Mining and Bioinformatics Analysis of the FPPS Gene in the Dysphania schraderiana Transcriptome Database (Postprint)

Authors: Fu Suhong, Lei Ming, Zhang Yongqun, Shi Jing, bean

Date: 2018-07-18T00:00:00+00:00

Abstract

The essential oil of Dysphania schraderiana exhibits potential medicinal value and is rich in sesquiterpenoid compounds, with farnesyl pyrophosphate synthase (FPPS) serving as the key enzyme at the biosynthetic branch point. To gain deeper insights into the farnesyl pyrophosphate synthase genes of Dysphania schraderiana, this study mined the transcriptome database of Dysphania schraderiana, obtaining two FPPS gene sequences (DsFPPS1 and DsFPPS2), and conducted comprehensive analyses of the physicochemical properties, structure, function, and phylogeny of the proteins encoded by DsFPPS1 and DsFPPS2. The results demonstrated that the DsFPPS1 and DsFPPS2 genes contain open reading frames of 1029 bp and 969 bp, respectively, encoding 342 and 322 amino acids. DsFPPS1 and DsFPPS2 are localized in mitochondria, with no detectable signal peptides or transmembrane structures; DsFPPS1 is a stable protein, whereas DsFPPS2 is an unstable protein. Amino acid sequence alignment revealed that DsFPPS1 and DsFPPS2 share 60.53% sequence similarity, both possessing five conserved domains and two aspartate-rich regions. The secondary structures of DsFPPS1 and DsFPPS2 are predominantly composed of -helices, with their tertiary structures forming an -helix bundle comprising eight -helices; however, the DsFPPS2 tertiary structure lacks one -helix A. In the phylogenetic tree, DsFPPS1 clusters with Chenopodiaceae plants, exhibiting close genetic distance to them, while DsFPPS2 forms a separate clade. This study, through mining and bioinformatics analysis of FPPS genes from the Dysphania schraderiana transcriptome database, provides a theoretical foundation for functional studies of Dysphania schraderiana FPPS and research on the biosynthesis of its sesquiterpenoid compounds.

Full Text

Mining and Bioinformatic Analysis of FPPS Genes from *Dysphania schraderiana* Transcriptome Database

FU Suhong[1], LEI Ming[2], ZHANG Yongqun[1]*, SHI Jing[1], HAO Doudou[1]
(1. Molecular Medical Laboratory, Hospital of Chengdu Office of People' s Government of Tibet Autonomous Region, Chengdu 610041, China;
2. School of Science, Tibet University, Lhasa 850000, China)

Abstract: The essential oil of *Dysphania schraderiana* (Chenopodiaceae) exhibits potential medicinal value and is rich in sesquiterpenoid compounds. Farnesyl pyrophosphate synthase (FPPS) represents the key branch-point enzyme in their biosynthetic pathway. To gain deeper insights into *D. schraderiana* FPPS genes, we mined the transcriptome database of this species, obtaining two FPPS gene sequences (DsFPPS1 and DsFPPS2), and subsequently analyzed the physicochemical properties, structural features, functions, and phylogenetic relationships of the encoded proteins. The results demonstrated that DsFPPS1 and DsFPPS2 contain open reading frames of 1,029 bp and 969 bp, encoding 342 and 322 amino acids, respectively. Both proteins were predicted to localize in mitochondria, lacking signal peptides and transmembrane domains. DsFPPS1 was identified as a stable protein, whereas DsFPPS2 was predicted to be unstable. Amino acid sequence alignment revealed 60.53% similarity between DsFPPS1 and DsFPPS2, with both proteins containing five conserved domains and two aspartate-rich motifs. The secondary structures of both proteins were predominantly composed of -helices, with their tertiary structures forming an -helix bundle comprising eight -helices; however, DsFPPS2 lacked one -helix in its tertiary structure. Phylogenetic analysis showed that DsFPPS1 clustered with Chenopodiaceae plants, indicating close genetic relationships, while DsFPPS2 formed a separate, independent clade. This study provides a theoretical foundation for functional investigation of *D. schraderiana* FPPS and for research on sesquiterpenoid biosynthesis in this species.

Keywords: farnesyl pyrophosphate synthase; transcriptome; *Dysphania schraderiana*; gene mining; bioinformatics

Introduction

Dysphania schraderiana, a member of the Chenopodiaceae family, is a traditional medicinal plant whose whole herb is used to treat asthma, spasms, toxicity, and skin conditions, with additional effects as an analgesic and antipruritic. Modern research has revealed broad biological activities in its essential oil. Our previous studies demonstrated significant concentration-dependent inhibition of *Escherichia coli* and *Bacillus subtilis* growth by *D. schraderiana* essential oil. Additionally, Lei (2015) reported concentration-dependent inhibition of the insect *Sitophilus zeamais*, and a patent has been filed for using this essential oil

to control plant mites (Liu et al., 2015). GC-MS analysis in our earlier work identified sesquiterpenoids as a major component (26.846%) of the essential oil, second only to aliphatic compounds (54.232%).

Sesquiterpenoids are a class of natural compounds containing 15 carbon atoms (three isoprene units) with diverse structural types, representing a primary source for new drug development (Wang, 2012). Plant terpenoids are synthesized primarily through the mevalonate (MVA) pathway in the cytoplasm and the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway in plastids. Both pathways converge to produce the common precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). One molecule of IPP then condenses with DMAPP to form geranyl diphosphate (GPP), which subsequently combines with another IPP molecule under the catalysis of farnesyl diphosphate synthase (FPPS) to produce farnesyl diphosphate (FPP)—the direct precursor for diverse sesquiterpenoid compounds (Sun, 2017).

Plant FPPS enzymes are encoded by gene families containing multiple isoforms. For example, *Arabidopsis thaliana* possesses two FPPS genes (FPPS1 and FPPS2) that encode three isoforms: FPPS1S, FPPS1L, and FPPS2. Simultaneous knockout of both genes is lethal, while single-gene knockouts do not affect plant development, suggesting that this multi-gene architecture ensures survival when one gene is compromised. Different isoforms exhibit distinct expression patterns and activities: FPPS1 is expressed throughout the plant lifecycle, whereas FPPS2 functions primarily during seed development (Closa, 2010). Moreover, FPPS2 demonstrates higher catalytic activity and thermostability, along with greater sensitivity to sodium chloride inhibition (Keim, 2012).

Given the critical role of FPPS as a branch-point enzyme in terpenoid biosynthesis and the absence of prior research on *D. schraderiana* FPPS genes, this study aimed to mine the flower and leaf transcriptome database of *D. schraderiana* to identify FPPS genes (DsFPPS) and conduct comprehensive bioinformatic analyses of the encoded proteins' physicochemical properties, subcellular localization, structural characteristics, functions, and evolutionary relationships, thereby establishing a theoretical basis for future structural and functional studies.

Materials and Methods

1.1 Materials

The two FPPS genes from *D. schraderiana* were obtained from our previous transcriptome study. The flower and leaf transcriptome datasets are deposited in the SRA database under accession numbers SRX3145242 and SRX3145241, respectively. FPPS protein sequences from other species were retrieved from the NCBI database, including: *Chenopodium quinoa* (CqFPPS: XP_021740746.1), *Spinacia oleracea* (SoFPPS: XP_021846881.1), *Beta vulgaris* (BvFPPS: XP_010675977.1), *Camellia sinensis* (CsFPPS: ANA11766.1), *Mangifera*

indica (MiFPPS: AFJ52720.1), *Ricinus communis* (RcFPPS: AMN82836.1), *Hevea brasiliensis* (HbFPPS: ANJ77846.1), *Euphorbia pekinensis* (EpFPPS: ACN63187.1), *Panax ginseng* (PgFPPS: AAY87903.1), *Panax japonicus* (PjFPPS: AKN52395.1), *Panax notoginseng* (PnFPPS: AGS79228.1), *Hedera helix* (HhFPPS: APV45530.1), *Panax quinquefolius* (PqFPPS: ADJ68004.1), *Mentha × piperita* (MpFPPS: AAK63847.1), *Salvia officinalis* (SolFPPS: AQY54371.1), *Salvia miltiorrhiza* (SmFPPS: ABV08819.1), *Lavandula angustifolia* (LaFPPS: AGQ04160.1), *Leucosceptrum canum* (LcFPPS: ALT07952.1), *Astragalus membranaceus* (AmFPPS: AID51444.1), *Glycyrrhiza uralensis* (GuFPPS: ADE18770.1), *Medicago sativa* (MsFPPS: ADC32809.1), *Trichoderma reesei* (TrFPPS: AFX82678.1), *Emmonsia crescens* (EcFPPS: POS87336.1), *Bacillus licheniformis* (BlFPPS: ARC74194.1), *Rhodobacter capsulatus* (RHcFPPS: ADE44162.1), *Bos taurus* (BtFPPS: AAL58886.1), and *Pan troglodytes* (PtFPPS: JAA28793.1).

1.2.1 Protein Basic Property Analysis

Open reading frames (ORFs) were identified using the ORF Finder tool from GenBank (https://www.ncbi.nlm.nih.gov/orffinder/). The encoded amino acid sequences were subsequently analyzed using multiple online platforms: physicochemical properties were assessed with ProtParam (http://web.expasy.org/protparam/); hydrophilicity/hydrophobicity profiles were generated using ProtScale (http://web.expasy.org/protscale/); signal peptide prediction was performed with SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/); transmembrane domains were identified using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/); and subcellular localization was predicted with Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/).

1.2.2 Protein Structural and Functional Analysis

Secondary structures were predicted using the SOPMA online server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). Tertiary structures were modeled using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) with homology threading. Conserved domains were analyzed using the NCBI Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/cdd/).

1.2.3 Protein Sequence Alignment and Phylogenetic Tree Construction

Homologous FPPS amino acid sequences were retrieved using BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignments were performed with ClustalW. Phylogenetic trees were constructed using the Neighbor-Joining method with 1,000 bootstrap replicates in MEGA 7.0.14.

Results

2.1 Mining of FPPS Genes from *D. schraderiana*

Through sequence alignment and functional annotation of nucleotide sequences from the *D. schraderiana* transcriptome, two candidate FPPS genes designated c13053_g1 and c12747_g1 were identified and renamed DsFPPS1 and DsFPPS2. DsFPPS1 spans 1,257 bp, with its complementary strand serving as the coding template. The ORF comprises 1,029 bp located at positions 1,185–155 bp, encoding a 342-amino-acid protein designated DsFPPS1. DsFPPS2 spans 1,667 bp and contains a 969-bp ORF from positions 296–1,264 bp, encoding a 322-amino-acid protein designated DsFPPS2 (Figure 1 [Figure 1: see original paper]).

2.2 Physicochemical Properties of DsFPPS Proteins

The physicochemical properties of DsFPPS1, DsFPPS2, and three Chenopodiaceae FPPS proteins (*C. quinoa* CqFPPS, *S. oleracea* SoFPPS, and *B. vulgaris* BvFPPS) were analyzed using ProtParam (Table 1). DsFPPS1 closely resembled other Chenopodiaceae FPPS proteins, with 342 amino acids, molecular weights of 39.67–39.74 kDa, isoelectric points (pI) of 5.01–5.12, and similar amino acid composition: acidic (15.5–16.4%), basic (13.8–14.7%), polar neutral (25.3–26.5%), and nonpolar (42.4–44.4%) residues. All Chenopodiaceae proteins exhibited instability indices below 40, indicating stability. In contrast, DsFPPS2 differed substantially: 322 amino acids, molecular weight of 36.76 kDa, pI of 5.65, lower acidic amino acid content (13.1%), higher nonpolar content (46.5%), and an instability index exceeding 40, classifying it as unstable.

Hydrophilicity/hydrophobicity profiles were generated using ProtScale, where positive scores indicate hydrophobicity and negative scores indicate hydrophilicity. DsFPPS1 showed maximum hydrophobicity at position 202 (score: 2.611) and maximum hydrophilicity at position 101 (score: -2.733). DsFPPS2 exhibited maximum hydrophobicity at position 68 (score: 2.800) and maximum hydrophilicity at position 118 (score: -2.578) (Figure 2 [Figure 2: see original paper]).

2.3 Signal Peptide, Transmembrane Domain, and Subcellular Localization Analysis

Signal peptides are short N-terminal sequences (15–30 residues) that target secreted proteins and are cleaved post-translationally. SignalP 4.1 analysis revealed no signal peptides in either DsFPPS1 or DsFPPS2, indicating they are non-secreted proteins (Figure 3 [Figure 3: see original paper]A and 3B). TMHMM analysis further confirmed the absence of transmembrane domains in both proteins (Figure 4 [Figure 4: see original paper]A and 4B). Subcellular localization prediction using Cell-PLoc 2.0 indicated mitochondrial localization for both DsFPPS1 and DsFPPS2.

2.4 Functional Domain Analysis of DsFPPS Proteins

Conserved domains were identified using NCBI CDD, ClustalW alignment, and literature data (Chen, 1994; Guo, 2015; Li, 2012). FPPS family proteins contain five conserved regions (I–V) and two aspartate-rich motifs (DDXXD, where X represents any amino acid): the First Aspartic Rich Motif (FARM) and Second Aspartic Rich Motif (SARM) (Figure 5 [Figure 5: see original paper]). DsFPPS1 and DsFPPS2 share 60.53% sequence identity and both possess the characteristic FARM and SARM motifs and conserved regions I–V. These aspartate-rich motifs serve as Mg^{2} binding sites that coordinate substrate phosphate groups for catalysis (Christianson, 2017). In DsFPPS1, the five conserved regions are: GGKLNR (45–50, I), EWLQAYFLVLDDIMDNSHTRRG (83–104, II), GQMIDL (160–165, III), KT (190–191, IV), and GIYFQVQDDYLDCFGDPEFIGKIGTDIEDFK (225–255, V), with FARM = DDIMD (93–97) and SARM = DDYLD (232–236). In DsFPPS2, the corresponding regions are: SRKLNR (25–30, I), QWLVGCVLVLDDLLDASHTRRG (63–84, II), GEMIDL (140–145, III), KT (170–171, IV), and GIYFQAQDDYLDCFGDPKKSGKIGSDIEDFK (205–235, V), with FARM = DDLLD (73–77) and SARM = DDYLD (212–216).

2.5 Structural Characteristics of DsFPPS Proteins

Secondary structure prediction using SOPMA revealed similar compositions for both proteins, predominantly -helices (59.36% in DsFPPS1; 64.29% in DsFPPS2) (Figure 6 [Figure 6: see original paper]). DsFPPS1 also contained 19.88% random coils, 11.70% extended strands, and 9.06% -turns, while DsFPPS2 comprised 15.53% random coils, 10.87% extended strands, and 9.32% -turns.

Tertiary structures were modeled using Phyre2 with *Artemisia spiciformis* FPPS (PDB ID: 4KK2) as template, yielding 79% and 63% similarity for DsFPPS1 and DsFPPS2, respectively (Figure 7 [Figure 7: see original paper]). DsFPPS1 contains 13 -helices: 10 long helices (A–J) forming a hairpin structure between helices A and B, with helices C–J creating an -helical bundle surrounding a central cavity, plus three small helices (-1, -2, -3) inserted between helices H and I. The catalytic reaction occurs at the cavity base. DsFPPS2 exhibits a similar architecture but contains only 12 -helices (9 long helices B–J and 3 small helices), lacking helix A and consequently the A–B hairpin structure.

2.6 Phylogenetic Analysis of DsFPPS Proteins

A phylogenetic tree was constructed using MEGA 7.0.14 with DsFPPS1, DsFPPS2, and 24 additional FPPS protein sequences from GenBank. The Neighbor-Joining analysis revealed four major clades: plants, fungi, animals, and bacteria, demonstrating clear taxonomic specificity (Figure 8 [Figure 8: see original paper]). Bacteria formed a separate major branch, while animals, fungi, and plants clustered together. Within the plant clade, Lamiaceae, Araliaceae, Chenopodiaceae, Iridaceae, and Euphorbiaceae each formed distinct subclades. DsFPPS1 grouped within the Chenopodiaceae clade, indicating close phylogenetic rela-

tionships, whereas DsFPPS2 formed an independent, separate branch.

Discussion and Conclusion

Sesquiterpenoids, as plant secondary metabolites, exhibit strong biological activities (Piao, 2012). Notable examples include artemisinin (C H O), a sesquiterpene lactone and potent antimalarial drug (Tu, 2011); germacrone (C H O) with anti-ulcer, anti-tumor, anti-inflammatory, antibacterial, and antitussive properties (Wu, 2017); -eudesmol (C H O) which stimulates appetite via TRPA1 and the autonomic nervous system (Ohara, 2017); cedrol (C H O) that dose-dependently promotes fibroblast growth and collagen synthesis (Mu, 2012); and valerenic acid (C H O) which exerts anxiolytic effects through allosteric modulation of GABA-A receptors (Becker, 2014).

FPPS serves as a critical branch-point enzyme in plant terpenoid biosynthesis, catalyzing the formation of the sesquiterpenoid precursor FPP. The biosynthesis of sesquiterpenoids in *D. schraderiana* essential oil requires this enzyme. Leveraging transcriptomic data, we mined two FPPS genes (DsFPPS1 and DsFPPS2) from the *D. schraderiana* transcriptome. Bioinformatic analysis revealed that DsFPPS1 and DsFPPS2 share 60.53% sequence identity and both contain the characteristic five conserved domains and two aspartate-rich motifs (FARM and SARM) of FPPS proteins. However, they differ in stability, with DsFPPS2 being unstable and lacking -helix A, which eliminates the A–B hairpin structure that may affect protein stability. Phylogenetically, DsFPPS2 forms a separate branch within the plant clade, suggesting it may represent a novel FPPS isoform.

Plant FPPS enzymes are typically encoded by multi-gene families with multiple isoforms that function in different tissues and developmental stages. In *Arabidopsis*, loss of FPPS2 in seeds increases HMGR (3-hydroxy-3-methylglutaryl-CoA reductase) activity (Closa, 2010), thereby regulating carbon flux and controlling isoprenoid composition. Our previous transcriptome analysis indicated that DsFPPS1 expression exceeds DsFPPS2 in both flower and leaf tissues, and DsFPPS1 exhibits greater stability. We hypothesize that DsFPPS1 is constitutively and abundantly expressed throughout *D. schraderiana*, while DsFPPS2 expression may be temporally and spatially specific. However, the precise functions, activities, expression patterns, and regulatory mechanisms of each *D. schraderiana* FPPS isoform require further investigation. In summary, this study provides a comprehensive mining and bioinformatic analysis of FPPS genes from the *D. schraderiana* transcriptome, establishing a theoretical foundation for future basic research and applied development of these genes.

References

BECKER A, FELGENTREFF F, SCHRÖDER H, et al., 2014. The anxiolytic effects of a Valerian extract is based on valerenic acid [J]. *BMC Complementary and Alternative Medicine*, 14(1): 1-5.

CHEN A, KROON PA, POULTER CD, 1994. Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure [J]. *Protein Science*, 3(4): 600-607.

CHRISTIANSON DW, 2017. Structural and chemical biology of terpenoid cyclases [J]. *Chemical Reviews*, 117(17): 11567-11601.

CLOSA M, VRANOVÁ E, BORTOLOTTI C, et al., 2010. The *Arabidopsis thaliana* FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest [J]. *The Plant Journal*, 63(3): 512-525.

GUO D, LI HL, PENG SQ, 2015. Structure conservation and differential expression of farnesyl diphosphate synthase genes in Euphorbiaceous plants [J]. *International Journal of Molecular Sciences*, 16(9): 22402-22414.

KEIM V, MANZANO D, FERNÁNDEZ FJ, et al., 2012. Characterization of *Arabidopsis* FPS isozymes and FPS gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds [J]. *PLoS ONE*, 7(11): e49109.

LEI M, HE H, ZHANG PF, et al., 2015. Study on the extraction of *Chenopodium foetidum* essential oil and its inhibition against insect activity [J]. *Journal of Anhui Agricultural Sciences*, 43(28): 64-66. [in Chinese]

LI YB, FAN QQ, WANG BL, et al., 2012. Advances in the study of plant farnesyl pyrophosphate synthase (FPPS) gene [J]. *Chinese Journal of Agricultural Biotechnology*, 20(3): 321-330. [in Chinese]

MU HJ, SUN GP, HWANG YL, et al., 2012. Cedrol enhances extracellular matrix production in dermal fibroblasts in a MAPK-dependent manner [J]. *Annals of Dermatology*, 24(1): 16-21.

OHARA K, FUKUDA T, ISHIDA Y, et al., 2017. -Eudesmol, an oxygenized sesquiterpene, stimulates appetite via TRPA1 and the autonomic nervous system [J]. *Scientific Reports*, 7(1): 15785-15801.

PIAO YH, PIAO HS, 2012. Research progress in biological activity of sesquiterpene compounds [J]. *Occupation and Health*, 28(18): 2291-2293. [in Chinese]

SUN LC, LI SY, WANG FZ, et al., 2017. Research progresses in the synthetic biology of terpenoids [J]. *Biotechnology Bulletin*, 33(1): 64-75. [in Chinese]

TU YY, 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine [J]. *Nature Medicine*, 17(10): 1217-1220.

WANG J, YOU S, ZHOU LN, 2012. Progress in research of sesquiterpenoids

biotransformation [J]. *Journal of Shenyang Pharmaceutical University*, 29(2): 76-84. [in Chinese]

WU J, FENG Y, HAN C, et al., 2017. Germacrone derivatives: synthesis, biological activity, molecular docking studies and molecular dynamics simulations [J]. *Oncotarget*, 8(9): 15149-15158.

LIU ZL, LIU XC, SHI WP, et al., 2015-07-22. Essential oil of *Dysphania schraderiana* in the prevention and control of plant mites [P]. China Patent: CN104782667A. [in Chinese]

Note: Figure translations are in progress. See original paper for figures.