS gene family presented in the *O. sanctum* putative functional TPS genes were shown. All the conserved amino acids were highlighted in the grey shades. 47 *OsaTPS* were shown and classified into TPS gene subfamily. Here RR(X)₈W conserved motif is conserved in *OsaTPS*-b subfamily except for *OsaTPS19* (angiosperm monoterpene synthase). Variation of the RR(X)₈W motif is found in *OsaTPS*-a subfamily of sesquiterpene and diterpene synthase that have putative N-terminal plastid peptides, *OsaTPS*-c and -e lack twin-arginine residues but consist only tryptophan residue. NSE/DTE motifs were continuously seen conserved except *OsaTPS*-c subfamily. Motif RXR in all *OsaTPS* are shown genes either in same or modified form except some *OsaTPS*. Motif DDXXD and DXDD were shown highly conserved in all putative functional *O. sanctum* TPS genes, but in two genes, i.e., *OsaTPS6* and *OsaTPS25* DDXXD motif is absent, which might be due to sequence assembly error.

<https://doi.org/10.1371/journal.pone.0207097.g007>

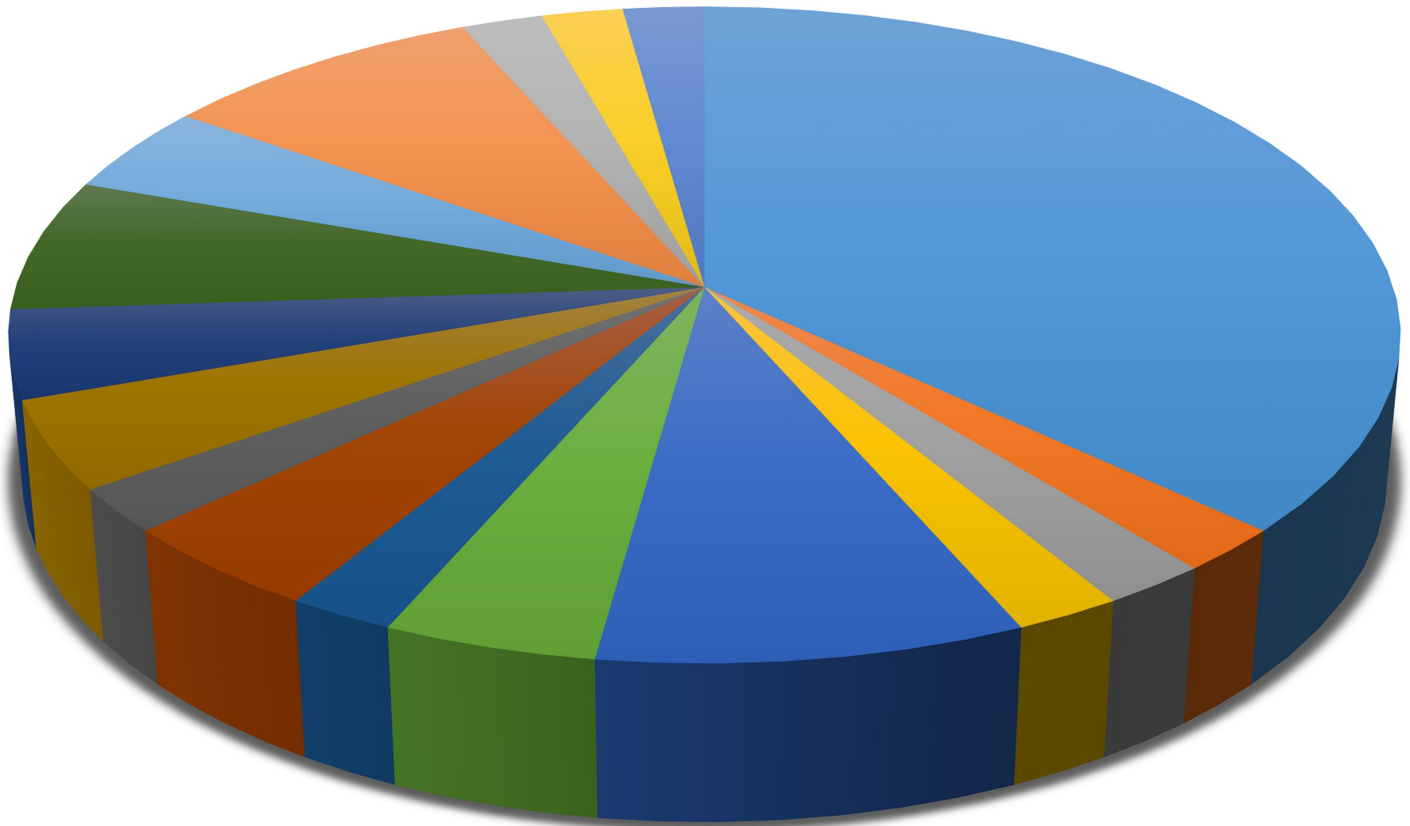
the sequences have variation of this motif as R(D/H)(R/K/D/V/Q) and tree sequences were shown missing this motif due to sequencing error. The additional metal cofactor binding motif NSE/DTE, is less conserved in TPS, evolved from a second aspartate-rich region in prenyl transferases to form a consensus sequence of (L,V)(V,L,A)(N,D)D(L,I,V)x(S,T)xxxE [66]. In our prediction motif: NSE/DTE form 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)xxx(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), Kyoto 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 monoterpene biosynthesis pathway, total 15 *OsaTPS* genes were found, which may be involved in the synthesis of the particular secondary metabolite. Similarly, in the diterpene pathway, 8 *OsaTPS* genes show their presence, whereas the maximum 23 *OsaTPS* genes were observed in the sesquiterpene biosynthesis pathway (Fig 9). *OsaTPS19* shows homology with the isoprene synthase gene, which is used to synthesize the isoprene from dimethylallyl diphosphate.



- | | | |
|----------------------|--------------------------|------------------------|
| ■ Ocimum basilicum | ■ Solanum lycopersicum | ■ Solanum tuberosum |
| ■ Origanum vulgare | ■ Pogostemon cablin | ■ Phyla dulcis |
| ■ Hyoscyamus muticus | ■ Mentha x piperita | ■ Populus tremuloides |
| ■ Ricinus communis | ■ Lavandula angustifolia | ■ Salvia officinalis |
| ■ Pisum sativum | ■ Nicotiana tabacum | ■ Arabidopsis thaliana |
| ■ Antirrhinum majus | ■ Fragaria x ananassa | |

Fig 8. Annotation result of 47 *OsaTPS* protein sequences, show higher sequence homology with *Ocimum basilicum* followed by *Nicotiana tabacum*, *Pogostemon cablinsesamum* and *Salvia officinalis* species. The pie chart represents the sequence similarity matching of *OsaTPS* with different plant species.

<https://doi.org/10.1371/journal.pone.0207097.g008>

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

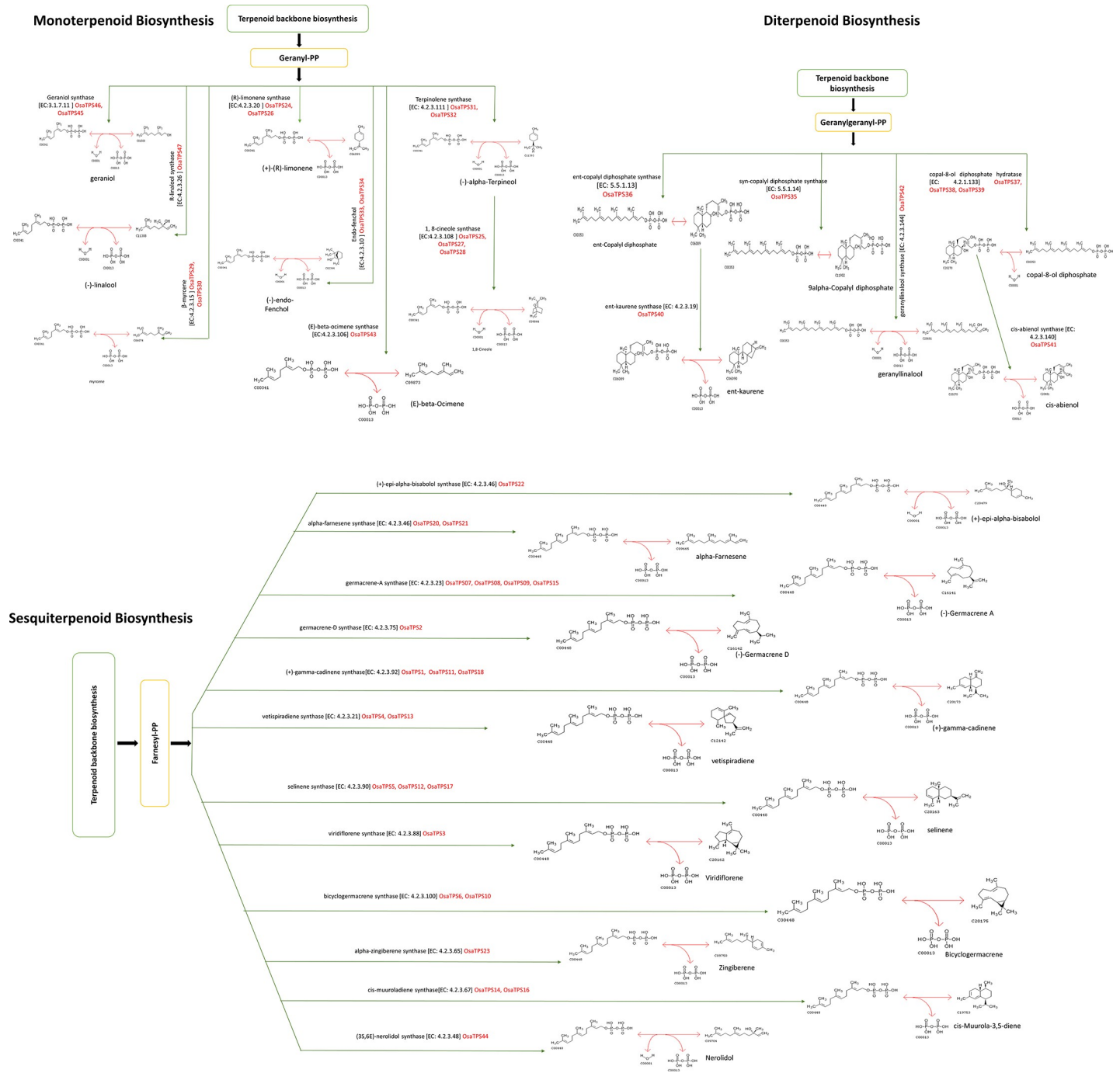


Fig 9. Proposed secondary metabolite synthesis pathway was constructed for *O. sanctum*. Monoterpenoid, Diterpenoid and Sesquiterpenoid pathways show the presence of putatively functional *OsaTPS* genes.

<https://doi.org/10.1371/journal.pone.0207097.g009>

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 synthase 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 synthase 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 *OsaTPS1* to *OsaTPS18*. This clade is composed explicitly of angiosperm sesquiterpene and diterpene synthases and contains conserved DDxxD motif in the c-terminal domain of TPS which is supposed to combining metal ions, with less conserved RR(X)₈W 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)₈W 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., cadinen synthase, Germacrene-D-synthase, Viridiflorene synthase, Selinene synthase, Bi-cyclogermacrene synthase and Cis-muroladiene synthase. The previous subfamily classification suggests that TPS-b contains conserved RR(X)₈W belongs to monoterpene synthases and synthesizes angiosperm monoterpenoids, while in *O. sanctum* it was seen that *OsaTPS20* to *OsaTPS28* and *OsaTPS31* genes contain highly conserved RR(X)₈W motif which confirms the predicted *OsaTPS* genes to be monoterpene 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 *OsaTPS19* to *OsaTPS34* 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 *OsaTPS35* to *OsaTPS39*, these TPS genes contain only Dx(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 be 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-e/f contains all valid kaurene synthase proteins from angiosperms which encode and annotate *OsaTPS40*, the presence of DDxxD motif but not DxDD 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 RR(X)₈W 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 know as terpene synthases. In this study, we discovered a large number of terpene synthase genes in *O. 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. *OsaTPS46* and *OsaTPS45* mapped with geraniol synthase which synthesises geraniol, also reported in *Ocimum basilicum* species. Similarly we identified R-linalool synthase (*OsaTPS47*), (-)-Endo-fenchol synthase (*OsaTPS33*, and *OsaTPS34*), Selinene synthase (*OsaTPS5*, *OsaTPS12* and *OsaTPS17*), Terpinolene synthase (*OsaTPS31*, *OsaTPS32*), β -myrcene synthase (*OsaTPS29* and *OsaTPS30*), α -zingiberene synthase (*OsaTPS23*), γ -cadinene synthase (*OsaTPS1*, *OsaTPS11* and *OsaTPS18*), and Germacrene-D synthase (*OsaTPS2*). 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 (*OsaTPS36*), Syn-copalyl diphosphate synthase (*OsaTPS35*), Ent-kaurene synthase (*OsaTPS40*), Copal-8-ol diphosphate synthase (*OsaTPS37*, *OsaTPS38*, and *OsaTPS39*), and Cis-abienol synthase (*OsaTPS41*), these all diterpene synthase genes were still unexplored in this species. Similarly in sesquiterpenoid biosynthesis *OsaTPS* genes mapped with (+)-Epi-alpha bisabolol synthase (*OsaTPS22*), α -Farnesene synthase (*OsaTPS20* and *OsaTPS21*), Germacrene-A synthase (*OsaTPS07*, *OsaTPS08*, *OsaTPS09* and *OsaTPS15*), Vetivipadiene synthase (*OsaTPS4*, *OsaTPS13*), Viridiflorene synthase (*OsaTPS3*), Bicyclogerma-crene synthase (*OsaTPS6* and *OsaTPS10*), Cis-muurooladiene synthase (*OsaTPS14* and *OsaTPS16*) and (3S, 6E)-Nerolidol synthase (*OsaTPS44*) (Fig 9).

The chemical composition of *O. 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

structure, phylogenetic subfamily classification, motif localization and enzyme pathway mapping. 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* TPSs 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 $\geq 80\%$ were designated in red color triangles and values supporting $\geq 95\%$ 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> 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 BIOSYNTHESIS IN PLANTS. *Annu Rev Plant Physiol Plant Mol Biol*. 1999; 50: 47–65. <https://doi.org/10.1146/annurev.arplant.50.1.47> PMID: 15012203
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. Külheim 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)- α -bisabolene synthase from grand fir (*Abies*). *Natl Acad Sci*. Available: <http://www.pnas.org/content/95/12/6756.short>
9. Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc Natl Acad Sci*. 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: 21443633
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, Hudock MP, et al. Diterpene cyclases and the nature of the isoprene fold. *Proteins Struct Funct Bioinforma*. 2010; 78: 2417–2432. <https://doi.org/10.1002/prot.22751> PMID: 20602361
14. Aubourg S, Lecharny 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, Schouwey 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, Schouwey 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, Schauvinhold I, et al. The Tomato Terpene Synthase Gene Family. *PLANT Physiol*. 2011; 157: 770–789. <https://doi.org/10.1104/pp.111.179648> PMID: 21813655
18. Harris LJ, Saparno A, Johnston A, Pristic 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, Bharti 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.2022> 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, Bhattacharya 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 Phenylpropenes 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-arplant-043014-114639> PMID: 25621517
28. Ringer KL, Davis EM, Croteau R. Monoterpene Metabolism. Cloning, Expression, and Characterization of (-)-Isopiperitenol/(-)-Carveol Dehydrogenase of Peppermint and Spearmint. *PLANT Physiol.* 2005; 137: 863–872. <https://doi.org/10.1104/pp.104.053298> PMID: 15734920
29. Crock J, Wildung M, Croteau R. Isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (*Mentha x piperita*, L.) that produces the aphid alarm pheromone (E)-beta-farnesene. *Proc Natl Acad Sci U S A.* 1997; 94: 12833–8. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9371761> PMID: 9371761
30. Prosser IM, Adams RJ, Beale MH, Hawkins ND, Phillips AL, Pickett JA, et al. Cloning and functional characterisation of a cis-muroladiene synthase from black peppermint (*Mentha x piperita*) and direct evidence for a chemotype unable to synthesise farnesene. *Phytochemistry.* 2006; 67: 1564–1571. <https://doi.org/10.1016/j.phytochem.2005.06.012> PMID: 16083926
31. 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.* 2006; 454: 123–136. <https://doi.org/10.1016/j.abb.2006.08.006> PMID: 16970904
32. Croteau R, Munck SL, Akoh CC, Fisk HJ, Satterwhite DM. Biosynthesis of the sesquiterpene patchoulol from farnesyl pyrophosphate in leaf extracts of *Pogostemon cablin* (patchouli): mechanistic considerations. *Arch Biochem Biophys.* 1987; 256: 56–68. Available: <http://www.ncbi.nlm.nih.gov/pubmed/3038029> PMID: 3038029
33. Crowell AL, Williams DC, Davis EM, Wildung MR, Croteau R. Molecular cloning and characterization of a new linalool synthase. *Arch Biochem Biophys.* 2002; 405: 112–21. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12176064> PMID: 12176064
34. Gao W, Hillwig M, Huang L, Cui G, . . . XW-O, 2009 undefined. A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *ACS Publ.* Available: <https://pubs.acs.org/doi/abs/10.1021/ol902051v>
35. Cui G, Duan L, Jin B, Qian J, Xue Z, . . . GS-P, 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>
36. 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* gibberellin biosynthetic. *nature.com.* Available: <https://www.nature.com/articles/srep23057> <https://doi.org/10.1038/srep23057> PMID: 26971881
37. Li J, Chen Q, Jin Q, Gao J, Zhao P, Lu S, et al. *leCPS2* is potentially involved in the biosynthesis of pharmacologically active *Isodon* diterpenoids rather than gibberellin. Elsevier. Available: <https://www.sciencedirect.com/science/article/pii/S003194221100611X>
38. Jin B, Cui G, Guo J, Tang J, Duan L, . . . HL-P, 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>
39. Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, Zerbe P, et al. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolol, 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. Božić D, Papaefthimiou D, Brückner K, de Vos RCH, Tsoleridis 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* clerodienyl diphosphate synthase. Wiley Online Libr. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.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: 587–603. <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; 134: 370–379. <https://doi.org/10.1104/pp.103.032946> 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 Phenylpropenes 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, Coghill 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, Difazio S, Jansson S, Bohlmann 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. Külheim 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, Filipowski 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 “incomparable” & “elixir of life” 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, Bohlmann 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.2011.04520.x>