BMC Plant Biology

RESEARCH ARTICLE   **Open Access**

CrossMark

Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*)

Avinash Pandreka[1,2†], Devdutta S. Dandekar[1†], Saikat Haldar[1†], Vairagkar Uttara[1], Shinde G. Vijayshree[1], Fayaj A. Mulani[1], Thiagarayaselvam Aarthy[1] and Hirekodathakallu V. Thulasiram[1,2*] (iD)

Abstract

Background: Neem tree (*Azadirachta indica*) is one of the richest sources of skeletally diverse triterpenoids and they are well-known for their broad-spectrum pharmacological and insecticidal properties. However, the abundance of Neem triterpenoids varies among the tissues. Here, we delineate quantitative profiling of fifteen major triterpenoids across various tissues including developmental stages of kernel and pericarp, flower, leaf, stem and bark using UPLC-ESI (+)-HRMS based profiling. Transcriptome analysis was used to identify the initial genes involved in isoprenoid biosynthesis. Based on transcriptome analysis, two short-chain prenyltransferases and squalene synthase (AiSQS) were cloned and functionally characterized.

Results: Quantitative profiling revealed differential abundance of both total and individual triterpenoid content across various tissues. RNA from tissues with high triterpenoid content (fruit, flower and leaf) were pooled to generate 79.08 million paired-end reads using Illumina GA II platform. 41,140 transcripts were generated by *de novo* assembly. Transcriptome annotation led to the identification of the putative genes involved in isoprenoid biosynthesis. Two short-chain prenyltransferases, geranyl diphosphate synthase (AiGDS) and farnesyl diphosphate synthase (AiFDS) and squalene synthase (AiSQS) were cloned and functionally characterized using transcriptome data. RT-PCR studies indicated five-fold and ten-fold higher relative expression level of AiSQS in fruits as compared to leaves and flowers, respectively.

Conclusions: Triterpenoid profiling indicated that there is tissue specific variation in their abundance. The mature seed kernel and initial stages of pericarp were found to contain the highest amount of limonoids. Furthermore, a wide diversity of triterpenoids, especially C-seco triterpenoids were observed in kernel as compared to the other tissues. Pericarp, flower and leaf contained mainly ring-intact triterpenoids. The initial genes such as AiGDS, AiFDS and AiSQS involved in the isoprenoids biosynthesis have been functionally characterized. The expression levels of AiFDS and AiSQS were found to be in correlation with the total triterpenoid content in individual tissues.

Keywords: *Azadirachta indica*, Triterpenoids, Quantitative profiling, Transcriptome

* Correspondence: hv.thulasiram@ncl.res.in
†Equal contributors
[1]Chemical Biology Unit, Division of Organic Chemistry, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India
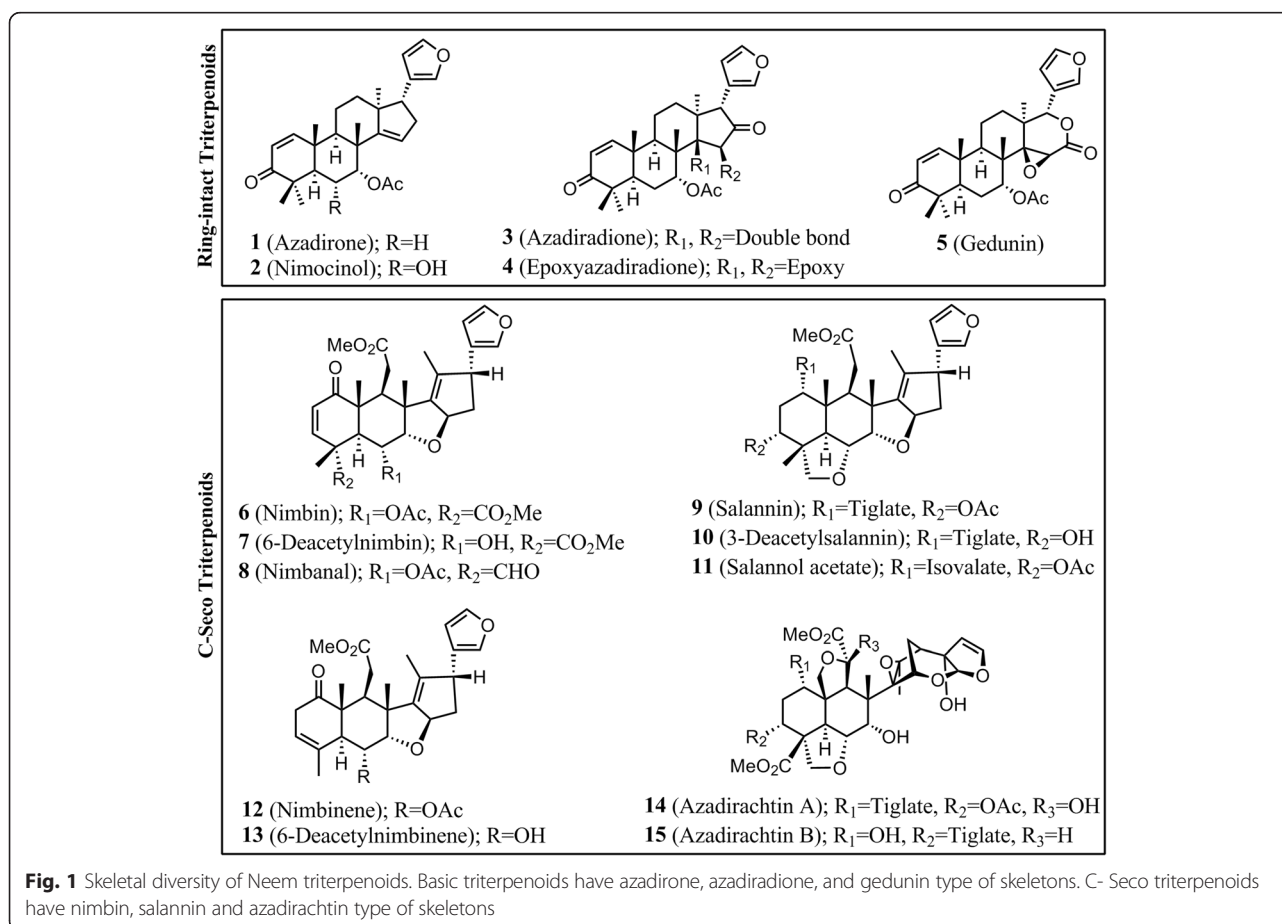[2]CSIR-Institute of Genomics and Integrative Biology, Mall Road, New Delhi 110007, India

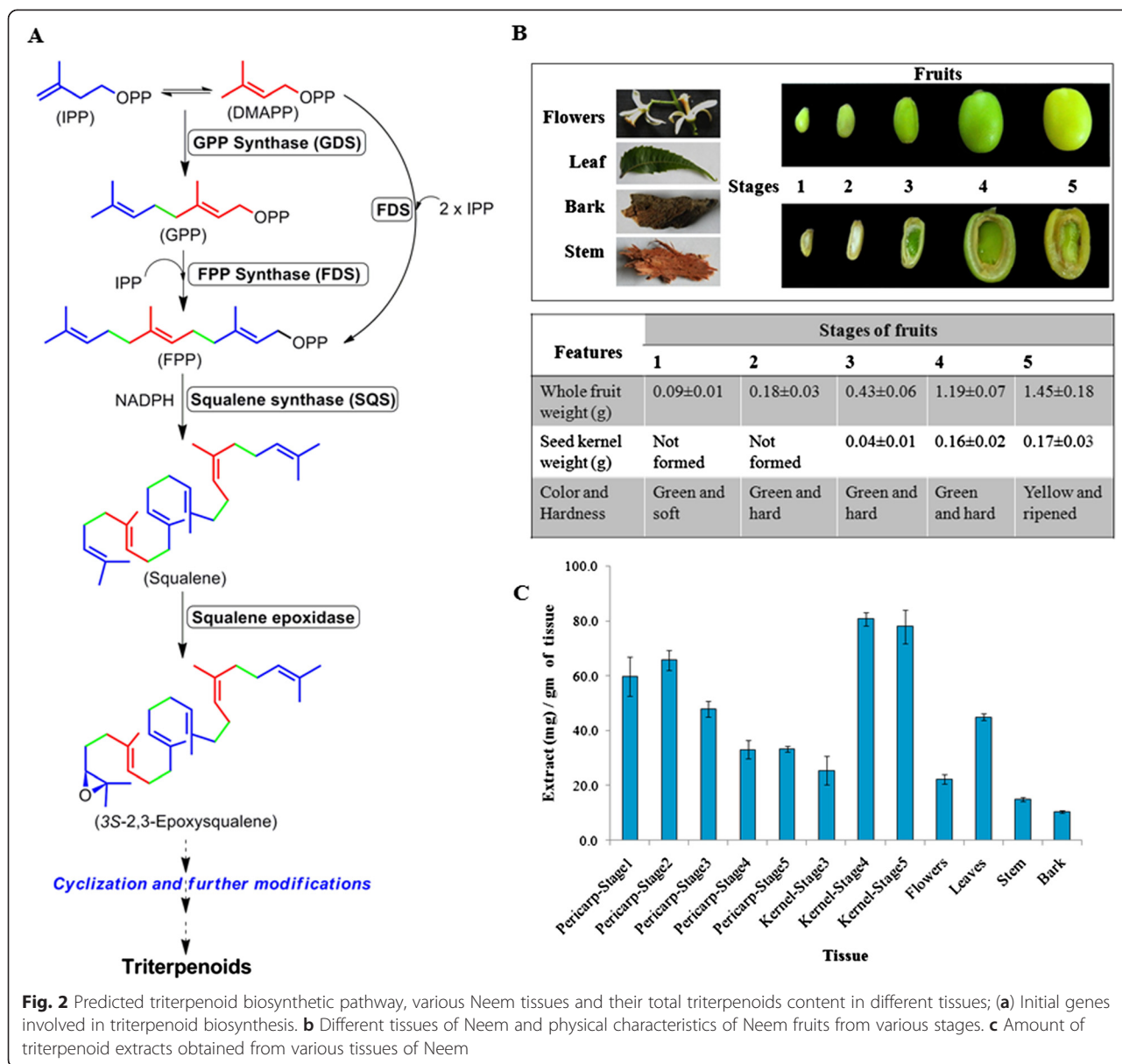Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 2 of 14

Background

Neem tree is one of the richest reserves of secondary metabolites, mainly tetranortriterpenoids (limonoids), which are known to be responsible for insecticidal and wide pharmaceutical activities [1, 2]. Various parts of this evergreen tree have been used as traditional medicine in day-to-day household remedies from ancient time. In addition to its therapeutic potential, Neem is being widely used in eco-friendly commercial pesticides and agrochemicals [3–5]. Over 150 structurally complex, highly oxygenated and skeletally diverse tetranortriterpenoids [2] have been isolated and characterized from different parts of the tree. Depending on the skeletal modifications, they can be categorized into two groups; ring-intact (basic) triterpenoids and C-seco triterpenoids [2, 6]. Ring-intact triterpenoids encompass 4,4,8-trimethyl-17-furanylsteroidal skeleton such as azadirone, azadiradione, and gedunin (1-5) type of structures (Fig. 1). C-seco triterpenoids are generated by the opening and further rearrangements of C-ring thus producing nimbin, salannin and azadirachtin (6-15) type of skeletons (Fig. 1). Although the biosynthetic pathway leading to the formation of triterpenoids (Fig. 2a) in Neem plant has been predicted [1, 7]

genes involved in triterpenoid biosynthesis have not been characterized till date [8].

Secondary metabolites are the final outcome of omics cascade and their distribution pattern is typical characteristic of every life in nature, which can be considered as an intrinsic signature of that species. Targeted metabolomics is all about identification and quantification of known metabolites and their time and space resolved distribution in a specific biological system [9–13]. Hyphenated mass spectrometry is a powerful and most utilized analytical technique in metabolomics due to its high sensitivity, accuracy, resolution, low sample requirement and ability to monitor broad range of metabolites [9, 12–14]. Triterpenoids in Neem are diverse in skeletal architecture, huge in count and their abundance is highly tissue-specific [1, 2]. Except few discrete studies [15, 16], there are no systematic investigations on the tissue- and stage-specific quantitative variation of Neem triterpenoids. It will be of great importance to investigate the targeted metabolic profiling of major triterpenoids in Neem plant, which may enlighten the differential tissue specific abundance of skeletally diverse triterpenoids. Further, correlation of metabolic profiling

Fig. 1 Skeletal diversity of Neem triterpenoids. Basic triterpenoids have azadirone, azadiradione, and gedunin type of skeletons. C- Seco triterpenoids have nimbin, salannin and azadirachtin type of skeletons

Pandreka et al. BMC Plant Biology (2015) 15:214

Page 3 of 14

Fig. 2 Predicted triterpenoid biosynthetic pathway, various Neem tissues and their total triterpenoids content in different tissues; (**a**) Initial genes involved in triterpenoid biosynthesis. **b** Different tissues of Neem and physical characteristics of Neem fruits from various stages. **c** Amount of triterpenoid extracts obtained from various tissues of Neem

with transcriptome helps in analysis and identification of genes involved in Neem triterpenoid biosynthesis.

Terpenoid biosynthesis starts with basic building blocks such as Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) which are in turn synthesized through the mevalonate (MVA) or methylerythritol phosphate (MEP) pathways [17, 18]. Allylic diphosphate, DMAPP undergoes condensation with one or more IPP in head-to-tail fashion to produce linear diphosphates such as geranyl diphosphate (C_{10}, GPP), farnesyl diphosphate (C_{15}, FPP) and geranylgeranyl diphosphate (C_{20}, GGPP) catalyzed by short-chain prenyltransferases such as geranyl diphosphate synthase (GDS), farnesyl diphosphate synthase (FDS) and geranylgeranyl diphosphate synthase (GGDS), respectively

[19–21]. Two molecules of FPP undergo 1-1′ head to head condensation to form squalene via NADPH dependent reduction of presqualene diphosphate intermediate catalyzed by squalene synthase (SQS) [22]. Thus squalene is the first committed precursor for the biosynthesis of triterpenoids [23]. This molecule is also well known to serve as a precursor for the primary metabolites such as steroids required for cell growth and division. Squalene thus acts as an important intermediate governing the balance between primary and secondary metabolism. Squalene undergoes further oxidation to form 2,3-epoxysqualene mediated by squalene epoxidase, followed by cyclization catalyzed by triterpene cyclases to form basic triterpene skeletons [24, 25]. Structural diversity of triterpenoids arises from the modifications of functional groups and

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 4 of 14

rearrangements on the parental backbone of these triterpenes (Fig. 1) [26].

Short-chain prenyltransferases, such as FDS and SQS are shown to play key regulatory role in triterpenoid and phytosterol biosynthesis. To show some instances, when hairy root culture of *Panax ginseng* was treated with methyl jasmonate (MJ) to enhance the production of triterpenoids, FDS was up-regulated [27]. Over expression of mevalonate-5-pyrophosphate decarboxylase and FDS in *Panax ginseng* hairy root culture resulted in increased accumulation of phytosterols and triterepenes [28]. In *Centella asiatica*, overexpression of *Panax ginseng* FDS resulted in overexpression of dammarenediol synthase and cycloartenol synthase and when induced with MJ, enhanced production of triterpenes was observed [29]. Similarly, overexpression of SQS in *Panax ginseng*, *Eleutherococcus senticosus*, *Withania coagulans* and *Arabidopsis thaliana* showed increased production of phytosterols and triterpenoids [30–33]. Therefore, identification and functional characterization of short-chain prenyltransferases and SQS will assist in understanding of triterpenoid biosynthesis.

In this study, fifteen major triterpenoids were quantified in six different Neem tissues including kernel, pericarp, flower, leaf, stem and bark using UPLC-ESI(+)-HRMS based targeted profiling. Tissue specific profiling of triterpenoids delineated the variation in the abundance of triterpenoids across various tissues. This information was further utilized for the selection of tissues for transcriptome analysis followed by identification of initial genes involved in isoprenoid biosynthesis. Amongst the predicted genes from this pathway, here we report, molecular cloning and functional characterization of full-length geranyl diphosphate synthase (AiGDS), farnesyl diphosphate synthase (AiFDS) and squalene synthase (AiSQS) from Neem. Furthermore, using real-time PCR analysis, we showed that the expression level of one of the important genes in the pathway, AiSQS correlates with the triterpenoid content in respective tissues (fruit, leaf and flower).

Results and discussion
Tissue specific quantitative profiling of triterpenoids
The levels of individual fifteen triterpenoids (Fig. 1) were determined in different tissues of Neem including flowers, leaves, stem, bark, five developmental stages of pericarp and three stages of kernel (Additional file 1: Figure S5). The developmental stages of the fruits were classified on the basis of kernel formation, weight, hardness and colour (Fig. 2b). The crude mixture of triterpenoids was extracted from fresh tissues of Neem using solvent partition technique and were analyzed by UPLC-ESI(+)-HRMS in a gradient solvent program of methanol-water. Amount of crude extract obtained was directly correlated with the triterpenoid content of the corresponding tissue (Fig. 2c).

Quantification of the crude extract revealed that kernel of stages 4 and 5 contained the highest amounts of triterpenoids (~80 mg/g of the tissue) followed by pericarp of stages 1, 2 and 3 (~48-66 mg/g). Pericarps of stages 4, 5 and kernel of stage 3 were found to possess comparatively lower amount of triterpenoids in the range of ~25-35 mg/g. Flowers and leaves have been shown to contain 22 and 45 mg/g of triterpenoids (including chlorophyll and other pigments), while stem and bark furnished 15 and 10 mg/g of the tissue respectively.

Standard graphs were prepared for each of the fifteen isolated triterpenoids within the concentration range of 0.04 to 0.003 mg/mL with injection volume 5 μL in UPLC-ESI(+)-HRMS (Additional file 1: Figure S4). They were further utilized for the quantification of individual molecules in the extracts of different tissues of Neem by correlating with the area under respective peaks of extracted ion chromatograms (Additional file 1: Figures S2 and S3). The quantitative level of individual fifteen triterpenoids across various tissues of Neem has been represented in Additional file 1: Figures S3 and S6. Among the fifteen triterpenoids under investigation, azadirachtin A (**14**), a well-studied Neem triterpenoid was found to be highly abundant in seed kernels, especially in the stages 4 and 5 (~3.6 mg/g of the tissue). Pericarp, flowers and leaves showed 100-500 fold lower levels (~0.004-0.04 mg/g) of azadirachtin A as compared to the kernel, whereas bark and stem contained negligible quantities (≤0.005 mg/g, 1000 fold lesser than seed kernel). Similar distribution was observed with the levels of azadirachtin B (**15**). Highest level of azadirachtin B was observed in kernel of stages 4 and 5 (0.5-0.6 mg/g), whereas pericarp and flowers showed 100-150 fold lesser amounts in comparison. Stem and bark were found to possess negligible levels (<0.005 mg/g, 1000 fold lesser than seed kernel) of azadirachtin B. Salannin (**9**) showed highest levels in kernel of stages 4 and 5 (1.2-1.4 mg/g). Salannin content was 4 fold less (~0.3 mg/g of the tissue) in stem as compared to that in kernel. Salannin content in bark was ~0.04 mg/g which was 35 fold lesser in comparison to seed kernel. Flowers, leaves and pericarp showed negligible levels of salannin (≤0.02 mg/g). Highest percentage of 3-deacetylsalannin (**10**) was observed in kernel of stages 4, 5 and stem with 0.01 mg/g of the tissue. Other tissues showed traceable amounts of 3-deacetylsalannin. Nimbin (**6**) was mainly present in kernels in the range of 0.1-0.2 mg/g and in negligible quantities in other tissues. 6-Deacetylnimbin (**7**) was found to be present in kernel of stages 4, 5 and leaves (0.08-0.23 mg/g). Nimbinene (**12**) and 6-deacetylnimbinene (**13**), two pentanortriterpenoids exhibited similar pattern of distribution across different tissues. Highest level was observed in seed kernels of stages 4, 5 and stem within the range of 0.15-0.25 mg/g. Flowers and leaves showed minor quantity (0.02-0.06 mg/g), whereas bark and pericarps

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 5 of 14

exhibited negligible level. Nimbanal (**8**) was present in higher level in kernel of stages 4, 5 and stem (0.05-0.10 mg/ g) and traceable levels were observed in other parts. Salannol acetate (**11**) was found to be abundant in seed kernels and stem with ~0.15 mg/g and in other tissues in minor amounts. Ring-intact triterpenoids (basic limonoids) such as azadirone, azadiradione, epoxyazadiradione and gedunin were found to be present at higher levels in pericarps. Azadiradione (**3**) showed highest level (3.0-8.0 mg/g) in all five developmental stages of pericarps especially in the stages 2 and 3 (7.0-8.0 mg/g), during which the seed kernel formation is about to start. These levels were about 100-200 fold higher than that in seed kernels (kernel stage 4 and 5) and flowers (0.01-0.05 mg/g). Other tissues contained negligible amounts of it (<0.001 mg/g). Similarly, epoxyazadiradione (**4**) showed 400-500 folds higher level in pericarps (9.0-12.0 mg/g; in stages 2 and 3) in comparison to that in the seed kernels (0.01-0.04 mg/g) and 50 folds higher than in flowers (~0.20 mg/g). Azadirone (**1**) was also found to be most abundant in all the developmental stages of pericarps (0.3-0.7 mg/g) especially in the stages 2 and 3 (0.6-0.7 mg/ g) and flowers (0.5 mg/g). Leaves showed very less quantity (~0.08 mg/g) of **1** whereas other tissues contained traceable amounts (<0.001 mg/g). Gedunin (**5**), a potent anticarcinogenic triterpenoid was abundantly present in pericarps, especially in the stages 2 and 3 (~1.0 mg/g). Negligible amount of **5** was present in other tissues (<0.002 mg/g). Nimocinol (**2**), 6α-hydroxy derivative of azadirone was observed to be abundant in leaves (2.9 mg/g), 15 fold higher than flowers (0.18 mg/g) and 50-150 times higher than pericarps (0.02-0.08 mg/g). Other tissues such as kernel, bark and stem showed very less amount of nimocinol (<0.001 mg/g).

Metabolic profiling data (Fig. 3 and Additional file 1: Figure S6) depicted the kernel to be rich in quantity and diversity of triterpenoids especially C-seco triterpenoids of azadirachtin (**14**, **15**), salannin (**9**, **11**), nimbin (**6**, **7**, **8**) and nimbinene (**12**, **13**) skeletons. However, pericarps were found to be rich in triterpenoids mainly consisting of ring-intact (basic) structures such as azadirone (**1**), azadiradione (**3**), epoxyazadiradione (**4**) and gedunin (**5**). Flowers and leaves showed relatively lower levels of triterpenoids and mostly of ring-intact skeletons (**1**, **2**, **3**, **4**). Stem and bark contained very low levels of triterpenoids; majorly C-seco metabolites of salannin (**9**, **11**) and nimbinene (**12**, **13**) type. In essence, profiling data revealed C-seco triterpenoids (**6-10**) to be the major constituents of triterpenoid pool from seed kernel, stem and bark whereas ring-intact skeletons (**1-5**) were observed to be major metabolites of the triterpenoid content obtained from pericarp, flower and leaf.

Transcriptome analysis

For extensive coverage, RNA isolated from triterpenoid rich tissues such as fruit stage 4, leaves and flowers were pooled and used for transcriptome sequencing. A total of 79,079,412 (79.08 million) paired-end reads each of 72 bp length were generated by Illumina GA II platform. 71,537,895 (90.46 %) high quality reads were obtained with more than 20 phred score and reads of low quality were trimmed and used for further analysis. Total 27,390 contigs were generated using Velvet with a hash length of 41. These contigs were given as input for Oases to generate 41,140 transcripts. The average length of transcripts obtained was 1331 bp and the N50 length was 1953 bp (Table 1).

All the transcripts were submitted to Blastx against non-redundant database available at NCBI with an E-value cutoff of 10^{-5}, where, a total of 32,856 (79.8 %) transcripts were annotated (Fig. 4a). Pathway annotation
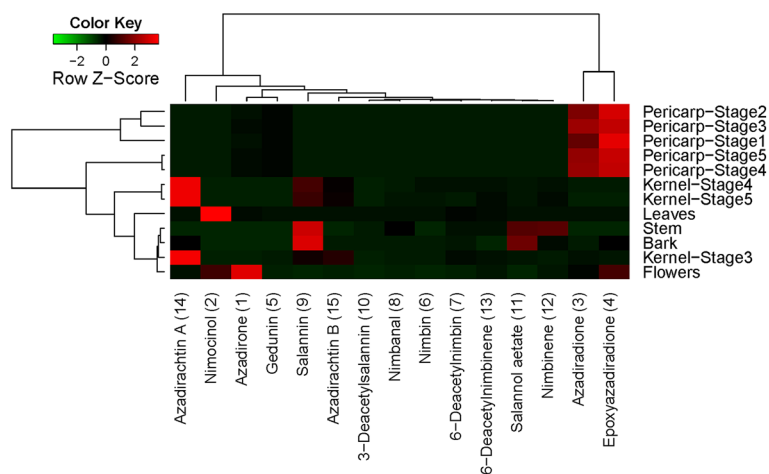

Fig. 3 Quantitative abundance of major triterpenoids in different tissues of Neem. Basic and C-seco triterpenoids are highly abundant in Pericarp and Kernel respectively as compared to other tissues

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 6 of 14

Table 1 Summary of transcriptome sequencing and assembly

Total number of reads	79079412
Total number of HQ reads	71537895
Hash length	41
Contigs generated	27390
Average contig length	897.431
N50 length of contigs	1479
Transcripts generated	41140
N50 length of transcripts	1953
Number of reads assembled	68871778

was carried out by KAAS (KEGG Automatic Annotation Server) with *Arabidopsis thaliana* (thale cress) and *Oryza sativa japonica* (Japanese rice) as the reference database. Out of the 41,140 transcripts only 6281 transcripts were assigned 2749 unique KO numbers, which covered 223 pathways (Fig. 4b). Virtual ribosome, a web based server, was used for finding the Open Reading Frame (ORF) of transcripts. 27,368 transcripts had an ORF with length more than 99 amino acids and 67 transcripts without any ORF (Fig. 4c). The peptide sequences of transcripts with length more than 99 amino acids were submitted to Pfam analysis. 18,807 transcripts were assigned different Pfam IDs. A total of 3467 different Pfam IDs were assigned to the transcripts (Fig. 4d). Based on

transcriptome annotation, all the genes involved in triterpenoid back-bone biosynthesis from isoprene units (MVA pathway and MEP pathway) to triterpene cyclase were found (Additional file 1: Table S1). A total of 134 transcripts predicted as cytochrome P450 monooxygenases and two transcripts as cytochrome P450 reductases were identified. Based on BLAST results, with reference to *Arabidopsis thaliana* cytochrome P450, Neem CYP450s were classified into 39 families and 78 subfamilies, out of which most of the CYP450 belonged to CYP71 family. Seven transcripts were related to plant steroid biosynthesis and six transcripts related to triterpenoid biosynthesis were predicted (Additional file 1: Table S1). Recently, Neem draft genome and transcriptome of fruit, stem, leaf and flower [34], and suppression subtractive hybridization of transcripts between fruit mesocarp and endocarp [35] have been reported. However, there are no reports regarding functional characterization of the genes involved in Neem triterpenoid biosynthesis. To further explore this pathway, two short-chain prenyltranferases and squalene synthase were selected for functional characterization based on the transcriptome data.

Heterologous expression and functional characterization of short-chain prenyltransferases (AiGDS and AiFDS)

Short-chain prenyltransferases function at the branching point of terpenoid metabolism and play regulatory role

Fig. 4 Functional annotation of transcriptome; (**a**) Based on Blastx analysis 80 % (32,856) transcripts had homologous proteins in NCBI nr database. **b** Based on KAAS analysis only 15.2 % (6281) transcripts were assigned 2749 KO numbers. **c** Based on virtual ribosome analysis 66.5 % (27,368) transcripts had ORF region length more than 100 amino acids and 0.001 % (67) Transcripts did not show ORF region. **d** Based on Pfam analysis 69.1 % (18,907) transcripts were assigned Pfam IDs

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 7 of 14

in the distribution of isoprene units into various terpenoids biosynthesis. In total, 12 short-chain prenyltranferases from Neem transcriptome were identified (Additional file 1: Table S1). Based on functional annotation studies, two geranyl diphosphate synthases (GDS), nine putative geranylgeranyl diphosphate synthases (GGDS) and one farnesyl diphosphate synthase (FDS) were identified. Sequence analysis using BLAST indicated that Neem_transcript_10912 was a homomeric GDS and Neem_transcript_10001 could be the smaller subunit of heteromeric GDS. TargetP analysis showed that both of these genes are localized in the mitochondria (Additional file 1: Table S5). For further study, Neem_transcript_10912 (AiGDS) and Neem_transcript_25722 (AiFDS) were selected for cloning and functional characterization.
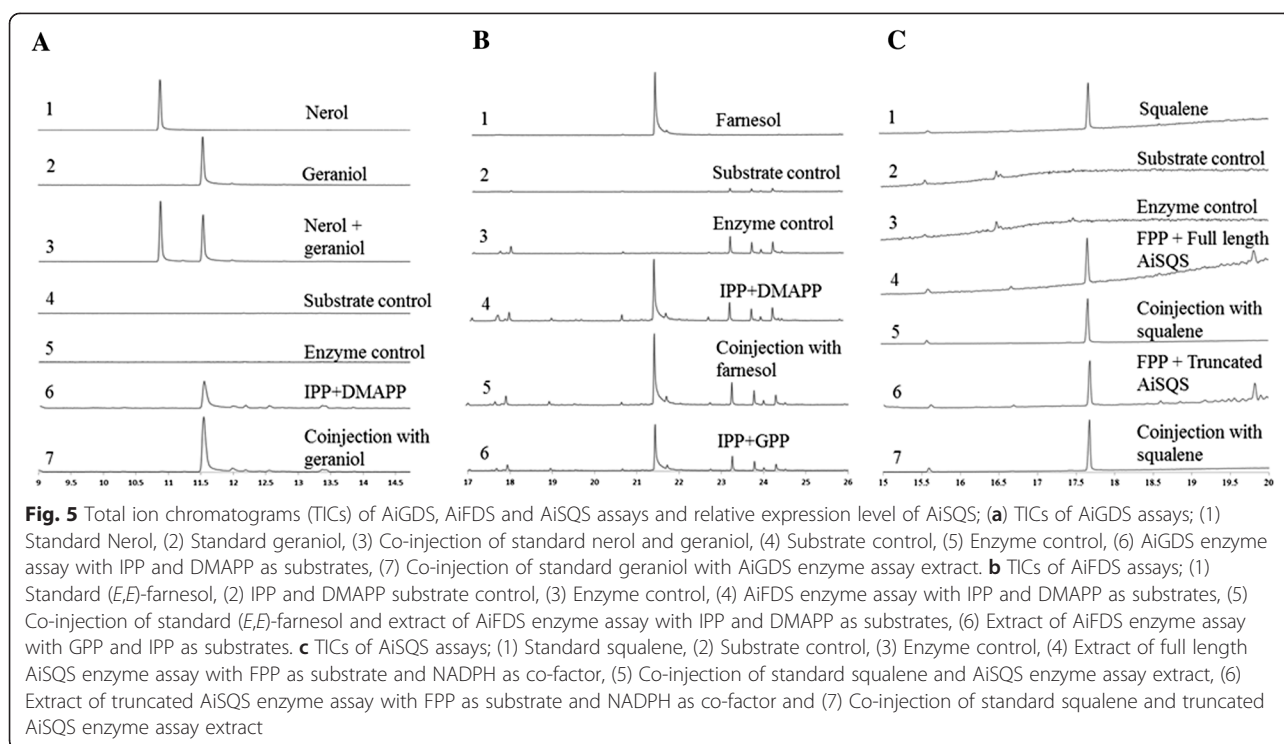
The ORF of AiGDS [GenBank: KM108315] was 1263 bp, which coded for a protein of 420 amino acids with theoretical molecular weight and calculated pI as 46.1 kDa and 6.33, respectively. AiGDS had maximum identity with several plant characterized homomeric GDSs such as 90 % identity to homomeric GDS from *Citrus sinensis* [GenBank: CAC16851] [36], 86 % identity to GDS from *Mangifera indica* [GenBank: AFJ52721] [37] and 76 % identity to GDS from *Catharanthus roseus* [GenBank: AGL91647] [38]. The percentage identity matrix of AiGDS with other plant homomeric GDS and heteromeric GDS larger subunits indicated that AiGDS possesses 71 % to 89 % identity with homomeric GDS (Additional file 1: Table S2). The multiple sequence alignment of AiGDS consisted of two aspartate rich motifs $DDX_{(2-4)}D$ and DDXXD which are highly conserved motifs in prenyltransferases and involved in substrate and metal ion binding (Additional file 1: Figure S7). CxxxC motifs were not observed in AiGDS, which play a key role in the interaction of heteromeric GDS [39]. The ORF of AiGDS was cloned into pET32a expression vector having an N-terminal thioredoxin domain and subsequently expressed in BL21 (DE3) cells. However recombinant AiGDS protein was found in inclusion bodies. To enhance solubility, AiGDS cloned construct was transformed into Lemo 21 (DE3) cells [40] and expression was carried out. Recombinant AiGDS protein remained solely in the insoluble portion in the pellet. Eventually we were able to obtain soluble active AiGDS by re-suspending the pellets in lysis buffer, then drop-wise addition of 0.1 M NaOH until pH 11.0 with constant swirling on ice till the solution became clear. The pH was then reduced to 7.0 using 0.1 M HCl under similar conditions [41]. The resulting solution was centrifuged at $10,000 \times g$ and subjected to SDS-PAGE analyses (Additional file 1: Figure S11A). The AiGDS was found to be in soluble form in the supernatant, which was subjected to purification by Ni-NTA affinity chromatography. The recombinant protein was over 94 % pure as analysed by SDS-PAGE (Additional file 1: Figure S11A). Purified recombinant AiGDS was incubated with

equimolar concentration of IPP and DMAPP followed by treatment with alkaline phosphatase to hydrolyze the diphosphate esters to their corresponding alcohols. The extracted assay mixture was analyzed by GC-MS and the products formed were confirmed by comparing the retention time and coinjection studies with standard geraniol (Fig. 5a). GC-MS analyses of the extracts of alkaline phosphatase treated assay mixture of AiGDS with GPP/FPP and IPP indicated that AiGDS failed to synthesize chain elongation products FPP (C_{15}) or GGPP (C_{20}) suggesting that AiGDS can catalyse the chain elongation reaction to produce GPP (C_{10}) as sole enzymatic product.

AiFDS [GenBank: KM10831] ORF of 1029 bp length was found to be encoding for a protein of 342 amino acids. The theoretical molecular weight and pI for this polypeptide were 39.5 kDa and 5.59 respectively. The sequence comparison of AiFDS exhibited 83 % identity with FDS from *Mangifera indica* [GenBank: AFJ52720] [37], 82 % identity with that from *Santalum album* [GenBank: AGV01244.1] and 81 % identity with FDS from *Catharanthus roseus* [GenBank: ADO95193.1] [42]. The multiple sequence alignment of AiFDS consisted of two aspartate rich motifs $DDX_{(2-4)}D$ and DDXXD (Additional file 1: Figure S8) which were highly conserved motifs in prenyltransferases. AiFDS was cloned into pET32a expression vector. The cloned construct was transformed into BL21 (DE3) cells and expressed. AiFDS was obtained as soluble form and purified by Ni-NTA affinity column chromatography. The recombinant protein was over 98 % pure as analyzed by SDS-PAGE (Additional file 1: Figure S11B). Buffers used for AiGDS and AiFDS protein purification are given in Addition file 1: Table S4. The purified short-chain prenyltransferase was incubated with DMAPP/ GPP and IPP followed by treatment with alkaline phosphatase. GC-MS analyses of the assay extracts indicated the formation of FPP which was further confirmed by comparing the retention time, mass fragmentation pattern and coinjection studies with standard (*E,E*)-farnesol (Fig. 5b). Further GC-MS analysis of alkaline phosphatase treated assay mixture of AiFDS with FPP and IPP did not show formation of geranylgeraniol indicating that AiFDS catalyses the chain elongation reaction to produce FPP as the sole enzymatic product.

Heterologous expression and functional characterization of squalene synthase (AiSQS)

An ORF of 1176 bp encoding a polypeptide of 396 amino acids was identified as AiSQS [GenBank: JQ327160]. The theoretical pI of protein was found to be 8.18 and molecular weight of 44 kDa. The amino acid sequence of AiSQS shared 86 % identity with squalene synthase from *Diospyros kaki* [GenBank: ACN69082], 85 % identity with *Camellia oleifera* [GenBank: AGB05603], 84 %

Fig. 5 Total ion chromatograms (TICs) of AiGDS, AiFDS and AiSQS assays and relative expression level of AiSQS; (**a**) TICs of AiGDS assays; (1) Standard Nerol, (2) Standard geraniol, (3) Co-injection of standard nerol and geraniol, (4) Substrate control, (5) Enzyme control, (6) AiGDS enzyme assay with IPP and DMAPP as substrates, (7) Co-injection of standard geraniol with AiGDS enzyme assay extract. **b** TICs of AiFDS assays; (1) Standard (*E,E*)-farnesol, (2) IPP and DMAPP substrate control, (3) Enzyme control, (4) AiFDS enzyme assay with IPP and DMAPP as substrates, (5) Co-injection of standard (*E,E*)-farnesol and extract of AiFDS enzyme assay with IPP and DMAPP as substrates, (6) Extract of AiFDS enzyme assay with GPP and IPP as substrates. **c** TICs of AiSQS assays; (1) Standard squalene, (2) Substrate control, (3) Enzyme control, (4) Extract of full length AiSQS enzyme assay with FPP as substrate and NADPH as co-factor, (5) Co-injection of standard squalene and AiSQS enzyme assay extract, (6) Extract of truncated AiSQS enzyme assay with FPP as substrate and NADPH as co-factor and (7) Co-injection of standard squalene and truncated AiSQS enzyme assay extract

identity with *Euphorbia tirucalli* [GenBank: BAH23428] and 84 % identity with that from *Glycyrrhiza glabra* [GenBank: BAA13084.1]. Eukaryotic SQSs have four conserved regions and are important for catalysis as indicated by biochemical characterization of site-directed mutants and crystal structure of human squalene synthase [43] (Additional file 1: Figure S9). The aspartate rich motifs found in region 1 and 3 are involved in binding of the diphosphate moiety of FPP via bridging Mg^{2+} ions. Careful analysis of AiSQS sequence with TMHMM program showed the presence of transmembrane motif YNTTM IIMLFIILAIIFAYLSAN at the C-terminus. Although transmembrane domain exhibits low level of sequence homology with other SQS enzymes, this domain is highly hydrophobic and consistent with the putative endoplasmic reticulum anchoring function.

Squalene synthase has been characterized previously from human [43], rodents [44, 45], plants [46–48], protozoa [49] and fungi [50]. All these SQS enzymes were obtained in soluble form by deletion of a putative C-terminal membrane-spanning motif [51]. In the present study we have cloned the full-length ORF of AiSQS, as well as a truncated AiSQS by deletion of 15 amino acids from N-terminal and 63 amino acids from the C-terminal end into pRSET-C and pET28c vectors respectively. The truncated AiSQS was transformed into BL21 (DE3) cells, expressed and purified by subjecting to Ni-NTA affinity column chromatography. Purified truncated AiSQS was analyzed by SDS-PAGE which showed a single band

(>90 % purity) at ~35 kDa, consistent with the predicted molecular mass for the $(His)_6$-tagged enzyme (Additional file 1: Figure S11D).

The full-length recombinant AiSQS protein was expressed in BL21 star (DE3) cells. Majority of the protein was found to be insoluble (Additional file 1: Figure S11C). Lee and Poulter observed that adding glycerol to the lysis and purification buffers helped in solubilization of the insoluble *T. elonatus* BP-1 SQS [52]. Induced cell pellets were disrupted in lysis buffer containing 50 % (v/v) glycerol and 1 % CHAPS. The glycerol concentration in cell lysate obtained was reduced to 20 % (v/v) by adding lysis buffer (without glycerol). This lysate was subjected to Ni-NTA affinity column chromatography. The purified full length AiSQS, when analyzed by SDS-PAGE, exhibited a single band (90 % purity) at approximately 44 kDa, consistent with the predicted molecular mass for the $(His)_6$-tagged enzyme (Additional file 1: Figure S11C). Purified proteins were flash-frozen in liquid nitrogen and stored at -80 °C until further use. Buffers used for AiSQS full length and truncated protein purification are given in Addition file 1: Table S4.

GC-MS analyses of the assay extracts of full length and truncated AiSQS with FPP in the presence of NADPH indicated the formation of squalene. The formation of squalene was further confirmed by comparing the retention time, mass fragmentation pattern and co-injection studies with standard squalene (Fig. 5c). This confirms that AiSQS catalyzes the condensation of two
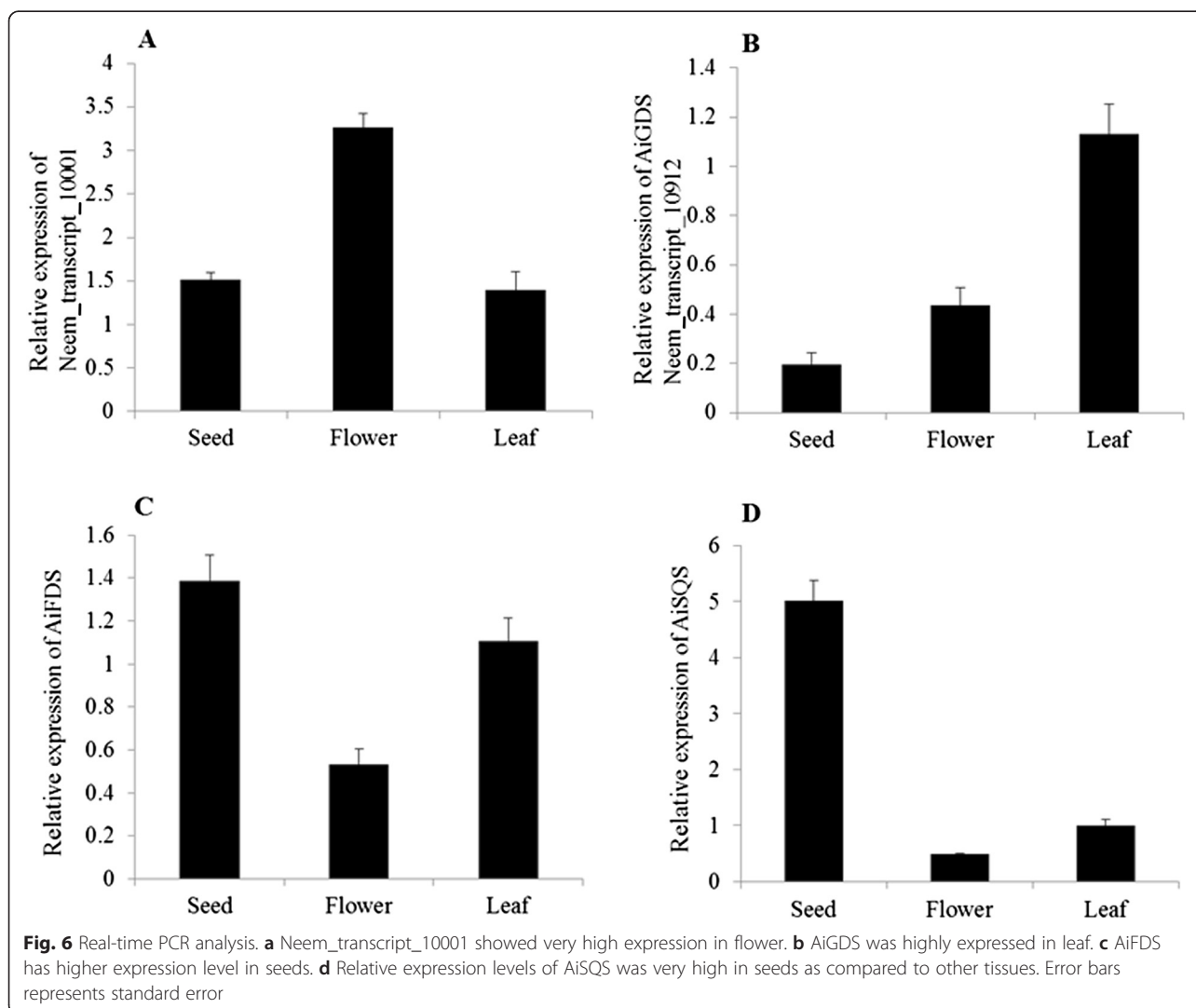
Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 9 of 14

molecules of farnesyl diphosphate (FPP) to form squalene through a NADPH-dependent rearrangement of C1′-2-3-linked triterpene intermediate, presqualene diphosphate [52].

Real time PCR analysis

To determine the role of short-chain prenyl diphosphate synthases and squalene synthase in triterpenoid biosynthesis, real time PCR analysis of the Neem_transcript_10001 (smaller subunit of heteromeric geranyl diphosphate synthase), AiGDS, AiFDS, and AiSQS was carried out.

AiSQS is the first committed enzyme involved in triterpene biosynthesis in Neem. Real time PCR was carried out for AiSQS from flowers, leaves and fruit and normalized with 18S rRNA expression level. Neem fruit showed fivefold higher expression level in comparison with the leaves and tenfold higher relative expression level than flowers (Fig. 6d). The results

were in correlation with profiling of triterpenoids from different tissues. Neem fruits as a whole, not only showed structurally diverse triterpenoids but also showed very high levels of these metabolites. On the other hand, flowers and leaves exhibited lesser skeletal diversity and quantity of abundant triterpenoids. Squalene is the precursor of primary metabolites such as membrane sterols and steroid hormones required for cell division and growth. Also, it serves as precursor for triterpenoids found in Neem, which assign squalene, a crucial branch point between primary and secondary metabolism. Transgenic *Panax ginseng* overexpressing squalene synthase has previously shown to produce higher levels of triterpene and phytosterols than wild type strains which depict the key role of intracellular squalene flux between primary and secondary metabolism [31]. High expression levels of AiSQS in fruits indicated considerable amount of squalene flux might get diverted towards triterpenoids formation in Neem fruits.

Fig. 6 Real-time PCR analysis. **a** Neem_transcript_10001 showed very high expression in flower. **b** AiGDS was highly expressed in leaf. **c** AiFDS has higher expression level in seeds. **d** Relative expression levels of AiSQS was very high in seeds as compared to other tissues. Error bars represents standard error

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 10 of 14

AiFDS (Fig. 6c), compared to other tissues, showed very high expression levels in seeds. Similar expression patterns of AiFDS and AiSQS suggest that both these genes could be involved in triterpenoid biosynthesis. On the contrary, AiGDS (Fig. 6b) and Neem_transcript_10001 (Fig. 6a) showed very high expression in leaf and flower, respectively, compared to other tissues. These results indicate that AiGDS may not be involved in triterpene biosynthesis in Neem.

Phylogenetic analysis

Neighbour joining phylogenetic tree was constructed based on the deduced amino acid sequences of AiGDS, AiFDS and AiSQS with corresponding enzymes from different organisms, which were retrieved from the NCBI GenBank database (Additional file 1: Figure S10). The degree of relatedness correlated well with the amino acid similarity among the plant proteins, which indicated AiGDS, AiFDS and AiSQS belonged to the clade of plant kingdom. These enzymes from Neem were classified into one cluster revealing their closest evolutionary relationships with the plant group.

Conclusions

Due to immense significance of Neem as a wonder tree and known to synthesize biologically and commercially important triterpenoids having highly complex carbon skeleton with diverse functional groups, it is of great interest to study their biosynthetic pathway. Levels of total triterpenoid and fifteen major individual triterpenoids were quantified in various tissues of the Neem plant. Tissue specific variation in the abundance of triterpenoids has been observed. The mature seed kernel and pericarp of initial stages were found to contain the highest amount of triterpenoids. Furthermore, a wide diversity of triterpenoids, especially C-seco triterpenoids were observed in kernel as compared to the other tissues. Pericarp, flower and leaf contained mainly ring-intact triterpenoids. From transcriptome analysis, short-chain prenyl trasnferases, squalene synthase, squalene expoxidase, triterpene synthases and putative cytochrome P450 genes were predicted. The genes involved in the initial steps of isoprenoid biosynthesis, such as AiGDS, AiFDS and AiSQS were cloned and functionally characterized. Furthermore, AiFDS and AiSQS expression levels were found to be nicely correlating with the triterpenoids content of various tissues of Neem.

Methods

Materials and chemicals

Neem tissues for the profiling of triterpenoids were collected from Pune region, Maharashtra, India in the period March to May. Fifteen reference triterpenoids were isolated and characterized as reported earlier [53, 54, 6] and described briefly in Additional file 1. For extraction, HPLC grade solvents were purchased from Sigma (St. Louis, MO, USA). For UPLC-ESI(+)-MS experiments LC-MS grade solvents were procured from Avantor Performance Materials, JT Baker (PA, USA). SuperScript® III First-Strand Synthesis System (Invitrogen) was used for cDNA synthesis. For PCR amplification, AccuPrime™ (Invitrogen) polymerase was used. For Restriction digestion, NEW ENGLAND BioLabs®inc(NEB) restriction enzymes were used. Gel extraction of restricted product and vector were carried out by GenElute™ Gel Extraction Kit from Sigma. T_4 DNA ligase from Invitrogen was used for ligation. TOP10 cells (Invitrogen) were used for cloning. Lemo21 (DE3) cells (NEB), BL21 (DE3) cells (NEB) and BL21 Star (DE3) cells (Invitrogen) were used as expression cells. Ni-NTA agarose (Invitrogen) was used for protein purification. Enzyme assay samples were analyzed on Agilent 7890A GC coupled with 5975C mass detector. Geraniol, nerol, (*E,E*)-farnesol, squalene standards were purchased from Sigma Aldrich. IPP, FPP, GPP, and DMAPP were synthesized as reported previously [55, 56].

Extraction of total triterpenoids

Fresh Neem tissues (0.5 g) were extracted with methanol (10 mL × 3), by continuous stirring for 3 h. The pooled methanol layer after concentration under reduced pressure at 50 °C was partitioned between ethyl acetate (20 mL) and water (20 mL). The organic layer was separated, passed through anhydrous sodium sulphate and concentrated under similar conditions to obtain the crude triterpenoid extract. Extraction of individual tissues was performed in triplicates.

UPLC-ESI(+)-HRMS profiling of triterpenoid extract

For triterpenoids profiling, UPLC-ESI(+)-HRMS runs were performed on Q Exactive Orbitrap associated with Accela 1250 pump (Thermo Scientific, MA, USA). Mixture of triterpenoids were dissolved in a known volume of methanol (concentration ~0.2 mg/mL), centrifuged to remove the suspended particles and injected (10 μL) in UPLC-ESI(+)-HRMS (Additional file 1: Figure S5). Samples were resolved through Acquity BEH C18 UPLC column (2.1 × 100 mm) of particle size 1.7 μM with a flow rate of 0.3 mL/min and gradient solvent program of 35 min (0.0 min, 40 % methanol/water; 5.0 min, 50.0 % methanol/water; 10.0 min, 60 % methanol/water; 25.0 min, 65 % methanol/water; 30.0 min, 90 % methanol/water; 32.0 min, 90 % methanol/water; 34.0 min, 40 % methanol/water; 35.0 min, 40 % methanol/water). 0.1 % LC-MS grade formic acid was also added to water (mobile phase). Profiling experiments were performed in ESI-positive ion mode using the tune method as follows: sheath gas (nitrogen) flow rate 45 units, auxiliary gas (nitrogen) flow rate 10 units, sweep gas (nitrogen) flow rate 2 units, spray voltage (|KV|) 3.60, spray current (μA)

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 11 of 14

3.70, capillary temperature 320 °C, s-lens RF level 50, heater temperature 350 °C. ESI(+)-HRMS data were recorded in full scan mode within the mass range m/z 100 to 1000. Profiling data were analyzed through Thermo Xcalibur software. Retention times (R_t) and extracted ions for the individual studied triterpenoids have been listed in Table 2. UPLC-ESI(+)-HRMS chromatograms for individual standard triterpenoids and their corresponding ESI(+)-HRMS spectra have been provided in Additional file 1: Figures S2 and S3. R version 3.1.2 was used for generating heatmap.

Transcriptome analysis

Total RNA was isolated using Spectrum Plant total RNA isolation kit (Sigma-Aldrich). Equal quantity of RNA from each tissue was mixed. Transcriptome library was constructed using TruSeq RNA Sample Preparation Guide (Illumina). Quality of the prepared library was analyzed by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent). 79,079,412 paired end raw reads were generated with the length of 72 bp by Illumina GA II analyzer. *De novo* assembly was carried out by Velvet (version- 1.1.05) with hash length 41 [57]. A total of 27,390 contigs were generated with average contig length of 897 and N50 value of 1479. These contigs were then submitted to Oases (version- 0.2.01) to generate a total of 41,140 transcripts [58]. Neem transcripts were submitted to Blastx against non-redundant database available at NCBI with E-value cutoff of 10^{-5}. Pathway annotation was done by bidirectional best hit method of KAAS (KEGG Automatic Annotation Server. http://www.genome.jp/kegg/kaas) with *Arabidopsis thaliana* (thale cress) and *Oryza sativa japonica* (Japanese rice) as the reference database [59].

Table 2 Retention times (R_t), extracted ions and corresponding molecular fragments for the studied triterpenoids

Triterpenoid	R_t (min)	Extracted ion (m/z)	Fragment
Azadirachtin A (14)	7.4	703	$[M + H-H_2O]^+$
Azadirachtin B (15)	8.6	645	$[M + H-H_2O]^+$
Salannin (9)	21.3	597	$[M + H]^+$
3-Deacetylsalannin (10)	18.5	555	$[M + H]^+$
Nimbin (6)	18.0	541	$[M + H]^+$
6-Deacetylnimbin (7)	14.7	499	$[M + H]^+$
Nimbinene (12)	20.7	483	$[M + H]^+$
6-Deacetylnimbinene (13)	16.5	441	$[M + H]^+$
Nimbanal (8)	17.8	511	$[M + H]^+$
Salannol acetate (11)	28.4	599	$[M + H]^+$
Azadiradione (3)	16.3	451	$[M + H]^+$
Epoxyazadiradione (4)	27.4	467	$[M + H]^+$
Azadirone (1)	31.7	437	$[M + H]^+$
Gedunin (5)	20.4	483	$[M + H]^+$
Nimocinol (2)	29.9	453	$[M + H]^+$

Virtual ribosome, (http://www.cbs.dtu.dk/services/Virtual Ribosome/) a web based server, was used for deducing the ORFs of these transcripts [60]. The peptide sequences of transcripts with length more than 99 amino acids were submitted to batch search of Pfam (http://pfam.xfam.org/search#tabview=tab1) [61].

Cloning and characterization of AiGDS, AiFDS and AiSQS

The Neem seed RNA was used for the synthesis of cDNA using SuperScript® III First-Strand Synthesis System (Invitrogen). Full length primers for AiGDS and AiFDS ORFs were designed using their transcripts as a template (Additional file 1: Table S3). Synthesized cDNA was used for PCR reaction using AccuPrime (Invitrogen). PCR products were cloned into pET32a expression vector using respective cloning sites. Full length and truncated primers for AiSQS were designed from Neem_transcript_33869 (Additional file 1: Table S3). PCR products were cloned into pCR Blunt vector. Further, the ORF was digested with EcoRI and the resulting fragment was ligated into pRSET-C vector for full length AiSQS and pET28c for truncated AiSQS. The expression of the recombinant plasmids containing AiGDS, AiFDS, truncated AiSQS were carried out in BL21 (DE3) cells except full length AiSQS, which was expressed in BL21 Star (DE3) cells.

Initially, AiGDS and full length AiSQS were found in inclusion bodies. Expression of AiGDS in Lemo 21 (DE3) cells did not show any improvement in the solubility. To obtain the soluble AiGDS protein, the pellet obtained after crude lysate centrifugation at $10,000 \times g$ was resuspended in lysis buffer, pH was increased to 11.0 with 0.1 M NaOH and then reduced to 7.0 with 0.1 M HCl (pH adjustment was done on ice with continuous stirring). The resulting solution was centrifuged at $10,000 \times g$ for 10 min at 4 °C [41]. The supernatant containing AiGDS protein was purified over Ni-NTA affinity chromatography by following user manual.

Purification of full length AiSQS was attempted under denaturing conditions in 50 mM Tris buffer containing 6 M guanidium hydrochloride as well as 8 M urea as denaturing agents. Refolding was attempted by stepwise slow removal of denaturants under dialysis. However, the protein obtained was not catalytically active. Purification under native conditions using buffer combinations of HEPES, TRIS, MOPS with non-ionic detergents like T ween 20, Triton X-100 also did not yield sufficient amount of soluble protein. A considerable amount of protein was found in soluble fractions using 50 % glycerol and 1 % CHAPS in Phosphate buffer. All the recombinant proteins were purified by Ni-NTA affinity chromatography. Buffers used for recombinant protein purifications were given in Addition file 1: Table S4. Protein estimation was performed by Bradford assay [62] and the protein purity was analyzed on SDS-PAGE (Additional file 1: Figure S11).

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 12 of 14

Enzyme assays for AiGDS and AiFDS were performed in HEPES buffer with DMAPP (100 μM)/GPP (100 μM) and IPP (100-200 μM) as substrates. 100 μM FPP was used as substrate for full length and truncated AiSQS with 1 mM NADPH as cofactor. The reaction mixtures were incubated at 30 °C for 2 h. AiSQS assay reaction was quenched by adding 1 M sodium hydroxide. For AiGDS and AiFDS assays, alkaline phosphatase (6 U) was added and further incubated at 37 °C for 1 h. Reaction mixtures were extracted thrice using n-hexane. Samples were concentrated with a stream of dry nitrogen and analysed by GC-MS on 30 m × 0.25 mm × 0.25 μm capillary columns (HP-5 and HP-5 MS, J & W Scientific). Functional characterization of AiFDS and AiGDS was carried out on GC-MS using the program: 70 °C for 1 min, 5 °C/min rise till 150 °C, 10 °C/min rise till 270 °C and hold for 5 min (Program 1). For the functional characterization of AiSQS, the program used was: initial temperature of 150 °C for 2 min followed by increase in temperature to 320 °C at the rate of 10 °C/min and hold at 320 °C for 11 min (Program 2). Product formation was confirmed by co-injection with authentic standards and comparing the mass fragmentation pattern and retention time (Fig. 5).

RT-PCR analysis

Real time PCR was carried out using Super Script III platinum SYBR green one-step qRT-PCR kit (Invitrogen, USA). In brief, for AiSQS quantification, 100 ng of DNase treated total RNA was added with AiSQS primers and for 18S intrinsic control, 18S primers were used (Additional file 1: Table S3). cDNA synthesis and PCR were carried out in a single tube reaction. cDNA synthesis was performed at 50 °C for 5 min followed by denaturation at 95 °C for 5 min and subsequent 40 cycles of denaturation step at 95 °C for 3 s, combined annealing and extension step at 60 °C for 30 s per cycle.

Quantification of AiGDS, AiFDS and Neem_Transcript_ 10001, was performed as follows: Initial cDNA synthesis was performed at 50 °C for 20 min, followed by 95 °C for 5 min, 40 cycles of 95 °C for 10 s and 60 °C for 30 s. GAPDH primers were used as an endogenous control to normalize the expression levels between different tissues. Threshold (Ct) values were obtained and ΔCt was calculated as Ct target gene – Ct endogenous reference gene. Relative fold difference was calculated using $2^{\Delta Ct}$. Experiments were carried out using three biological replicates with five technical replicates each.

Phylogenetic analysis

Reference protein sequences were obtained from GenBank database. Sequences were aligned using ClustalW using default parameters [63]. Neighbour joining tree was constructed with MEGA version 6.06 software [64]. Bootstrap analyses with 1000 replicates were also conducted in order to obtain confidence levels for the branches.

Availability of supporting data

The Illumina RNA-seq data generated from pooled RNA from leaves, fruits and flowers of *Azadirachta indica* are available in the NCBI SRA (http://trace.ncbi.nlm.nih.gov/Traces/sra) with accession SRR2145149.

Additional file

Additional file 1: Methods 1. Isolation of Neem triterpenoids from seed kernel and pericarp. **Methods 2.** Characterization of purified Neem triterpenoids. **Figure S1.** TLC profile of crude extracts and purified triterpenoids (developed in 70 % ethyl acetate in n-hexane for twice). **Figure S2.** UPLC-ESI(+)-quadrupole/orbitrap-MS extracted ion chromatograms of the fifteen pure triterpenoids from Neem. Chromatograms have been arranged in the order of increasing retention time. **Figure S3.** ESI(+)-quadrupole/orbitrap-MS spectra of the fifteen pure triterpenoids from Neem. **Figure S4.** Standard graphs for the purified triterpenoids prepared in UPLC-ESI(+)-quadrupole/orbitrap-MS; concentration range 0.040-0.003 mg/mL, injection volume 5 μL. **Figure S5.** Representative UPLC-ESI(+)-quadrupole/orbitrap-MS chromatograms of various Neem tissue extracts (× denotes non-triterpenoids with molecular mass less than 350). **Figure S6.** Quantitative abundance of individual triterpenoids in different tissues of Neem. **Figure S7.** Multiple sequence alignment of *A. indica* geranyl diphosphate synthases (AiGDS). **Figure S8.** Multiple sequence alignment of *A. indica* farnesyl diphoshate synthase (AiFDS). **Figure S9.** Multiple sequence alignment of *A. indica* Squalene synthase (AiSQS); Amino acid sequence alignment of *C. annuum* (CaSQS, AAD20626), *N. tabacum* (NtSQS, AAB08578), *A. indica* (AiSQS, AFJ15526), *L. japonicas* (LjSQS, BAC56854), *G. max* (GmSQS, NP_001236365), *P. vulgaris* (PvSQS, AHA84150). The solid lines indicate four highly conserved regions 1, 2, 3 and 4 which are considered to be the catalytic sites of squalene synthases. **Figure S10.** Phylogenetic analysis of AiGDS, AiFDS and AiSQS. **Figure S11.** Purification of recombinant AiGDS, AiFDS and AiSQS. **Table S1.** Predicted genes for Triterpenoid back bone biosynthesis. **Table S2.** Present Identity Matrix of AiGDS with plant homomeric GDS and heteromeric GDS Larger subunits. **Table S3.** Primers and vectors used for cloning of AiGDS, AiFDS and AiSQS and RT-PCR primers of 18S rRNA, GAPDH, Neem_transcript_10001, AiGDS and AiSQS. **Table S4.** Buffers used for AiGDS, AiFDS and AiSQS protein purification. **Table S5.** TargetP analysis Neem_transcript_10912 (AiGDS) and Neem_Transcript_10001. (DOCX 3387 kb)

Abbreviations

MVA: Mevalonate pathway; MEP: Methylerythritol phosphate pathway; GPP: Geranyl diphosphate; FPP: Farnesyl diphosphate; GGPP: Geranylgeranyl diphosphate; GDS: Geranyl diphosphate synthase; FDS: Farnesyl diphosphate synthase; GGDS: Geranylgeranyl diphosphate synthase; SQS: Squalene synthase; MJ: Methyl jasmonate; ORF: Open Reading Frame.

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 13 of 14

CSC0106 and CSC0130). Authors thank Dr. Dhanashekaran Shanmugam for helping in analyzing the transcriptome data.

References
1. Champagne DE, Koul O, Isman MB, Scudder GGE, Towers GHN. Biological-activity of limonoids from the rutales. Phytochemistry. 1992;31(2):377–94.
2. Tan QG, Luo XD. Meliaceous limonoids: chemistry and biological activities. Chem Rev. 2011;111(11):7437–522.
3. Jacobson M. Focus on Phytochemical Insecticides: The Neem Tree. Boca Raton: CRC Press; 1988.
4. Morgan ED. Azadirachtin, a scientific gold mine. Bioorg Med Chem. 2009;17(12):4096–105.
5. Schmutterer H. The Neem Tree: Source of Unique Natural Products for Integrated Pest Management, Medicine, Industry and Other Purposes. Weinheim, Germany: VCH; 1995.
6. Haldar S, Phapale PB, Kolet SP, Thulasiram HV. Expedient preparative isolation, quantification and characterization of limonoids from Neem fruits. Anal Methods. 2013;5(20):5386–91.
7. Siddiqui S, Siddiqui BS, Faizi S, Mahmood T. Tetracyclic triterpenoids and their derivatives from *Azadirachta indica*. J Nat Prod. 1988;51(1):30–43.
8. Ekong DEU, Ibiyemi SA, Olagbemi EO. The meliacins (limonoids). Biosynthesis of nimbolide in the leaves of *Azadirachta indica*. J Chem Soc Chem Commun. 1971;18:1117–8.
9. Ellis DI, Goodacre R. Metabolomics-assisted synthetic biology. Curr Opin Biotechnol. 2011;23(1):22–8.
10. Hall RD. Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol. 2006;169(3):453–68.
11. Hollywood K, Brison DR, Goodacre R. Metabolomics: Current technologies and future trends. Proteomics. 2006;6(17):4716–23.
12. Kueger S, Steinhauser D, Willmitzer L, Giavalisco P. High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. Plant J. 2012;70(1):39–50.
13. Nguyen Q-T, Merlo ME, Medema MH, Jankevics A, Breitling R, Takano E. Metabolomics methods for the synthetic biology of secondary metabolism. FEBS Lett. 2012;586(15):2177–83.
14. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. J Pharmaceut Biomed. 2013;87:12–25.
15. Johnson S, Morgan ED, Peiris CN. Development of the major triterpenoids and oil in the fruit and seeds of Neem (*Azadirachta indica*). Ann Bot - London. 1996;78(3):383–8.
16. Sidhu OP, Kumar V, Behl HM. Variability in triterpenoids (nimbin and salanin) composition of neem among different provenances of India. Ind Crop Prod. 2004;19(1):69–75.
17. Kuzuyama T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. Biosci Biotechnol Biochem. 2002;66(8):1619–27.
18. Vranova E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol. 2013;64:665–700.
19. Dewick PM. The biosynthesis of C5-C25 terpenoid compounds. Nat Prod Rep. 2002;19(2):181–222.
20. Thulasiram HV, Erickson HK, Poulter CD. A common mechanism for branching, cyclopropanation, and cyclobutanation reactions in the isoprenoid biosynthetic pathway. J Am Chem Soc. 2008;130(6):1966–71.
21. Thulasiram HV, Poulter CD. Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity. J Am Chem Soc. 2006;128(49):15819–23.
22. Thulasiram HV, Erickson HK, Poulter CD. Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. Science. 2007;316(5821):73–6.
23. Hill RA, Connolly JD. Triterpenoids. Nat Prod Rep. 2013;29(7):780–818.
24. Phillips DR, Rasbery JM, Bartel B, Matsuda SP. Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol. 2006;9(3):305–14.
25. Xu R, Fazio GC, Matsuda SPT. On the origins of triterpenoid skeletal diversity. Phytochemistry. 2004;65(3):261–91.
26. Roy A, Saraf S. Limonoids: overview of significant bioactive triterpenes distributed in plants kingdom. Biol Pharm Bull. 2006;29(2):191–201.
27. Kim OT, Bang KH, Jung SJ, Kim YC, Hyun DY, Kim SH, et al. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. Biologia Plantarum. 2010;54(1):47–53.
28. Kim Y-K, Kim YB, Uddin MR, Lee S, Kim S-U, Park SU. Enhanced triterpene accumulation in *Panax ginseng* hairy roots overexpressing mevalonate-5-pyrophosphate decarboxylase and farnesyl pyrophosphate synthase. ACS Synth Biol. 2014;3(10):773–9.
29. Kim OT, Kim SH, Ohyama K, Muranaka T, Choi YE, Lee HY, et al. Upregulation of phytosterol and triterpene biosynthesis in *Centella asiatica* hairy roots overexpressed ginseng farnesyl diphosphate synthase. Plant Cell Rep. 2010;29(4):403–11.
30. Johnson EE, Jetter R, Wasteneys G. Rapid induction of the triterpenoid pathway in *Arabidopsis thaliana* mesophyll protoplasts. Biotechnol Lett. 2014;36(4):855–8.
31. Lee MH, Jeong JH, Seo JW, Shin CG, Kim YS, In JG, et al. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. Plant Cell Physiol. 2004;45(8):976–84.
32. Mirjalili MH, Moyano E, Bonfill M, Cusido RM, Palazon J. Overexpression of the *Arabidopsis thaliana* squalene synthase gene in Withania coagulans hairy root cultures. Biologia Plantarum. 2011;55(2):357–60.
33. Seo JW, Jeong JH, Shin CG, Lo SC, Han SS, Yu KW, et al. Overexpression of squalene synthase in *Eleutherococcus senticosus* increases phytosterol and triterpene accumulation. Phytochemistry. 2005;66(8):869–77.
34. Krishnan NM, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. BMC Genomics. 2012;13(464):1471–2164.
35. Narnoliya LK, Rajakani R, Sangwan NS, Gupta V, Sangwan RS. Comparative transcripts profiling of fruit mesocarp and endocarp relevant to secondary metabolism by suppression subtractive hybridization in *Azadirachta indica* (neem). Mol Biol Rep. 2014;41(5):3147–62.
36. Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J. 2000;24(2):241–52.
37. Kulkarni R, Pandit S, Chidley H, Nagel R, Schmidt A, Gershenzon J, et al. Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit. Plant Physiol Biochem. 2013;71:121–31.
38. Rai A, Smita SS, Singh AK, Shanker K, Nagegowda DA. Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. Mol Plant. 2013;6(5):1531–49.
39. Wang G, Dixon RA. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci U S A. 2009;106(24):9914–9.
40. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008;105(38):14371–6.
41. Gennadios HA, Gonzalez V, Di Costanzo L, Li A, Yu F, Miller DJ, et al. Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. Biochemistry. 2009;48(26):6175–83.
42. Thabet I, Guirimand G, Courdavault V, Papon N, Godet S, Dutilleul C, et al. The subcellular localization of periwinkle farnesyl diphosphate synthase provides insight into the role of peroxisome in isoprenoid biosynthesis. J Plant Physiol. 2011;168(17):2110–6.
43. Pandit J, Danley DE, Schulte GK, Mazzalupo S, Pauly TA, Hayward CM, et al. Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. J Biol Chem. 2000;275(39):30610–7.
44. Inoue T, Osumi T, Hata S. Molecular cloning and functional expression of a cDNA for mouse squalene synthase. Biochim Biophys Acta. 1995;1260(1):49–54.
45. McKenzie TL, Jiang G, Straubhaar JR, Conrad DG, Shechter I. Molecular cloning, expression, and characterization of the cDNA for the rat hepatic squalene synthase. J Biol Chem. 1992;267(30):21368–74.
46. Gupta N, Sharma P, Santosh Kumar RJ, Vishwakarma RK, Khan BM. Functional characterization and differential expression studies of squalene synthase from *Withania somnifera*. Mol Biol Rep. 2012;39(9):8803–12.
47. Uchida H, Yamashita H, Kajikawa M, Ohyama K, Nakayachi O, Sugiyama R, et al. Cloning and characterization of a squalene synthase gene from a petroleum plant, *Euphorbia tirucalli* L. Planta. 2009;229(6):1243–52.

Pandreka *et al. BMC Plant Biology* (2015) 15:214

Page 14 of 14

48. Kribii R, Arro M, Del Arco A, Gonzalez V, Balcells L, Delourme D, et al. Cloning and characterization of the *Arabidopsis thaliana* SQS1 gene encoding squalene synthase-involvement of the C-terminal region of the enzyme in the channeling of squalene through the sterol pathway. Eur J Biochem. 1997;249(1):61–9.

49. Bhargava P, Kumar K, Chaudhaery SS, Saxena AK, Roy U. Cloning, overexpression and characterization of *Leishmania donovani* squalene synthase. FEMS Microbiol Lett. 2010;311(1):82–92.

50. Zhang D, Jennings SM, Robinson GW, Poulter CD. Yeast squalene synthase: expression, purification, and characterization of soluble recombinant enzyme. Arch Biochem Biophys. 1993;304(1):133–43.

51. Thompson JF, Danley DE, Mazzalupo S, Milos PM, Lira ME, Harwood Jr HJ. Truncation of human squalene synthase yields active, crystallizable protein. Arch Biochem Biophys. 1998;350(2):283–90.

52. Lee S, Poulter CD. Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongatus* BP-1. J Bacteriol. 2008;190(11):3808–16.

53. Alam A, Haldar S, Thulasiram HV, Kumar R, Goyal M, Iqbal MS, et al. Novel anti-inflammatory activity of epoxyazadiradione against macrophage migration inhibitory factor: Inhibition of tautomerase and proinflammatory activities of macrophage migration inhibitory factor. J Biol Chem. 2012;287(29):24844–61.

54. Haldar S, Mulani FA, Aarthy T, Dandekar DS, Thulasiram HV. Expedient preparative isolation and tandem mass spectrometric characterization of C-seco triterpenoids from Neem oil. J Chromatogr A. 2014;2014(31):1–14.

55. Davisson VJ, Woodside AB, Neal TR, Stremler KE, Muehlbacher M, Poulter CD. Phosphorylation of isoprenoid alcohols. J Org Chem. 1986;51(25):4768–79.

56. Jo Davisson V, Woodside AB, Dale Poulter C, John H. Law HCR. Synthesis of allylic and homoallylic isoprenoid pyrophosphates. Method Enzymol. 1985;110:130-44.

57. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.

58. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012;28(8):1086–92.

59. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:W182–5.

60. Wernersson R. Virtual Ribosome - a comprehensive DNA translation tool with support for integration of sequence feature annotation. Nucleic Acids Res. 2006;34:W385–8.

61. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2012;42(D1):D222–30.

62. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248–54.

63. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

64. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.