

# Comparative transcripts profiling of fruit mesocarp and endocarp relevant to secondary metabolism by suppression subtractive hybridization in *Azadirachta indica* (neem)

Lokesh K. Narnoliya · Raja Rajakani ·  
Neelam S. Sangwan · Vikrant Gupta ·  
Rajender S. Sangwan

Received: 19 May 2013 / Accepted: 16 January 2014 / Published online: 30 January 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** *Azadirachta indica* (neem) is a medicinally important plant that is valued for its bioactive secondary metabolites. Higher levels of the bioactive phytochemicals are accumulated in fruits than in other tissues. In the present study, a total of 387 and 512 ESTs, respectively, from endocarp and mesocarp of neem fruits were isolated and analyzed. Out of them 318 ESTs (82.17 %) clones from endocarp and 418 ESTs (81.64 %) from mesocarp encoded putative proteins that could be classified into three major gene ontology categories: biological process, molecular function and cellular component. From the analyses of contigs, 73 unigenes from the forward subtracted library and 35 unigenes from the reverse subtracted library were obtained. The ESTs from mesocarp encoded cytochrome P<sub>450</sub> enzymes, which indicated hydroxylation to be a major metabolic event and that biogenesis of hydroxylated neem fruit phytochemicals was differentially

regulated with developmental stage-specificity of synthesis. Through this study, we present the first report of any gene expression data in neem tissues. Neem hydroxy-methyl glutaryl-coenzyme A reductase (NHMGR) gene was used as expressing control *vis-a-vis* subtracted tissues. NHMGR was present in fruit, endocarp and mesocarp tissues, but absent in subtractive libraries, revealing that it was successfully eliminated during subtraction. Eight genes of interest from subtracted libraries were profiled for their expression in fruit, mesocarp and endocarp. Expression profiles validated the quality of the libraries and functional diversity of the tissues. The subtractive cDNA library and EST database described in this study represent a valuable transcript sequence resource for future research aimed at improving the economically important medicinal plant.

**Keywords** *Azadirachta indica* · Cytochrome P<sub>450</sub> · Legumin · Limonoids · Mesocarp and endocarp · Suppressive subtraction hybridization · Neem, plant secondary metabolism

**Electronic supplementary material** The online version of this article (doi:10.1007/s11033-014-3174-x) contains supplementary material, which is available to authorized users.

L. K. Narnoliya · R. Rajakani · N. S. Sangwan (✉) ·  
V. Gupta · R. S. Sangwan  
CSIR-Central Institute for Medicinal and Aromatic Plants  
(CIMAP), Kukrail Picnic Spot Road,  
P.O. CIMAP, Lucknow 226015, India  
e-mail: nss.cimap@gmail.com

R. S. Sangwan  
e-mail: sangwan.lab@gmail.com

**Present Address:**

R. S. Sangwan  
Center of Innovative and Applied Bioprocessing (CIAB), A  
National Institute under DBT (Govt. of India), C-127, Phase-8,  
Industrial Area, S.A.S. Nagar, Mohali 160071, Punjab, India

## Abbreviations

ACPs	Acyl carrier proteins
CTAB	Cetyl trimethyl ammonium Bromide
CYP	Cytochrome P <sub>450</sub>
EST	Expressed sequence Tag
GO	Gene ontology
IPS	Inter pro scan
KEGG	Kyoto encyclopedia of genes and genomes
NCBI	National Centre for Biotechnology Information
NHMGR	Neem 3-hydroxy-3-methyl glutaryl coenzyme A reductase
RT-PCR	Real time-polymerase chain reaction
RQ	Relative quantification

SSH	Suppressive subtraction hybridization
SSR	Short sequence repeat
UTR	Un translated region

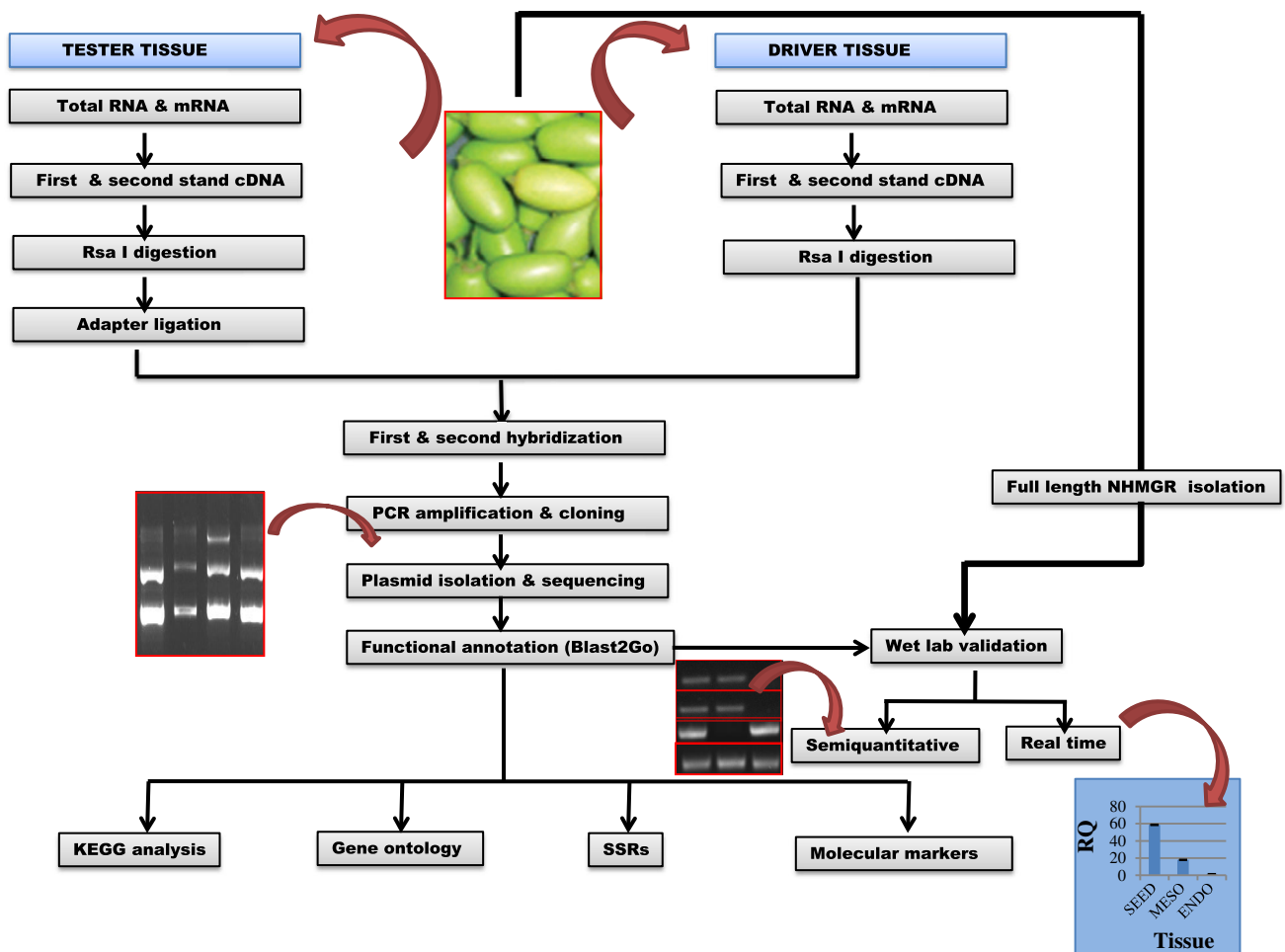
## Introduction

*Azadirachta indica* (neem) belongs to family Meliaceae (mahogany family) and is native to Indian sub-continent. The plant is highly valued for its medicinal and other bioactive properties. Phytochemical extracts from various part of the plant such as leaves, fruits, seeds, flowers, bark, stem and roots have been recognized to possess medicinal properties validating the value of its extensive use by human being in diverse traditional health and other practices [1]. Among different parts of the plant, fruits have been shown to have most prodigal amounts and diversified sets of phytochemicals. Fruits are globrous, olive-like drupes, which are green when young and with advancement of ripening stage progressively turn to be yellowish-red at fully ripened stage. Epicarp, mesocarp and endocarp of the fruit remain distinct althrough the stages of fruit growth and development. Epicarp of the fruit is thin that turns green to yellow with ripening and mesocarp is thick and changes from bitter taste white milky state to slightly fibrous-fleshy that becomes sweet with ripening. Endocarp contains one or more elongated seeds covered with a shell that is initially soft but turns to a white hard kernel. Applications of the neem products are very wide and records of beginning of use are dated to ancient ages in human history. It is used for medicines, pesticides, fuel wood, timber, food, hygiene etc. The most important application of neem is in formulation of bio-insecticides with azadirachtin as principal component responsible for the bioactivity. As such, neem is known to possess more than 300 complex secondary metabolites, a situation parallel to other medicinal plants of distinct repute like *Withania somnifera*, *Centella asiatica* [2–5]. Although these specialized phytochemicals of neem belong to several chemical classes such as terpenoids, alkaloids, phenolics, sulphur compounds, limonoids (metabolic descendants from triterpenoids) are the most abundant and diversified secondary metabolites of the plant. Terpenoids are synthesized with the recruitment of two distinct isoprenogenic routes at the upstream part of the biosynthetic pathway—mevalonate pathway and non-mevalonate pathway (methyl erythritol phosphate, MEP pathway). Accordingly, hydroxymethyl glutaryl coenzyme A reductase (HMGR), deoxy-xylulose synthase (DXS) and deoxy-xylulose reductase (DXR) are key enzymes which play an important role in governing limonoids biosynthesis at the

isoprenogenesis stage [6]. HMGR is key regulatory enzyme of cytosolic MVA pathway whilst DXS and DXR are the enzymes that play a regulatory role in the plastidic MEP pathway [2, 3, 6–10]. Most of these limonoid metabolites have one or more of the biomedical activities ranging from anti-feedant, growth regulating, anti-viral, antipyretic, antimalarial, anti-bacterial, anti-fungal, anti-inflammatory, to antiarthritic, hypoglycaemic, anti-ulcer, diuretic, anti-protozoal and anti-sickling properties [11–15]. Therefore, identification and characterization of genes involved in the biosynthesis of these important metabolites present in different tissues would be of immense importance. The information generated may be used to strategize the production of a particular metabolite using synthetic biology approach and/or enhance the production of a metabolites *in planta*. In order to address this issue, the relevant subsets of differentially expressed genes need to be identified.

Analysis of expressed sequence tags (ESTs) offers a rapid and cost effective approach to elucidate the transcriptome of an organism leading to identification of genes that could be involved in a specialized metabolic pathway. Successful examples of EST based gene discovery include the isolation of genes encoding enzymes that are involved in the secondary metabolites production, regulation and catabolism. Large scale EST sequencing and analyses have been reported on different tissues for genomic investigation of several important plants including *Arabidopsis*, rice and poplar etc. [16–18]. However, many of the genes, particularly non-housekeeping ones like those related to secondary metabolism, are expected to express only in specific tissues and at particular developmental stages. Therefore, suppressive subtraction hybridization (SSH) is a powerful technique for identification of such genes which are differentially expressed [19–21]. Recently, transcriptomes of fruits, leaves, flowers, stem and roots from neem were analysed which reveal that most of the secondary metabolites/metabolism group candidates were more predominantly present in the fruits [22].

Cytochromes P<sub>450</sub> oxido-reductases represent one of the biggest gene super families in plant genomes. Total estimated number of cytochrome P<sub>450</sub> genes is computed to be up to 1 % of the total gene annotations in each plant species. Plant cytochrome P<sub>450</sub> catalyze a wide variety of monooxygenation/hydroxylation reactions leading to the generation of diverse oxo-functionalized primary and secondary metabolites such as phenylpropanoids, alkaloids, terpenoids, lipids, cyanogenic glycosides, glucosinolates and plant hormones etc. [23]. Diversifications within cytochrome P<sub>450</sub> gene superfamilies have led to the emergence of new metabolic activities and pathways during the course of evolution and, thereby, to the new



**Scheme 1** Graphical presentation: schematic diagram of PCR-based suppression subtractive hybridization, functional analysis of ESTs and validation of libraries

member metabolites of the chemical groups as well as new chemical entities. Species-specific cytochrome P<sub>450</sub> families and members are one of the most important cues to the differential biosynthetic competences of plants and tissues and to the generation of species-specialized secondary metabolites [24].

Therefore, we have recruited the subtractive cDNA hybridization approach to generate a forward and reverse subtracted library with enriched transcripts for identification and analysis of differentially expressed genes in endocarp and mesocarp of neem fruit. A total of 387 clones from the forward subtracted library and 512 clones from the reverse subtracted library were identified. Annotations, contigs assembly and mapping of these clones were performed followed by validation of both the libraries by semiquantitative and quantitative real time PCR assays of the selected ESTs among them with focus on secondary metabolism (Scheme 1).

## Materials and methods

### Plant material

Mature fruits were collected from the tree maintained at the campus of CSIR-Central Institute of Medicinal and Aromatic Plant, Lucknow (India). Samples were immediately flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until use. The collection time was 3rd and 4th week of the month of May.

### Chemicals

Oligo-(dT)-cellulose for mRNA isolation was obtained from Sigma-Aldrich (USA). PCR Select hybridization kit was purchased from Clontech. All other kits and chemicals were procured from Promega, Applied Biosystems, Sigma, Fermentas and Himedia (India).

### Total RNA and mRNA extraction

Frozen fruits were taken out from liquid nitrogen, immediately immersed in RNA-Later solution and kept therein for 30 min at room temperature. Mesocarp and endocarp were separated by a sterilized scalpel and the separated tissues were immediately processed for RNA isolation. Total RNA was extracted by modified CTAB method [25]. RNA preparations were purified following DNase treatment and quality of RNA was ascertained by electrophoresis on 1.2 % agarose gel. RNA qualitative and quantitative analyses were performed by Nanodrop spectroscopy. Purification of poly (A)<sup>+</sup> mRNA from total RNA was carried out using oligo-(dT)-cellulose columns. Poly-(A)<sup>+</sup>-mRNA was used for PCR-Select cDNA subtraction.

### Suppression subtractive hybridization

Both forward and reverse subtraction libraries were prepared using PCR-Select cDNA subtraction kit (Clontech, USA) according to the manufacturer's instructions. For forward library, endocarp RNA was used as tester and mesocarp RNA was used as driver and *vice versa* for the reverse subtracted library. Tester and driver cDNAs were synthesized from 0.5 µg of mRNA extracted from mesocarp and endocarp tissues. Several rounds of suppressive subtractive hybridization were carried out to reduce the number of common transcripts. The resulting PCR products were subsequently ligated into pGEMT T vector using Easy "TA" cloning kit (Promega, USA).

### Library construction and sequencing

Electro-competent *Escherichia coli* (*E. coli*) DH5α cells were transformed with recombinant pGEM-T vector. Distinct white colonies (~1,000) were selected from each forward and reverse library for sequencing. Plasmids were isolated by alkaline lysis (Sambrook) method using 48 deep well blocks. The libraries were maintained in 96 well plates in two replicates, one set was used for sequencing and the other was stored at -80 °C. DNA sequencing was performed using a Big Dye Terminator Cycle Sequencing Ready Reaction Kit and an automated DNA sequencer (ABI Prism 3130XL; Applied Biosystems).

### Gene ontological (GO) analysis of differentially expressed genes

The resulting unique transcripts were subjected to Blast2GO tool [26] which searches against public protein database to produce BLASTx hits. Based on these hits, annotation and interproscan were performed. The results of

these searches were mapped to their corresponding GO accession and GO terms [5, 27].

### Phylogenetic analysis

Seed storage proteins were the most abundant gene in both the libraries therefore they were selected for phylogenetic analysis. Twenty-seven seed storage protein sequences of *Arabidopsis thaliana* (AAA32778.1), *Citrus sinensis* (AAB52963.1), *Populus trichocarpa* (XP\_002329472.1), *Ricinus communis* (XP\_002530185.1), *Pistacia vera* (ABG73109.1), *Gossypium herbaceum* (AEO27678.1), *Vernicia fordii* (AFJ04523.1), *Ficus pumila* (ABK80753.1), *Magnolia salicifolia* (CAA57846.1), *Elaeis guineensis* (AAF05770.1), *Vitis vinifera* (CBI37121.3), *Actinidia chinensis* (ABB77213.1), *Amaranthus hypochondriacus* (CAA57633.1), *Chenopodium quinoa* (ABI94736.1), *Carya illinoensis* (ABW86979.1), *Juglans regia* (AAW29810.1), *Castanea crenata* (AAM93194.1), *Quercus robur* (CAA67879.1), *Lupinus angustifolius* (AEB33709.1), *Glycine max* (ACT53400.1), *Pisum sativum* (S26688), *Anacardium occidentale* (AAN76862.1), *Vicia faba* (CAA32455.1), *Medicago truncatula* (XP\_003590688.1), *Asarum europaeum* (CAA64762.1), *Corylus avellana* (AAL73404.1) and *Sesamum indicum* (AAK15087.1) were downloaded from NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>). Phylogenetic and molecular evolutionary analyses were conducted [28] of selected plants with neem mesocarp and endocarp seed storage protein sequences.

### Isolation of putative NHMGR from fruit tissues

Degenerate primer set HMGRDGF and HMGRDGR amplified a 430 bp amplicon under following PCR cycling conditions: one cycle of 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 55 °C for 40 s and 72 °C for 2 min followed by a final extension of 72 °C for 7 min in a thermal cycler (Eppendorf). Amplicon was cloned into pJET1.2 vector (Fermentas) and then transformed into *E. coli* DH5α host strain. Cloned amplicons were sequenced using an automated DNA sequencer (ABI Prism 3130XL; Applied Biosystems). The nucleotide sequences obtained were analysed using the similarity search BLAST program and subsequently used for designing RACE primers.

### 5' and 3' RACE for isolation of full length NHMGR

To amplify putative 3' region of Neem hydroxy-methyl glutaryl-coenzyme A reductase (NHMGR), PCR was performed with a 3' RACE primer pair comprising of NHMGRDF1 and 3' AP under following conditions of PCR assay: 94 °C for 3 min, 35 cycles of 94 °C for 30 s,

**Table 1** List of primers from selected ESTs for gene expression validation

S.no.	Primer name (gene name)	Primer sequence
1	NM1CSPF (cytochrome P <sub>450</sub> )	CCCACCTGACAATGTCTTCC
2	NM1CSPR (cytochrome P <sub>450</sub> )	GACTTGAGAGGTGTAGTATAAGTGG
3	NM2CSPF (cytochrome P <sub>450</sub> )	TCCCGGAATAAGTAGAATAACG
4	NM2CSPR (cytochrome P <sub>450</sub> )	TGTCTACTGTCCGTTGTTCC
5	NM3CYPF (cytochrome P <sub>450</sub> )	GCATTTGGCTACCTTAAGAGAGTC
6	NM3CYPR (cytochrome P <sub>450</sub> )	GCTCACGCAATGTGATTTCTGC
7	NM4CYPF (cytochrome P <sub>450</sub> )	ACCTGACAATGTCTTCCG
8	NM4CYPR (cytochrome P <sub>450</sub> )	GACTTGAGAGGTGTGGTATAAGTG
9	NM5PEF (pectinesterase)	TGCAGACTGTCTGGACGAGATGG
10	NM5PER (pectinesterase)	TGCTAGGGCATTGCTGGTGACC
11	NE1CSPF (cytochrome P <sub>450</sub> )	AGTAGTGGTATTCACTTTCCG
12	NE1CSPR (cytochrome P <sub>450</sub> )	CGCATGAATGGATTAACGAGATCC
13	NE2ESF (embryo specific)	AAGAGCGGGCTGCTGTTTCATGC
14	NE2ESR (embryo specific)	CTGCCCCGTCTCTCGTCATCGCC
15	NE3ACF (acyl carrier)	TTCAGTCACGGGAACCTTCTGCC
16	NE3ACR (acyl carrier)	GAAGCGAGTGGGTGCAGACC
17	RTNHMGRF (HMGR)	ACAATGATGGAAGCCGTTA
18	RTNHMGRR (HMGR)	AACTGTGCCAACCTCAAT
19	LKCA3F (actin)	TTGCCGAGATAATGGCCTAC
20	LKCA3F (actin)	TACCGCGACTTCGATCTTTT

55 °C for 30 s and 72 °C for 1 min followed by a final extension of 72 °C for 7 min in a thermal cycler. First strand cDNA synthesis for 5'RACE was carried out using SMARTer RACE cDNA amplification kit (Clontech) according to manufacturer's instructions. Primary PCR product obtained in the reaction was used as template for the nested 5'RACE-PCR reaction in which NHMGRUR2 was used. Both initial and nested PCR reactions were carried under following conditions: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 55 °C for 50 s and 72 °C for 2 min followed by a final extension of 72 °C for 7 min in a thermal cycler. All PCR reactions were performed in a 50 µl assay volume containing 1.0 µl cDNA as template (except for secondary PCR where amplified products of primary PCR were used as template), 2 µl of 10 pmol respective primers and 45 µl master mix (34.5 µl PCR-grade water, 50 mM KCl, 4 mM MgCl<sub>2</sub>, 100 mM dNTPs, 2.5 U taq DNA polymerase). The nested amplified fragments of both 3' and 5' RACE were cloned in pJET1.2 cloning vector, transformed in DH5α cells and sequenced (ABI Prism 3130XL; Applied Biosystems).

#### Full length cloning of NHMGR

By comparing and aligning the sequences of the core fragments, 5' RACE and 3' RACE products, the full-length cDNAs of NHMGR were generated and subsequently

amplified with primers NHMFRFLF and NHMGRFLR containing BamHI and EcoRI restriction sites respectively. A high fidelity proof-reading DNA polymerase (Fermentas) was employed for amplification of full length gene under PCR conditions; 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 55 °C for 1 min and 72 °C for 3 min followed by a final extension of 72 °C for 7 min in a thermal cycler PCR. The resulted amplified products were ligated in pJET1.2 vector and sub-cloned into *E. coli* DH5α cells. The positive clones of NHMGR cDNA were confirmed by colony PCR and digestion with BamHI and EcoRI restriction enzymes. Sequences of all degenerate, RACE, full length and real time primers were listed in Table 1.

#### Validation of subtractive libraries by quantitative RT-PCR

First strand cDNAs of neem fruit, mesocarp and endocarp were synthesized using revert Aid<sup>TM</sup> cDNA synthesis kit (Fermentas). For cDNA synthesis, 5 µg total RNA isolated by modified CTAB method and treated with DNase (fermentas) sample were used from each tissue. Cytochrome P<sub>450</sub> and pectinesterase ESTs from fruit mesocarp and cytochrome P<sub>450</sub>, embryo specific protein and acyl carrier protein ESTs from fruit endocarp were used for expression. Quantitative PCR assays were performed using Step one<sup>TM</sup> Real time PCR system (Applied Biosystems) with absolute



SYBR green ROX mix (Applied Biosystems). The expression of the candidate genes was normalized against  $\beta$ -actin gene. Primers of target genes were designed from libraries according to their ESTs, with calculated  $T_m$  of 50–59 °C and amplification product not longer than 160 bp (Table 1). Quantitative real-time PCR was performed in a 20  $\mu$ l reaction volume, using above template, 10  $\mu$ l of SYBR Green master mix and 5 pmol of each primers. Reactions were conducted using the cycling conditions (95 °C for 10 s, 95 °C for 15 s and 60 °C 1 min for 40 cycles and for melting curve 95 °C for 15 s, 50 °C for 1 min and 95 °C for 15 s). After each reaction, which included a no-template control, dissociation curve analyses were carried out to verify the specificity of the amplification. All the real time PCR reactions were performed in triplicates and two biological repeats were performed. The relative gene expression levels were represented by relative quantification (RQ) values, which were calculated by  $2^{-\Delta\Delta C_t}$  method. Semi-quantitative PCRs were carried out in 20  $\mu$ l volume using first strand cDNA as template, 0.5 U Taq DNA polymerase and 10 pmole each primers under following set of PCR cycle conditions: one cycle of 94 °C for 3 min, 30 cycles of 94 °C for 30 s, 50–55 °C for 30 s and 72 °C for 1 min followed by a final extension of 72 °C for 7 min in a thermal cycler (Eppendorf).

## Results

### Construction of suppressive subtractive hybridization (SSH) cDNA libraries

Two SSH cDNA libraries from endocarp and mesocarp tissues of neem fruits were constructed. Two mRNA preparations were prepared from total RNA isolated from neem fruit mesocarp and endocarp tissues (Supplementary Fig. 1a, b). One mRNA population, isolated from endocarp tissue, containing targeted differentially expressed genes in forward library, was used as the tester and the second population of mRNAs, isolated from neem seed mesocarp constituted the driver, containing common transcripts in both populations that should be eliminated during the subtraction. After conversion of mRNA populations into double-stranded cDNA, the tester cDNA was subtracted by hybridization with driver cDNA and in reverse library *vice versa*. The resulting subtracted cDNA pool from each library that contain differentially expressed genes was further enriched (Supplementary Fig. 1c, d) and cloned into the pGEM-T Easy vector, generating subtracted cDNA libraries of more than 2,000 positive clones. These subtractive cDNA libraries, allowed the isolation of differentially expressed cDNAs from fruit mesocarp and endocarp.

The range of insert size obtained was 100 bp–1 kb for both libraries.

### EST sequencing and clustering

More than 2,000 recombinant clones from these libraries were selected for plasmid isolation (Supplementary Fig. 2a, b) and sequencing. Sequenced clones were analysed to identify secondary metabolism related ESTs. Total 387 significant ESTs from forward library (endocarp library) and 512 ESTs from reverse library (mesocarp library) with inserts size longer than 100 bp were selected. After trimming (removal of vector contamination, poor quality, ribosomal RNA and poly A regions) of the sequences, the average length of clones obtained were 309 and 343 bp for endocarp and mesocarp libraries, respectively. None of the ESTs showed any significant hits with any bacteria, animal and other genomes except with plant genomes/genes.

### Sequence annotation, gene ontology categorization and pathway analysis

ESTs of both libraries (387 from endocarp and 512 from mesocarp) were further analysed by BLAST2GO gene annotation tool for BLASTx, mapping and pathway analysis. BLASTx expect value set for analysis was  $1.0 \times 10^{-3}$ . From forward library 82.17 (318 clones) and 81.64 % (418 clones) ESTs from reverse library showed significant homology to proteins with known function and rest 17.83 % (69 clones) from forward library and 18.36 % (94 clones) from reverse library did not shown any significant similarity to any known proteins in public databases implying that these might be novel or less abundant ESTs.

In endocarp library, most of the ESTs were homologous to legumin and other seed storage proteins. Seed storage protein was the most abundant EST with 142 clones. Some other abundant ESTs were: legumin A (48 clones), legumin B (13 clones), senescence-associated protein (7 clones), embryo specific protein (7 clones), oleosin (5 clones) cytochrome P<sub>450</sub> protein (2 clones) and some allergen protein (22 clones). Some ESTs contained single clone like those for alanine-pyruvate aminotransferase, acyl carrier protein, vicilin-like antimicrobial peptides, succinate-semialdehyde mitochondrial-like protein, acyl binding protein, cell wall-associated partial, glyceraldehyde-3-phosphate dehydrogenase, soluble epoxide hydrolase and cell wall-associated hydrolase. Legumin A clone was the longest with 653 bp in endocarp library (Table 2). In mesocarp library, the most abundant ESTs were homologous to legumin and seed storage proteins. Seed storage legumin were the most abundant EST with 102 clones and with maximum length with 563 bp. Other abundant ESTs were

**Table 2** Neem forward library (endocarp) ESTs, EST description with top most blast hit, total clones of EST, EST length, mean E-value and mean similarity

S.no.	Sequence name	Sequence description	Number of clones	Sequence length	Minimum e-value	Mean similarity (%)
1	ScINSE-004-D11.b	Seed storage protein legumin partial	142	573	4.82e <sup>-65</sup>	73.65
2	ScINSE-002-H02.b	—NA—	71	515		
3	ScINSE-001-G05.b	Legumin A	48	653	2.98e <sup>-99</sup>	76.70
4	ScINSE-001-E08.b	11s globulin subunit beta	14	495	1.75e <sup>-39</sup>	65.45
5	ScINSE-004-B03.b	Legumin B	13	516	6.40e <sup>-68</sup>	84.35
6	ScINSE-001-F06.b	AF453947_1allergen ana o 2	9	333	1.24e <sup>-07</sup>	84.33
7	ScINSE-004-H08.b	Citritin	7	355	7.37e <sup>-07</sup>	70.05
8	ScINSE-002-C04.b	Legumin	7	357	1.30e <sup>-50</sup>	83.85
9	ScINSE-003-H04.b	11S legumin protein	6	347	5.07e <sup>-51</sup>	83.15
10	ScINSE-002-D02.b	Senescence-associated protein	6	249	3.55e <sup>-44</sup>	93.70
11	ScINSE-002-C08.b	Embryo-specific protein	5	381	4.28e <sup>-66</sup>	80.80
12	ScINSE-001-C08.b	Oleosin	5	430	5.10e <sup>-34</sup>	74.00
13	ScINSE-002-E10.b	Translationally controlled tumor protein	5	321	5.44e <sup>-42</sup>	92.35
14	ScINSE-002-D12.b	Pis v 2.0101 allergen11S globulin precursor	5	310	5.53e <sup>-34</sup>	75.20
15	ScINSE-004-H12.b	PREDICTED: uncharacterized protein LOC100246418	4	306	1.48e <sup>-58</sup>	85.40
16	ScINSE-003-D07.b	Pis v allergen 11s globulin precursor	3	286	4.89e <sup>-10</sup>	73.17
17	ScINSE-001-D09.b	Pis v 1 allergen 2s albumin	3	417	8.49e <sup>-09</sup>	65.00
18	ScINSE-003-B08.b	11s globulin	2	119	1.85e <sup>-10</sup>	88.50
19	ScINSE-001-D11.b	Oleosin 2	2	237	1.09e <sup>-13</sup>	84.80
20	ScINSE-001-H01.b	Pis v 2.0201 allergen 11S globulin precursor	2	460	3.26e <sup>-27</sup>	63.50
21	ScINSE-001-D03.b	Major latex	1	318	1.25e <sup>-12</sup>	70.40
22	ScINSE-001-B12.b	Cytochrome P <sub>450</sub> like_tbp	1	207	9.26e <sup>-40</sup>	94.85
23	ScINSE-001-G04.b	RNA-binding protein	1	263	5.13e <sup>-26</sup>	66.85
24	ScINSE-001-H04.b	Conserved domain protein	1	359	6.30e <sup>-70</sup>	74.20
25	ScINSE-001-B06.b	Beta-alanine-pyruvate aminotransferase	1	204	1.02e <sup>-21</sup>	78.60
26	ScINSE-001-H08.b	Ribosomal protein s2	1	248	3.27e <sup>-43</sup>	97.60
27	ScINSE-001-B09.b	Oleosin low molecular weight isoform	1	407	2.94e <sup>-14</sup>	84.50
28	ScINSE-001-C09.b	60S acidic ribosomal protein P1, putative	1	404	6.57e <sup>-32</sup>	78.35
29	ScINSE-001-E09.b	Chaperone protein	1	257	2.85e <sup>-22</sup>	94.20
30	ScINSE-001-G10.b	Mip tip subfamily	1	374	9.74e <sup>-43</sup>	83.95
31	ScINSE-001-B11.b	Translation factor	1	277	2.88e <sup>-26</sup>	94.75
32	ScINSE-001-C12.b	Hypothetical protein MTR_5g051050, partial	1	336	3.09e <sup>-36</sup>	84.80
33	ScINSE-001-E12.b	Elongation factor 1-	1	260	6.25e <sup>-51</sup>	96.70
34	ScINSE-001-F12.b	Acyl carrier protein	1	277	6.49e <sup>-12</sup>	64.87
35	ScINSE-002-G01.b	Vicilin-like antimicrobial peptides 2-3	1	120	8.26e <sup>-08</sup>	78.63
36	ScINSE-002-B04.b	Succinate-semialdehyde mitochondrial-like	1	179	1.36e <sup>-30</sup>	97.20
37	ScINSE-002-A06.b	Oleosin [ <i>Juglans regia</i> ]	1	129	3.95e <sup>-06</sup>	79.00
38	ScINSE-002-B08.b	Hypothetical protein MTR_4g091430 [ <i>Medicago truncatula</i> ]	1	202	1.34e <sup>-11</sup>	84.00
39	ScINSE-002-F08.b	Acyl- binding protein	1	325	4.07e <sup>-54</sup>	94.95
40	ScINSE-002-D10.b	Cell wall-associated partial	1	119	3.28e <sup>-07</sup>	63.78
41	ScINSE-003-B04.b	2s albumin	1	419	1.13e <sup>-10</sup>	69.00
42	ScINSE-003-E11.b	Glyceraldehyde-3-phosphate dehydrogenase	1	185	1.23e <sup>-31</sup>	96.00
43	ScINSE-003-F11.b	Low quality protein: protein	1	412	2.70e <sup>-27</sup>	64.05
44	ScINSE-004-A02.b	Class ii heat shock protein	1	312	2.12e <sup>-11</sup>	82.20
45	ScINSE-004-D07.b	Soluble epoxide hydrolase	1	383	2.31e <sup>-12</sup>	88.85
46	ScINSE-004-A12.b	Cell wall-associated hydrolase	1	394	7.70e <sup>-21</sup>	71.35
47	ScINSE-001-G07.b	AF240004_111s globulin	1	300	1.86e <sup>-09</sup>	58.75
48	ScINSE-003-A05.b	Pis v 5.0101 allergen 11S globulin precursor	1	284	3.85e <sup>-31</sup>	80.40

legumin A (23 clones), legumin B (73 clones), senescence-associated protein (70 clones), cytochrome P<sub>450</sub> (45 clones), oleosin isoform (13 clones), metalloproteinase inhibitor (6 clones) and some globulin (6 clones). Some ESTs contained single clone like, pectinesterase, citrin protein, calcium binding protein, allergen, albumin and globulin and some cytochrome clones (Table 3).

In endocarp library, total annotations of ESTs were 390 while those in mesocarp library were 443. Based on these annotations, members of both libraries were analysed in terms of three major GO categories; biological process, molecular function and cellular component. In both libraries, maximal number of the annotated ESTs grouped under GO category of molecular function. In endocarp library, more number of ESTs were covered under 2–6 GO level and in mesocarp library under 2–9 GO level (Supplementary Fig. 3a, b). There were more number of transcripts in endocarp library that categorized to post embryonic development, protein metabolic process and regulation of cellular process. For mesocarp library, it was so the category of metabolic and cellular process (Fig. 1a, b). In endocarp library, most transcripts pertained to binding and catalytic activity whereas in mesocarp library, the phenomenon of maximal representation was associated with genes involved in RNA binding activity, hydrolase activity and transferases activity (Fig. 1c, d). In endocarp library, maximal ESTs were assigned to cell protoplasm and organelles and only some to extracellular region and macromolecular complex. In mesocarp, a large number of ESTs were assigned to the intracellular region group and others were assigned to membranous group such as vesicles (Fig. 1e, f).

In endocarp library, 309 ESTs were with IPS, 78 ESTs were without IPS and about 230 were with GOs. Similarly in mesocarp library, 259 ESTs were with IPS, 253 ESTs were without IPS and about 200 ESTs were with GOs. 240 ESTs from endocarp library and 280 ESTs from mesocarp library could be fully annotated (Supplementary Fig. 4a, b). *Pistacia vera* and *Citrus sinensis* emerged as the most hit plant species for the neem endocarp ESTs whereas *Pistacia vera* and *Populus trichocarpa* were found to be most hit plant species with respect to matching of neem fruit mesocarp specific ESTs (Supplementary Fig. 5a, b). In terms of metabolic pathway assignment of the genes, five ESTs from endocarp library and 16 ESTs from mesocarp library could be assigned to specific steps in KEGG pathway (Supplementary Table 1). In endocarp library, genes for nine enzymes involved in eight different pathways were recognized (Supplementary Table 1) whereas in transcript set of mesocarp library genes of three enzymes involved in 15 different pathways were recognized (Supplementary Table 2).

In endocarp library, 49 ESTs contained SSR pertaining to the sub-groups of single, di- and tetra-nucleotides. Among them, eight were single nucleotide (A-5, T-3), 40 were dinucleotide (AT-10, GT-7, CT-13, AG and TA-1 each) and one was tetranucleotide (CATG). Average length of SSRs was 19 bases and maximum length was 36 bases. In mesocarp library, a total of 63 ESTs contained SSRs that were of two types- mono- and di-nucleotides. Among them, eight were single nucleotide type (T-7, A-1) and 55 were dinucleotide type (AG-11, AC-27, CT-8, TG-6, TA, GT and AT-1 each). Average length of the SSRs was 20 bases and maximum length was 35 bases.

For endocarp and mesocarp specific ESTs contig assembly and analysis, 15 assembled contigs and 58 singlets for endocarp and 11 assembled contigs and 24 singlets for mesocarp were obtained. Sequence lengths obtained after contigs assembly varied from 100 bases to about 1.8 kb with average length of 400 bases. In the endocarp library, maximum length of contig was 1.746 kb and in mesocarp it was 1.284 kb. The general features of both the libraries are presented in Table 4.

#### Phylogenetic analysis

A phylogenetic tree was constructed from aligned seed storage protein sequences of different plants and neem mesocarp and endocarp seed storage protein sequences from subtractive libraries as shown in Fig. 2.

#### Secondary metabolite metabolism related genes

The amino acid sequence of polypeptides encoded by the neem mesocarp and endocarp ESTs showed significant sequence homology with the genes related to metabolism of secondary phytochemicals. Interestingly, mesocarp contained more cytochrome P<sub>450</sub> function related genes as compared to endocarp. Five cytochrome P<sub>450</sub> ESTs (ScINSM-001-G02, ScINSM-003-C06, ScINSM-004-A05, ScINSM-005-E12 and ScINSM-006-A02) from mesocarp library and one cytochrome P<sub>450</sub> EST (ScINSE-001-B12) from endocarp library were identified.

#### Characterization of full length NHMGR

After assembling the sequences of 5' and 3' RACE products, the sequence of assembled full length cDNA when subjected to Blastx search, a very high homology with HMGR genes from several plant species such as *Dimocarpus longan* (86 %), *Litchi chinensis* (86 %), *Ricinus communis* (82 %) and *Hevea brasiliensis* (82 %) was obtained. This suggested that the cDNA was a NHMGR gene. The full-length cDNA was 2.233 kb with 5' and 3'



**Table 3** Details of ESTs from neem reverse library (mesocarp)

S.no.	Sequence name	Seq. description (top blast hit)	No of clones	Sequence length	E-value	Mean similarity (%)
1	ScINSM-001-A02.b	Seed storage protein legumin partial	102	563	6.01e <sup>-37</sup>	65.80
2		NA	94			
3	ScINSM-001-H04.b	Legumin B	73	440	1.15e <sup>-69</sup>	83.60
4	ScINSM-004-B05.b	Senescence-associated protein	70	295	2.64e <sup>-56</sup>	98.10
5	ScINSM-003-C06.b	Cytochrome P <sub>450</sub> like_tbp (Nicotiana)	41	539	3.61e <sup>-57</sup>	84.70
6	ScINSM-004-H11.b	Hypothetical protein MTR_5g051130	37	476	3.47e <sup>-39</sup>	81.65
7	ScINSM-002-C11.b	Legumin A	23	523	1.16e <sup>-46</sup>	67.60
8	ScINSM-002-A01.b	Elongation factor 1-alpha	13	240	4.06e <sup>-11</sup>	59.80
9	ScINSM-004-G07.b	Oleosin low molecular weight isoform	13	407	8.15e <sup>-16</sup>	89.70
10	ScINSM-002-D11.b	11s legumin protein	10	338	1.40e <sup>-23</sup>	72.60
11	ScINSM-001-C03.b	Legumin	9	379	2.69e <sup>-43</sup>	81.95
12	ScINSM-002-D09.b	Metalloprotease inhibitor	6	317	4.37e <sup>-31</sup>	85.70
13	ScINSM-004-D08.b	11s globulin	4	261	1.68e <sup>-11</sup>	81.65
14	ScINSM-003-D02.b	Oleosin 2	3	399	1.44e <sup>-13</sup>	89.53
15	ScINSM-001-F10.b	Calmodulin 24-like protein	2	158	8.94e <sup>-18</sup>	78.85
16	ScINSM-002-E06.b	11s globulin subunit beta	2	255	5.68e <sup>-09</sup>	63.90
17	ScINSM-001-G02.b	Cytochrome P <sub>450</sub> like TBP (Medicago)	1	291	1.45e <sup>-24</sup>	92.45
18	ScINSM-004-A05.b	Cytochrome P <sub>450</sub> like TBP Protein (Lilium)	1	147	1.01e <sup>-12</sup>	73.55
19	ScINSM-005-E12.b	Cytochrome P <sub>450</sub> like TBP (Citrus)	1	438	1.30e <sup>-28</sup>	89.10
20	ScINSM-006-A02.b	Probable cytochrome P <sub>450</sub> monooxygenase	1	369	5.27e <sup>-45</sup>	91.05
21	ScINSM-005-A09.b	Pectinesterase	1	377	1.18e <sup>-42</sup>	85.60
22	ScINSM-002-H01.b	Calcium-binding protein cml27	1	460	2.02e <sup>-32</sup>	80.60
23	ScINSM-003-B10.b	Citrus	1	215	2.48e <sup>-06</sup>	56.33
24	ScINSM-001-A04.b	Pis v 1 allergen 2s albumin	1	235	1.87e <sup>-08</sup>	64.00
25	ScINSM-003-C11.b	Pis v allergen 11s globulin precursor	1	256	2.91e <sup>-26</sup>	74.00
26	ScINSM-003-A07.b	Predicted protein [ <i>Populus trichocarpa</i> ]	1	318	4.21e <sup>-19</sup>	60.50

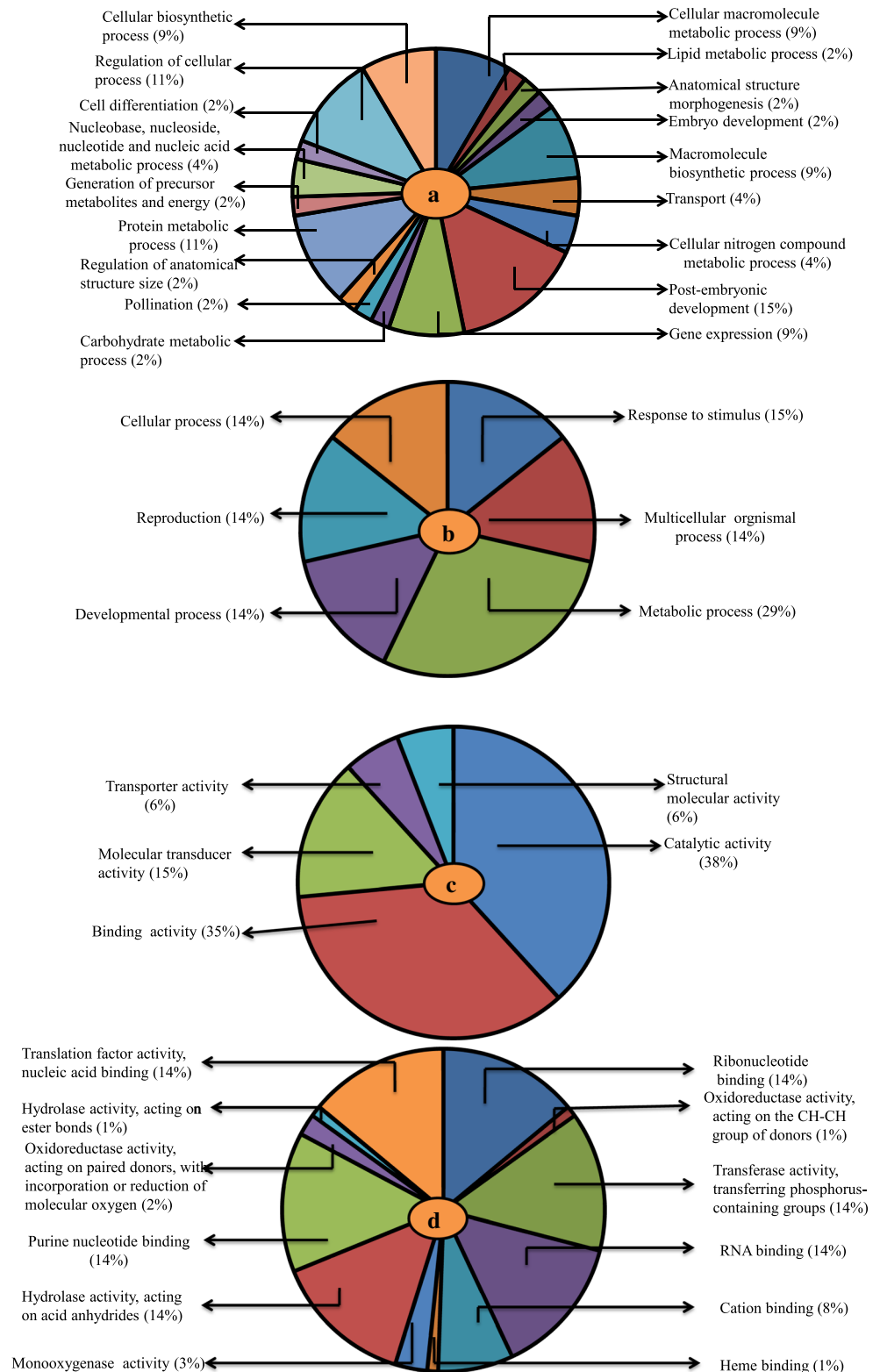
untranslated regions (UTRs) and a poly A tail. The ORF search result showed that NHMGR contained a 1.749 kb-ORF encoding a 583-amino-acid protein. The predicted NHMGR protein had a theoretical pI value of 9.11 and molecular mass of 62.01 kDa. The sequence analysis revealed that the functional motifs of NHMGR were very similar to those contained by other plant HMGRs. The highest homology regions were observed to be around the substrate binding sites. Plant HMGRs comprised of two HMG-CoA binding motifs (EMPIGYVQIP and TTEGCLVA) and two NADPH-binding motifs (DAMGMNM and GTVGGGT) [6–10, 29]. All the four motifs were also present in the NHMGR (Fig. 3). It was also observed that there was a 5' UTR of 262 bases upstream from the start codon and the coding region of NHMGR at 3' end was followed by a 222 bases long 3' UTR. Two transmembrane domains could be predicted to be present in NHMGR. It was further revealed by the analysis that one of these transmembranes regions was located between Pro31 (P) and Leu51 (L) and the other was located between

Gly74 (G) and Val 96 (V) along the polypeptide chain. On analysis of relative variability in the locale of sequence divergence/homology compared to HMGRs from other plants, it was noted that the amino acid sequence divergence in NHMGR was more predominant in length and composition in the N-terminus region, especially at two locations-amino acid 1–30 and 100–160. NHMGR was present in all three tissues, i.e. fruits, endocarp and mesocarp as shown by semi-quantitative PCR and real time PCR but it was absent in subtractive endocarp and mesocarp libraries (Fig. 4).

#### Validation of forward and reverse subtractive libraries by quantitative RT-PCR

To validate quality of the forward and reverse subtracted libraries, the expression of eight putative genes, five from mesocarp (four cytochrome P<sub>450</sub> genes and one pectinesterase gene) and three from endocarp (one cytochrome P<sub>450</sub> gene, an embryo specific and an acyl carrier protein gene)

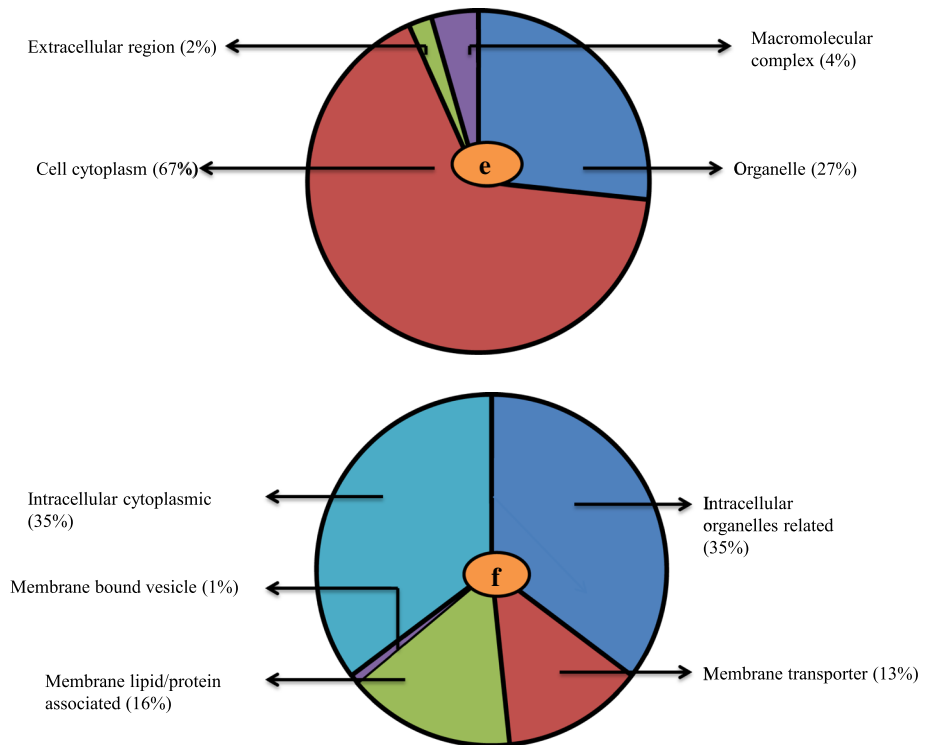
**Fig. 1** Gene ontology (GO) term-based functional categorization of endocarp and mesocarp ESTs, obtained by suppressive subtractive hybridization, based on biological process (1a, b); molecular function (1c, d) and cellular component (1e, f) respectively. The combined graph was generated based on ontology level 4 by B2G



were compared by semi-quantitative and real time PCR assay in whole fruit, mesocarp and endocarp tissues. Relative expression of the genes was analysed by  $\Delta\Delta CT$

methods (relative quantification) by real time PCR and on electrophoresis through agarose gel in case of semi-quantitative PCR (Fig. 5a, b). Putative cytochrome P<sub>450</sub> genes

Fig. 1 continued



and pectinesterase gene were found to be expressed in neem fruit mesocarp as well as in whole fruit whilst in endocarp it was noted to be down regulated. Relative mRNA abundance of endocarp specific cytochrome P<sub>450</sub> genes, embryo-specific gene and acyl carrier protein were observed to be more in fruit endocarp.

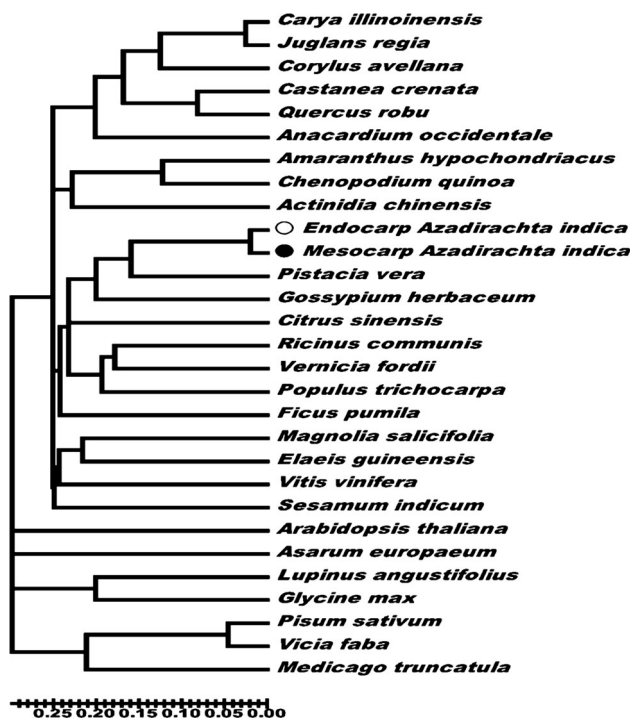
## Discussion

The reliability of the cDNA subtraction libraries has been tested by checking subtraction results for NHMGR, which was used as control in the subtraction based study. Full length cDNA sequence of NHMGR was isolated from fruit, endocarp and mesocarp tissues. Real time results showed that NHMGR gene was substantially expressed in both the tissues but the level of expression was higher in endocarp compared to mesocarp. As expected in the subtraction results, complete extinction of the constitutively expressed NHMGR gene in endocarp and mesocarp tissue implied that it was completely eliminated during subtraction of libraries in forward and reverse mode. This ensured that both the libraries were pure and did not contain any cross contaminating EST. Although a large number of metabolites are reported to be present in different parts of neem, fruits of the plant are known to be the major site for the presence of most of the compounds particularly limonoids [1, 25]. Fruit mesocarp and endocarp contained more than

**Table 4** An account of endocarp and mesocarp SSH libraries

S. no	Parameter	Endocarp	Mesocarp
1	Total no of sequence	387	512
2	Average sequence length	310	350
3	Homology to known proteins	82.17 % (318 seq)	81.64 % (418 seq)
4	Unknown ESTs	69	94
5	Number of clone of most abundant EST	142	102
6	Maximum length of EST	653	563
7	Type of ESTs involved in KEGG pathway	5	16
8	ESTs with SSRs	49	63
9	Unique transcripts	53	25
10	Cytochrome ESTs	1	5

80 % of the metabolites. Despite these metabolites of neem being commercially important and having diverse applications in practice, secondary metabolism of neem plant has not been explored. Further, there is no gene expression data available from the plant [22, 25]. Recently, transcriptomes of the fruit, stem, leaf, flower and root were reported but there is no report on differential expression of the genes related to secondary metabolism [30]. In this study, we have identified several differentially expressing genes from mesocarp and endocarp tissues of neem fruits



**Fig. 2** Phylogenetic relationship of seed storage proteins from *A. indica*. The phylogenetic tree was constructed using MEGA 5 version (maximum likelihood method)

by suppression subtraction hybridization which is involved in secondary metabolite biosynthesis. All the ESTs were unique and were associated with only particular tissue of neem fruits. As in other plants, neem seeds (endocarp) are also the site for food storage and that's why most of the ESTs from endocarp were related to seed storage protein such as legumin A, legumin B, citrin and allergen proteins that are accumulated during seed development to serve as carbon and nitrogen source during the follow-up seed germination phase. Several ESTs from mesocarp library were also related to storage legumin proteins because it is attached with endocarp and also contributes to seed development. As embryo is specifically located within endocarp, embryo related ESTs were present in only endocarp and were not detected in mesocarp.

Mapping of the mesocarp and endocarp transcripts correlated well with the physiological state of neem fruit. Embryonic development takes place in endocarp of neem fruit so embryonic related proteins are actively expressed in the tissues related with metabolic processes. The BLASTx hits that could not be assigned any GOs term are putatively tissue specific transcripts of neem fruits which need to be identified and validated by other experimental approaches. At least some of them could be useful in terms of their role in specific secondary metabolite biogenesis process/step(s) of the plant. *Pistacia vera* and *Citrus sinensis* are the most blast hit species because they are most

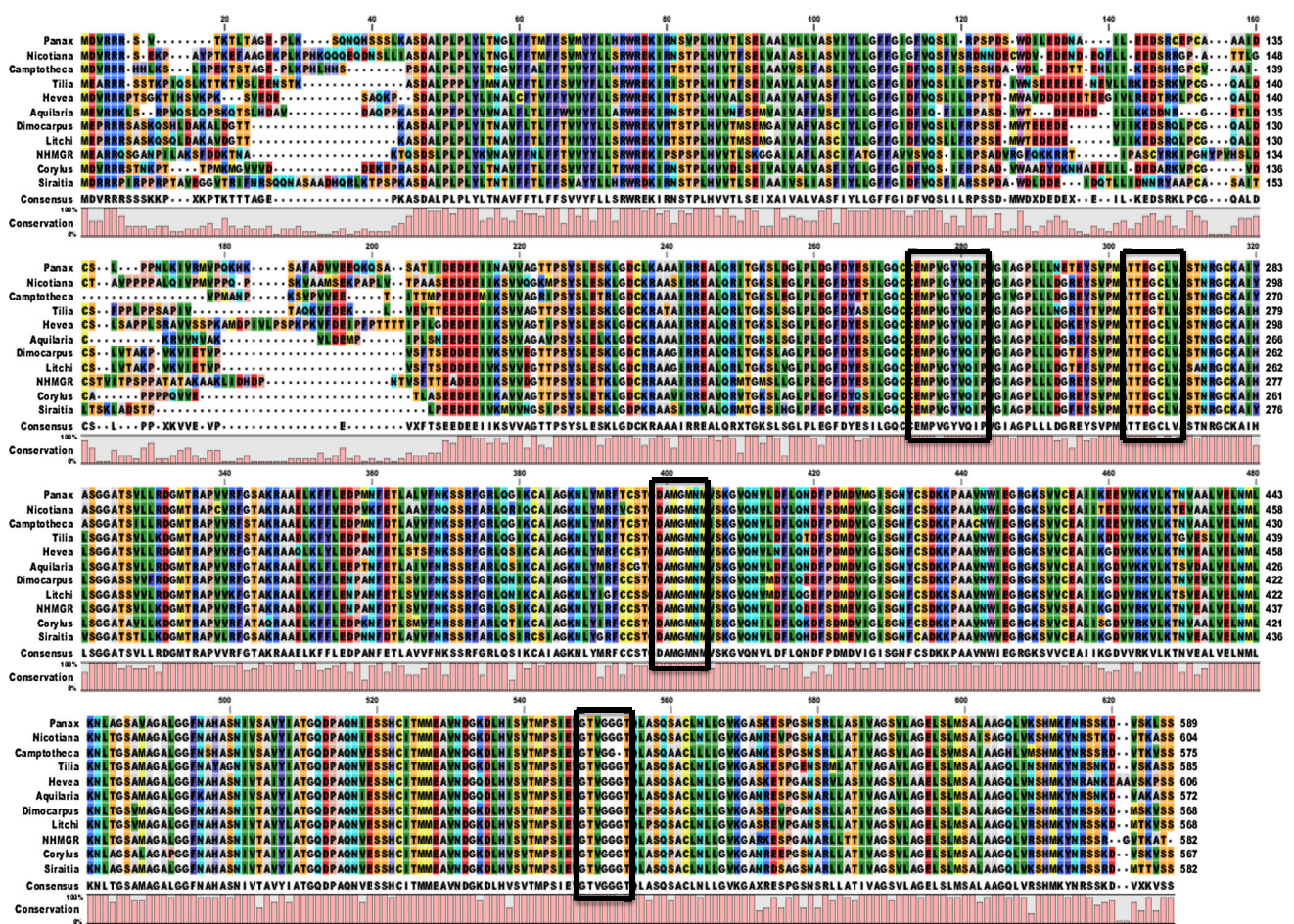
closely related to neem based on phylogenetic proximity under the order Sapindales/Rutales. Allergen and globulin mRNAs appear to be the most expressed transcripts in both the libraries but the number of ESTs was far greater in endocarp library. These proteins also fall under seed storage proteins and they accumulate during the process of seed development. Nevertheless, their observed presence in mesocarp is a moot point in terms of the fruit-seed physiology. A prediction of a share of mesocarp metabolism to the seed storage proteins need to be examined in future studies. These transcripts could also be ones remnant from the very early fruit development when physical separation of mesocarp from endocarp is least.

Of the clones that yielded high quality sequences and had inserts, major percentages (31.5 % from the forward subtracted library and 25.7 % from the reverse subtracted library) were unidentifiable on BLAST searches and comprised of novel and unclassified genes. Phylogenetic proximity based on most abundant EST seed storage protein of neem mesocarp and endocarp shows that it is closely related to *Pistacia vera* and *Citrus sinensis*. These plants *Citrus*, *Pistacia* belong to same order Sapindales and Rutaceae family to which *Citrus* is considered phylogenetically closest to neem family (Meliaceae). This proximity is also manifested in terms of their secondary metabolism predominated by limonoids group of terpenoidal phytochemicals.

Member diversity of neem compounds could be most prolifically diversified by secondary transformations involving mainly oxido-reductive and subsequent conjugative steps at the appropriate hydroxy-positions. Plant cytochromes P<sub>450</sub> enzymes catalyze a wide variety of monooxygenation/hydroxylation reactions to generate diversity in parental secondary metabolites pertaining to all chemical classes viz. phenylpropanoids, alkaloids, terpenoids etc. Thus, cytochrome P<sub>450</sub> hydroxylases play a lead role in metabolic processing of early progenitor intermediate from main pathway to facilitate biogenesis of several end products of the respective class of metabolites [23, 24, 31]. Tissue specific P<sub>450</sub> families are recruited for biosynthetic pathways of specialized metabolites. Neem contains many biologically active compounds like azadiradione, epoxyazadiradione, azadirachtin, salannin, nimbocinol, catechin, gallic acid etc. that were characteristically and/or multi-fold hydroxylated. The secondary metabolism related cytochrome P<sub>450</sub> sub-family members are likely to have a designated catalytic role in facilitating such specific hydroxylase/epoxidase catalyzed metabolic transformations [30].

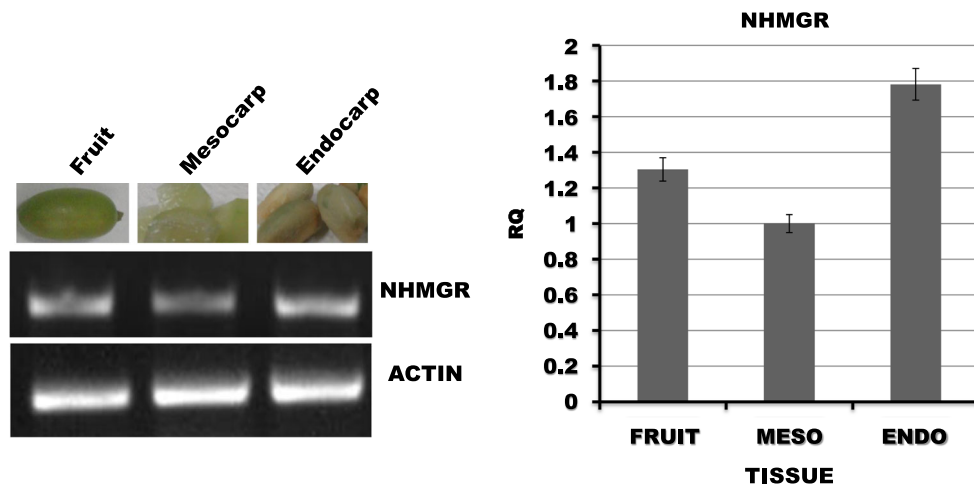
Real time and semi-quantitative expression analysis of cytochrome P<sub>450</sub> genes have revealed that cytochrome P<sub>450</sub> genes were up-regulated in a tissue-specific manner in different parts of the neem fruit. Occurrence of more





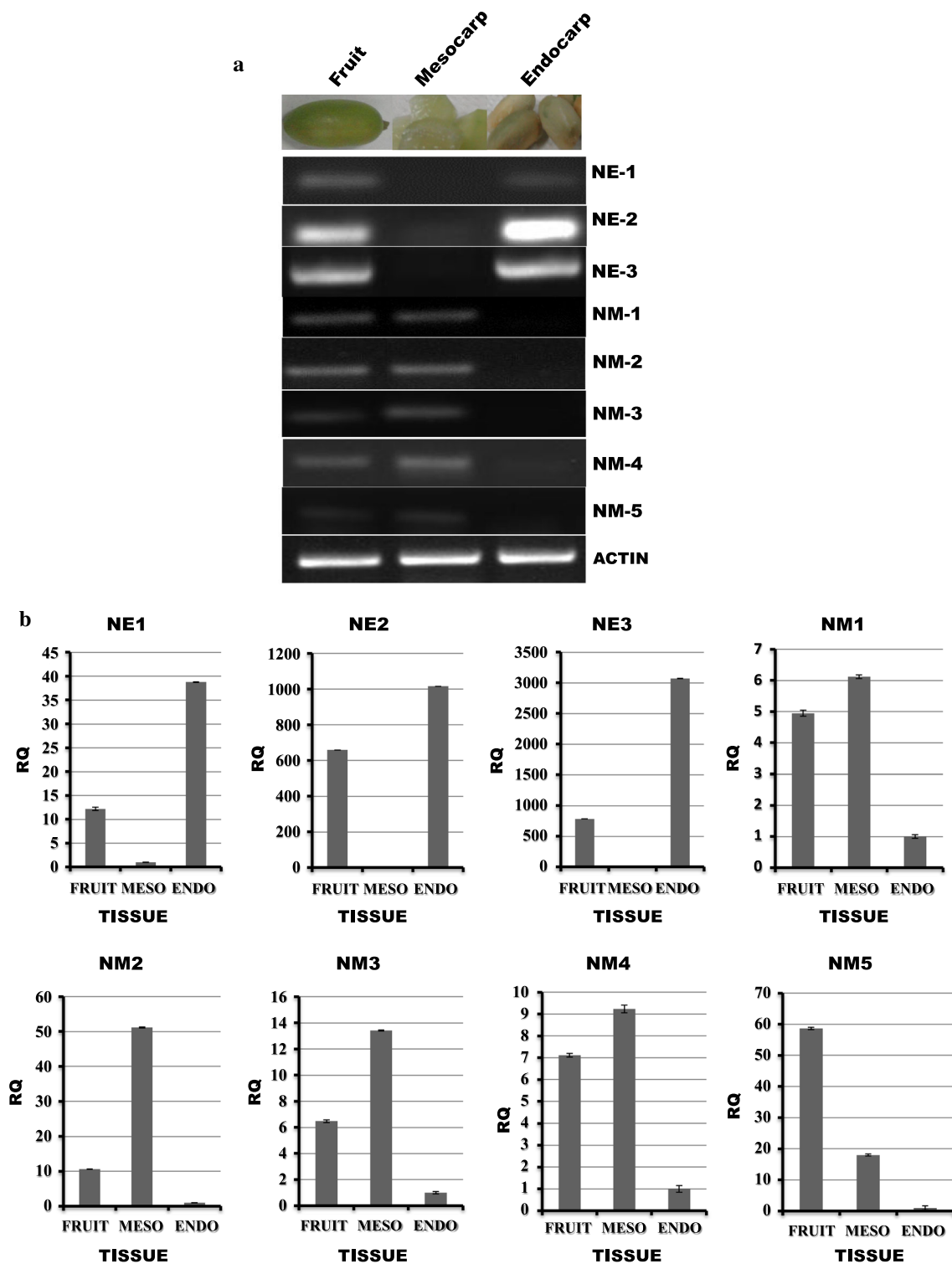
**Fig. 3** Multiple alignment of the deduced amino acid sequences of NHMGR and other plant HMGRs from *Panax quinquefolium* (Accession No. ACV65036.1), *Nicotiana tabacum* (Accession No. AAL54879.1), *Camptotheca acuminata* (Accession No. AAB69726.1), *Tilia miqueliana* (Accession No. AAY68034.1), *Hevea brasiliensis* (Accession No. BAF98281.1), *Aquilaria sinensis*

(Accession No. AFU75319.1), *Dimocarpus longan* (Accession No. AET72045.1), *Litchi chinensis* (Accession No. AET72043.1), *Corylus avellana* (Accession No. ABP04052.1), *Siraitia grosvenorii* (Accession No. AEM42971.1). Two putative HMG-CoA-binding sites (EMPIGVQIP and TTEGCLVA) and two NADPH-binding sites (DAMGMNM and GTVGGGT) are indicated within box



**Fig. 4** Expression profiles of NHMGR gene through semi quantitative and real time PCR in fruit, mesocarp and endocarp tissues. NHMGR was used as control gene which were expressed in fruit and unsubtractive endocarp and mesocarp libraries but absent in both subtractive libraries





**Fig. 5** Wet lab validation of the forward and reverse subtracted libraries from fruit, mesocarp and endocarp tissues of *A. indica* through **a** semi-quantitative PCR, **b** Quantitative real time PCR. Eight genes, three from endocarp; cytochrome P<sub>450</sub> (NE-1), embryo specific

(NE-2) and acyl carrier (NE-3) and five from mesocarp; cytochrome P<sub>450</sub> (NM-1, NM-2, NM-3 and NM-4) and pectin esterase (NM-5) were selected for expression analysis. Neem mesocarp (NM) and neem endocarp (NE)

member cytochrome P<sub>450</sub> genes in the mesocarp compared to endocarp tissue matches with the occurrence of a more complex repertoire of hydroxylated metabolites (limonoids) as synthesized and/or terminally processed for their likely subsequent fate of transportation to sequestration site(s) including endocarp. Neem green (immature) fruit is bitter in taste but after ripening its green epicarp gets converted to yellow coloration and hard white milky mesocarp transforms into flashy yellow sweet mesocarp. This may, besides other metabolic processes, also entails a limonoid transformation as similar process reversely occurs in citrus fruits juices on standing due to limonin (bitter) release from its glyco-conjugated (sweet) form. Pectinesterase plays an important role in cell wall metabolism during fruit ripening and, therefore, observed up-regulation of pectinesterase in mesocarp tissue might play an important role in ripening of neem fruit like other fruits. As such, its preponderance of expression in whole fruit indicates contribution of expression in other parts like epicarp of the fruit during ripening. It may bring about the observed loosening of the epicarp on maturity (acquiring and accessing sweet taste mesocarp). This facilitates seed dispersal by attracting crows as mature neem fruit is a feast for them. If sweetness is linked with the limonoid metabolism, probably neem forms a unique case of so called secondary (specialized) metabolites linked to seed dispersal- a primary activity essential for normal conclusion of reproductive process. Acyl carrier proteins (ACPs) are involved in both the pathways of primary and secondary metabolism as they transfer acylated intermediates/metabolites to their respective sites for processing/use. Several ACPs are found in different tissues of neem as its seed (kernel) contains abundant (~25–45 %) triacylglycerol. The observed up-regulated expression observed in the endocarp might also reflect secondary metabolite modification, in addition to the classical predominant ACPs involved in fatty acid synthesis.

## Conclusions

In the present report two cDNA libraries neem fruit endocarp and mesocarp were constructed by suppression subtraction hybridization approach. A total 387 ESTs from endocarp and 512 ESTs from mesocarp libraries were identified and analyzed. BLASTX data revealed that, 318 ESTs (82.17 %) from endocarp and 418 ESTs (81.64 %) from mesocarp encoded for polypeptides/proteins (known or putative proteins). Contigs assembly provided 73 unique clones from the forward subtracted library and 35 unique clones from the reverse subtracted library. Mesocarp contained more cytochrome P<sub>450</sub> genes than endocarp which reveals more hydroxylated fruit metabolites were synthesized or end

processed in mesocarp tissues. NHMGR was used as control, which was present in fruit, endocarp and mesocarp tissues, and absent in both subtractive libraries ascertaining the desired completion of the both the libraries. Further investigation of some of the selected ESTs 3 from endocarp (CYP, ACP and embryo specific protein) and five from mesocarp (four CYP and pectinesterase) revealed that these genes are expressed differentially in their respective tissues. Phylogenetic proximity based on most abundant EST seed storage protein of neem mesocarp and endocarp shows that it is closely related to *Pistacia vera* and *Citrus sinensis*. The results have manifested in interesting feature of tissue-specificity and insightful aspects of the neem secondary metabolism for further research including invoking secondary metabolites and metabolism in primary activity (chemical biology of seed dispersal).

**Acknowledgments** This work was financially supported by Council of Scientific and Industrial Research, Govt. of India, under the Grant HCP002. LKN is thankful to University Grants Commission, Govt of India, for financial assistance in form of senior research fellowship. The authors wish to express their sincere thanks to the Director, CSIR-CIMAP, for encouragement and providing necessary facilities.

## References

1. Atawodi SE, Atawodi JC (2006) *Azadirachta indica* (neem): a plant of multiple biological and pharmacological activities. *Phytochem Rev* 8:601
2. Chaurasiya ND, Sangwan NS, Sabir F, Misra LN, Sangwan RS (2012) Withanolide biosynthesis recruits both mevalonate and DOXP pathways of isoprenogenesis in ashwagandha *Withania somnifera* L. (Dunal). *Plant Cell Rep* 31(10):1889–1897
3. Sabir F, Mishra S, Sangwan RS, Jadaun JS, Sangwan NS (2012) Qualitative and quantitative variations in withanolides and expression of some pathway genes during different stages of morphogenesis in *Withania somnifera* Dunal. *Protoplasma* 250:539–549
4. Mishra SK, Sangwan NS, Sangwan RS (2013) Purification and physicochemical characterization of a gluconolactone inhibition insensitive B glucosidase from *Andrographis paniculata* leaf. *Prep Biochem Biotechnol* 43:481–499
5. Sangwan RS, Tripathi S, Singh J, Narnoliya LK, Sangwan NS (2013) De novo sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene* 525:58–76
6. Gupta P, Agarwal AV, Akhtar N, Sangwan RS, Singh SP, Trivedi PK (2013) Cloning and characterization of 2-C-methyl-d-erythritol-4-phosphate pathway genes for isoprenoid biosynthesis from Indian ginseng, *Withania somnifera*. *Protoplasma* 250: 285–295
7. Liao ZH, Tan QM, Chai YR, Zuo KJ, Chen M, Gong YF, Wang P, Pi Y, Tang F, Sun XF, Tang KX (2004) Cloning and characterization of the gene encoding HMG-CoA reductase from *Taxus media* and its functional identification in yeast. *Funct Plant Biol* 31:73–81
8. Akhtar N, Gupta P, Sangwan NS, Sangwan RS, Trivedi PK (2012) Cloning and functional characterization of 3-hydroxy-3-

- methylglutaryl coenzyme A reductase gene from *Withania somnifera*: an important medicinal plant. *Protoplasma* 250(2): 613–622
9. Shen GA, Pang YZ, Wu WS, Liao ZH, Zhao LX, Sun XF, Tang KX (2006) Cloning and characterization of a root-specific expressing gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Ginkgo biloba*. *Mol Biol Rep* 33:117–127
  10. Jiang JH, Kai GY, Cao XY, Chen FM, He DN, Liu Q (2006) Molecular cloning of a HMG-CoA reductase gene from *Eucommia ulmoides* Oliver. *Biosci Rep* 26:171–181
  11. Kumar ARV, Jayadevi HC, Ashoka HJ, Chandrashekara K (2003) Azadirachtin use efficiency in commercial neem formulations. *Curr Sci* 84:1459–1464
  12. Bhattacharyya N, Chutia M, Sarma S (2007) Neem (*Azadirachta indica* A. Juss), a potent biopesticide and medicinal plant: a review. *J Plant Sci* 2:251–259
  13. Biswas K, Chattopadhyay I, Banerjee RK, Bandyopadhyay U (2002) Biological activities and medicinal properties of neem (*Azadirachta indica*). *Curr Sci* 82:1336–1345
  14. Paul R, Prasad M, Sah NK (2011) Anticancer biology of *Azadirachta indica* L (neem). *Cancer Biol Ther* 12:467–476
  15. Bandyopadhyay U, Biswas K, Sengupta A, Moitra P, Dutta P, Sarkar D, Debnath P, Ganguly CK, Banerjee RK (2004) Clinical studies on the effect of neem (*Azadirachta indica*) bark extract on gastric secretion and gastroduodenal ulcer. *Life Sci* 75(24): 2867–2878
  16. Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res* 7:175–180
  17. Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. *Plant Mol Biol* 35:135–144
  18. Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc Natl Acad Sci* 95:13330–13335
  19. Ryang SH, Chung SY, Lee SH, Cha JS, Kim HY, Cho TJ (2002) Isolation of pathogen-induced Chinese cabbage genes by subtractive hybridization employing selective adaptor ligation. *Biochem Biophys Res Commun* 299:352–359
  20. Naika D, Dhanaraja AL, Arorab R, Rowlanda LJ (2007) Identification of genes associated with cold acclimation in blueberry (*Vaccinium corymbosum* L.) using a subtractive hybridization approach. *Plant Sci* 173:213–222
  21. Diatchenko L, Laut YFC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskayat N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci* 93:6025–6030
  22. Krishnan NM, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, Deepak S, Hariharan AK, Krishna PB, Nair J, Varghese L, Valivarathi NK, Dhas K, Ramaswamy K, Panda B (2012) A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genom* 13:464
  23. Mizutani M, Ohta D (2010) Diversification of P<sub>450</sub> genes during land plant evolution. *Annu Rev Plant Biol* 61:291–315
  24. Nelson DR, Ming R, Alam M, Schuler MA (2008) Comparison of cytochrome P<sub>450</sub> genes from six plant genomes. *Trop Plant Biol* 1:216–235
  25. Rajakani R, Narnoliya L, Sangwan NS, Sangwan RS, Gupta V (2013) Activated charcoal-mediated RNA extraction method for *Azadirachta indica* and plants highly rich in polyphenolics, polysaccharides and other complex secondary compounds. *BMC Res Notes* 6:1–10
  26. Conesa A, Gotz S, Garcia GZ, Teroz JM, Talón J, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
  27. Conesa A, Gotz S (2007) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. doi:10.1155/2008/619832
  28. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
  29. Kato ES, Higashi K, Hosoya K, Kobayashi T, Ezura H (2001) Cloning and characterization of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase in melon (*Cucumis melo* L. reticulatus). *Mol Genet Genomics* 265:135–142
  30. Krishnan NM, Swetansu P, Deepak SA, Hariharan AK, Gaur P, Choudhary R, Jain P, Vaidyanathan S, Krishna PGB, Panda B (2011) De novo sequencing and assembly of *Azadirachta indica* fruit transcriptome. *Curr Sci* 101:12
  31. Saxena A, Singh P, Yadav DK, Sharma P, Alam S, Khan F, Thul ST, Shukla RK, Gupta G, Sangwan NS (2013) Identification of cytochrome P<sub>450</sub> heme motif in plants proteome. *Plantomics* 6(1):1–12