International Journal of Biochemistry Research & Review
4(1): 1-15, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

Establishment of cDNA Library and EST Analysis from Leaves of *Phyllanthus amarus*

Sharmila Chattopadhyay[1*] and Aparupa Bose Mazumdar Ghosh[1]

[1]Plant Biology Laboratory, Drug Development/Diagnostics and Biotechnology Division; CSIR Indian Institute of Chemical Biology, 4, Raja, S. C. Mullick Road, Kolkata- 700 032, India.

Authors' contributions

This work was carried out in collaboration between all authors. Author SC designed the study and finally prepared the corrected version of the manuscript. Author ABMG performed the experiment and statistical analysis, literature searches, wrote the protocol, and the first draft of the manuscript. Both authors read and approved the final manuscript.

Research Article

Received 12th June 2013
Accepted 1st September 2013
Published 9th October 2013

ABSTRACT

DNA sequencing of randomly chosen clones from a cDNA library allows thousands of different transcripts to be identified. However, since the likelihood of observing a given transcript is proportional to the expression level of that transcript in the tissue from which the library is derived, often transcripts are represented by several EST sequences. An expressed sequence tags (EST) analysis was undertaken to identify the genes present in the leaves of *Phyllanthus amarus,* which is a small tropical, glabrous herb with several health benefits. Phyllanthin and hypophyllanthin, major bioactive components, present in highest amounts in the leaves, are of significant therapeutic importance like hepatoprotective, antioxidant, antiviral, hypoglycemic, etc. Taken together, sequencing of cDNA clones generated high-quality ESTs (Accession number: JK492908 to JK492964) with high similarities with genes from *Ricinus communis, Onchocerca volvulus*, *Eucalyptus globules, Gossypium hirsutum, Nicotiana tabacum, Solanum* spp. and many more. A BLASTN analysis along with BLASTX analysis of all the unique sequences was performed and was grouped according to the reported activities. Results represented here is the first reference collection of ESTs from this commercially important medicinal herb. This study indicated that the leaf transcriptome contains series of interesting sequences like ALBINO3, ribulose-1, 5 bisphosphate carboxylase/ oxygenase (RUBISCO), chloroplast

**Corresponding author: Email: sharmila@iicb.res.in;*

photosystem II chlorophyll A/B-binding protein, stress-responsive proteins like methionine sulfoxide reductase type, etc.

Keywords: cDNA library; expressed sequence tags (EST); Phyllanthus amarus.

ABBREVIATIONS

EST: Expressed Sequence Tag; NCBI: National Centre for Biotechnology Information BLAST: Basic Local Alignment Search Tool.

1. INTRODUCTION

The plant genus *Phyllanthus* (Euphorbiaceae) is widely distributed in most tropical and subtropical countries. It is a very large genus consisting of approximately 550 to 750 species and is subdivided into 10 or 11 subgenera. [1,2]. *Phyllanthus amarus* Schum & Thonn. popularly known as bhuiamlaki, is found in different parts of India.

Genus: *Phyllanthus*
Species: *amarus*

This important species of the genus *Phyllanthus* has been assigned common names in different languages in India as bhuiamla and sadahazurmani in Bengali, bhuiavala in Bombay, bhonyaanmali in Gujrati, bhuianvalah in Hindi, etc. [3]. In Spain this plant is best known by the common name chanca piedra, which means stonebreaker [4].

P. amarus has a long history of use as an herbal edible source for the treatment of a broad spectrum of diseases in Brazil, India and other countries [1,3]. Works done by various group of researchers globally have shown the huge potential of this herb in treating multi-faceted diseases like jaundice, hepatitis B, HIV, Herpex Simplex virus, cancer, diabetes, inflammation, oxidative stress, etc. [5-9]. The whole plant is used in gonorrhea, menorrhagia and other genital affections. It is useful in gastropathy, diarrhoea, dysentery, intermittent fevers, ophthalmopathy, scabies, ulcers and wounds. It also has antimutagenic activities [10], antinociceptive activities [11] and antilipidemic potentials [12,13]. The organic compounds of medicinal interest present in this herb include lignans – phyllanthin and hypophyllanthin [14,15], geraniin and 5 flavanoids -quercetin, astralgin, quercetrin, isoquercetin and rutin [16,4]. It also contains minor compounds like hydrolysable tannins like phyllanthusiin D [17], amariin [16], amarulone [18], amarinic acid and alkaloids like ent-norsecurinine, sobubbialine, epibubbialine [17]; diarylbutane, nyrphyllin, and a neolignan, phyllnirurin besides lactones, steroids, terpenoids, and so forth. The plant is bitter, astringent, cooling, diuretic, febrifuge and antiseptic.

Over the past years, construction of cDNA library and its analysis is considered to be an indispensable tool for functional genomic analysis as it provides much more detailed information on the genomic mechanisms underlying diverse processes of the organism [19]. Thus, cDNA library is a powerful and useful tool in the area of biotechnology. It is helpful in expressing eukaryotic genes in prokaryotes, which in turn helps in the transcription process of prokaryotes; facilitate the study of the repertoire of mRNAs expressed in different cells or tissues, and study of alternative splicing in different cells or tissues; helps in discovery of novel genes and cloning of full-length cDNA molecules for *in vitro* study of gene function.

Gubler and Hoffman developed the original method for generating cDNA libraries by combining classical first-strand synthesis with RNase H DNA polymerase I mediated second-strand synthesis. This allows the conversion of first-strand to second-strand ds cDNA by RNA-primed nick-translation without significant loss of sequence information [20]. Various modifications followed over the years to improve the method's versatility [21], but the procedure to synthesise cDNA has not been extensively changed, with the possible exception of the construction of uniform-abundance (normalised) cDNA libraries [22].

ESTs are fragments of mRNA sequences generated from randomly selected cDNA clones. Since they represent the expressed portion of a genome, ESTs have proven to be extremely useful for purposes of gene identification and verification of gene predictions. Over 45 million ESTs have been generated from over 1400 different species of eukaryotes [23]. Expressed sequence tag (EST) and complementary DNA (cDNA) sequences are currently the most important resources for transcriptome exploration. ESTs are short (usually 200–800 nucleotide bases in length); unedited, randomly selected single-pass sequence reads derived from cDNA libraries. High-throughput ESTs can be generated at a reasonably low cost from either the 5' or 3' end of a cDNA clone to get an insight into transcriptionally active regions in any organism. In 1991, ESTs were used as a primary resource for human gene discovery. Thereafter, there has been an exponential growth in the generation and accumulation of EST data in public databases for myriad organisms. At present, ESTs enable gene discovery, complement genome annotation, aid gene structure identification, establish the viability of alternative transcripts, guide single nucleotide polymorphism (SNP) characterization and facilitate proteome analysis [24,25]. Resources pertaining to EST analysis are maintained at http://biolinfo.org/EST/. Related to medicinal plants, EST analysis and cDNA libraries have been constructed from various plant species. Examples include *Panax ginseng* [26,27] and *Panax quinquefolius* [28]. Korean ginseng (*Panax ginseng* C.A. Meyer) and American ginseng (*Panax quinquefolius* L.) belong to the Araliaceae family and are herbaceous medicinal plants possessing pharmacologically active constituents known as ginsenosides[29,30]. High quality Expressed Sequence Tags (ESTs) from a saffron stigma cDNA library have been reported which represents the first reference collection for the genomics of Iridaceae family, for the molecular biology of stigma biogenesis, as well as for the metabolic pathways underlying saffron secondary metabolism [31]. Saffron (*Crocus sativus* Linn*)* is a flowering plant and the stigmas of the plant are mainly used in traditional medicine for various purposes, as an aphrodisiac, antispasmodic, expectorant, for treatment of stomach ailments, reducing stomachache and for relieving tension. Other examples include a large EST dataset generation for a traditional Chinese medicinal plant species *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim, aiming to provide sequence information for deciphering secondary metabolism, especially for flavonoid pathway in *Epimedium*. Out of the total 76,459 consensus sequences in the EST dataset a total of 2,810 EST-SSRs have been identified which together will be a powerful resource for further studies such as taxonomy, molecular breeding, genetics, genomics, and secondary metabolism in *Epimedium* species [32]. Reports of cDNA library construction and analysis of ESTs from roots of *Bupleurum chinense* DC.are also known. *Radix Bupleuri* (Chaihu), sourced from the dried roots of *Bupleurum* species, is a traditional Chinese medicine with anti-inflammatory, anti-pyretic, and anti-hepatotoxic efficacy. The library and EST data of this medicinal plant provide a platform to study the molecular mechanisms of various physiological phenomena of *Bupleurum* [33]. cDNA library has also been constructed for white flower *Salvia miltiorrhiza bge. f.alba.* along with partial EST sequence analysis in order to screen and isolate secondary metabolite biosynthesis related [34]. *Salvia miltiorrhiza Bge* is a well-known traditional Chinese herb and broadly planted in China and has been used clinically for

the treatment of various ailments such as cardiovascular, cerebrovascular, hyperlipidemia, and acute ischemic stroke diseases [35,36].

The main focus of this research article is the implementation of molecular biology methods by employing the techniques of cDNA library construction and expressed sequence tags (ESTs) analysis. Thus in this study we have created a cDNA library from fresh young leaves of *P. amarus* and used it for generation of EST database to explore the presence of metabolite synthesizing enzymes in the library. This is the first report of such an effort to establish a cDNA library from the leaves of *Phyllanthus amarus.*

2. MATERIALS AND METHODS

2.1 Plant Material

P. amarus cultivated in the medicinal plant garden of Indian Institute of Chemical Biology, Kolkata was collected and the fresh young leaves of the plant were used as the experimental material for mRNA isolation.

2.2 RNA Isolation and mRNA Purification

The fresh, young leaves of the plant were used for total RNA isolation with TRIzol agent (Invitrogen Company) following the manufacturer's instructions. The integrity and size of the isolated RNA were determined by formaldehyde agarose gel electrophoresis, and the quantity and purity of the RNA were assessed by nanodrop 1000 (Thermo Scientific).

2.3 Construction of cDNA Library

1 µg of DNase free total RNA, isolated from young leaves, was used for the synthesis of first strand cDNA according to the protocol of SMARTer™ cDNA Library Construction Kit (Clontech Lab Inc, USA). The first step involved mixing of 1µl of RNA with 1µl of 12µM 3' SMART CDS Primer II A in addition to 2.5 µl of deionized water in a total of 4.5 µl of reaction mixture and incubated at 72ºC for 3 minutes with a further reduction of temperature to 42ºC for 2 minutes. Meanwhile during the incubation period a master mix was prepared at room temperature to a total volume of 5.5µl using 2µl of 5X first strand buffer, 0.25µl of 100mM DTT, 1µl of 10mM dNTP mix, 1µl of 12µM SMARTer II A Oligonucleotide, 0.25µl of RNase inhibitor and 1µl of SMARTScribe™ Reverse Transcriptase. This total volume of the master mix was added to the reaction mixture of the previous step and the entire mixture was incubated at 42ºC for 90 minutes. Amplification of cDNA and preparation of double stranded cDNA (ds cDNA) was performed by long distance PCR (LD PCR) using 2µl of the total first strand cDNA synthesized storing the remaining first strand cDNA at -20ºC. The entire process of LD PCR was done in duplicates i.e. two separate 100µl PCR reactions were set up and each combining 2µl of first strand cDNA, 10µl of 10X Advantage 2 PCR Buffer, 2µl of 10mM 50X dNTP Mix, 2µl of 12µM 5' PCR Primer II A, 2µl of 50X Advantage 2 Polymerase Mix along with 82µl of deionized water. The reactions were placed in a thermal cycler preheated to 95ºC and each reaction was subjected to 15 cycles after which the PCR program was paused and 30µl from each tube was transferred to a second reaction tube named the "optimization tube" which was further used for the remaining additional PCR reactions. The other reaction tube containing 70µl of the 15^{th} cycle PCR reaction mix named the "experimental tube" was stored at 4ºC. 5µl from the 15 cycle PCR reaction (optimization tube) was transferred to another tube for agarose/EtBr gel analysis. To prevent distortion of

representation of cDNA molecules within the total cDNA population and to maintain the double-stranded cDNA in the exponential phase of amplification, the optimal number of PCR cycles was determined by sampling the remaining 25µl of PCR mixture of the optimization tubes by running three additional cycles (for a total of 18) and increments of three cycles thereafter, up to 27 cycles. 5µl from each of the PCR mixtures of 21st, 24th, and 27th cycle was aliquoted and kept separately and was electrophoresed along with the 18th cycle PCR aliquot and 1kb DNA marker on a 1.1% agarose/EtBr gel in 1X TAE buffer to visualize the amount and the size distribution of the cDNA (Fig. 1). Following the LD PCR, the ds cDNA was purified and size-fractioned using the CHROMA SPIN™ +TE-1000 Column. According to the protocol of the CHROMA SPIN +TE-1000 column was prepared and centrifuged at 700g for 5 minutes to purge the equilibration buffer and semi-dry the gel matrix. In the next step the cDNA sample was added to the center of the flat surface of the gel matrix carefully to prevent the flow of the sample into the inner wall of the column and the centrifuged at 700g for 5 minutes. The purified sample of cDNA obtained was pooled into a clean microcentrifuge tube and ethanol precipitated. According to the protocol 1.5µl of Glycogen (20µg/µl), 10µl of 3M Sodium Acetate (pH 4.8), 250µl of ice-cold ethanol (100%) were added one by one and the tube was placed in –20ºC freezer for 1 hr to precipitate the cDNA. After centrifugation at 14,000 rpm for 20 minutes and careful removal of the supernatant, the pellet was washed using 100µl 70% ethanol. The tube was again centrifuged at 14,000 rpm for 5 minutes at room temperature and carefully removing the remaining supernatant the pellet was air dried for 10 minutes. The cDNA was then resuspended in 15µl of TE buffer. 2-5µg of ds cDNA obtained was ligated into pGEM-T vector (Promega) using the TA cloning procedure and for the availability of the M13 forward and the reverse primers. Next after 18 hours of ligation at 4ºC the ligation mixture was transferred into the competent cells of *Escherichia coli* strain JM109 by the $CaC1_2$ transformation method. To screen white plaques with insert fragment, the transformed cells were plated on 140 mm Petri dishes containing LB/ampicillin/ IPTG/X-Gal and were placed overnight at 37ºC. The colonies were counted and the the ratio of white plaques (recombinants) to blue plaques (non-recombinants) was rapidly estimated as the recombination efficiency.

2.4 Sequencing and EST Analysis

The white bacterial colonies obtained from the library were randomly picked and cultured on selective LB media and grown overnight at 37^0C shaker. The plasmid DNA was isolated from the randomly picked colonies using QIAprep Plasmid DNA isolation kit. The presence of inserts was confirmed by restriction enzyme analysis done with EcoRI in the presence of suitable buffers at 37ºC for 1½ hours. The restriction enzyme products were analyzed on a 1.1% agarose/EtBr gel. After clones' analysis, they were sent to TCGA (The Center for Genomic Application, New Delhi, India) for sequencing. Single pass sequenced clones were used for downstream applications.

2.5 Editing of ESTs and BLASTN Analysis

The sequenced clones from the library were edited to remove vector sequences using Vecscreen software of NCBI (http://www.ncbi.nlm.nih.gov/Vecscreen/Vecscreen.html). The resulting sequences were then annotated by performing sequence similarity searches against the NCBI non-redundant database using the BLASTN analysis. The functional classification was based on BLASTX analysis. The sequencing resulted in 3' end high-quality expressed sequenced tags which were reported to NCBI and submitted to GenBank under the accession numbers JK492908 to JK492964.

3. RESULTS AND DISCUSSION

Total RNA analysis of *Phyllanthus amarus* on 1.5% agarose-formaldehyde gel showed the presence of two distinct bands of 28S and 18S rRNA bands in case of both the libraries. The ratio of intensities of the 28S and 18S rRNA bands was 1.5: 1 (Fig. 1). 18 LD-PCR cycles yielded low smear ranging in size from 0.1Kb to 4 kb. 3 additional cycles further up to 27th cycle yielded a prominent smear from 0.1 Kb to 4 Kb. Due to the high complexity of RNA of *P. amarus*, several distinct bands were absent in the smear (Fig. 2).

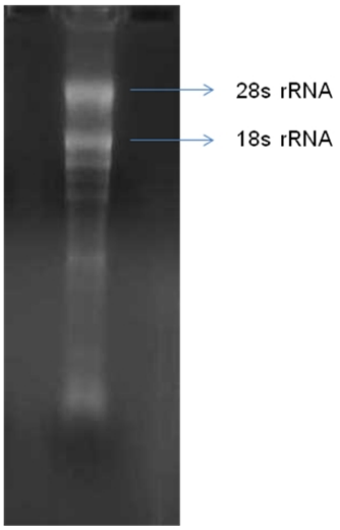

Fig. 1. 1.5% agarose-formaldehyde gel showing RNA isolated from fresh young leaves of *P. amarus*

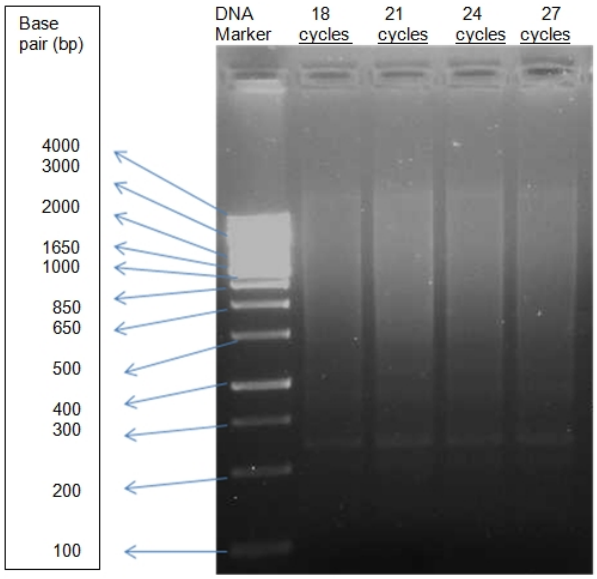

Fig. 2. LD PCR cycles yielding smear ranging in size from 0.1 to 4 Kb

Plasmid DNA isolated from randomly isolated clones of the cDNA library was digested with the restriction enzyme EcoRI. Agarose gel analysis showed the presence of insert ranging in size from 300 bp to 2 Kb and 200 to 800 bp respectively.

3.1 BLASTN Analysis and Functional Classification of cDNA Library

The sequenced clones after BLASTN analysis showed a total of 57 sequences as shown in Table 1. Functional categorization of these 57 clones was performed with BLASTX analysis. They have been divided into nine classes as shown in Fig. 3 - Ribosomal proteins, Chloroplast proteins, Chlorophyll-binding proteins, Stress-related proteins, Protein metabolism, Structural proteins, Proteins of unknown functions, Amino acid-binding proteins, and other types of proteins.

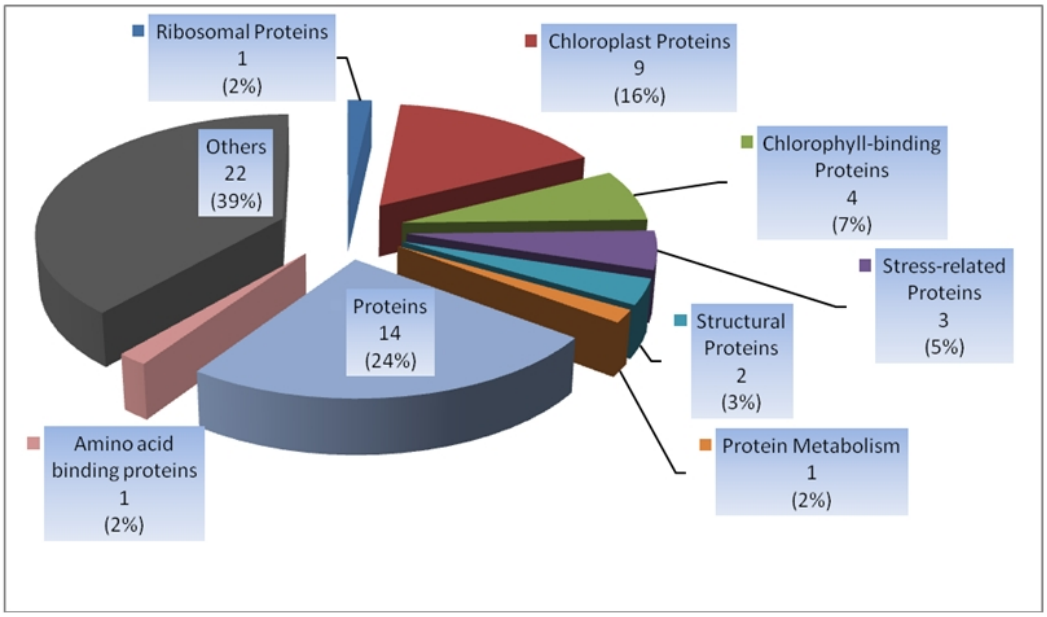

Fig. 3. Functional categorization of 57 sequences into nine classes according to BLASTX analysis

Ribosomal proteins/protein metabolism included S28 (JK492943) and rubber elongation factor protein (JK492942) belonging to protein metabolism type. The ribosomal proteins and protein metabolism types present in the library indicate the synthesis of proteins in leaves.

Chloroplast proteins mainly included ribulose-1, 5-bisphosphate carboxylase/oxygenase (JK492909, JK492964, JK492923, JK492926, JK492944, JK492931) which is one of the important chloroplastic proteins related to photosynthesis. Besides chloroplast membrane protein ALBINO 3 (JK492915) and chloroplast ferredoxin I (fdn-1) (JK492920) mRNA were also obtained.

Chlorophyll-binding proteins included chlorophyll A/B binding (JK492933, JK492936, JK492946) and LHCII type II chlorophyll a/b-binding (CipLhcb2) (JK492935) protein types. Chlorophyll-binding proteins constitute a large family of proteins and have diverse functions in both light-harvesting and photoprotection [37].

Table 1. Some of the ESTs from cDNA library characterized from leaves of *P. amarus*

Sl. No.	Clone no.	BLAST hit	Accession no. of reported sequences	Accession no. of EST clones	Insert length (bp)	%Identity of ESTs with the reported sequences	% of the ESTs having putative matches with the known cDNAs.	e-value
Ribosomal proteins								
1.	PA lib 96	*Ricinus communis* ribosomal protein S28, putative, mRNA	XM002531546	JK492943	700	83%	34%	2e-53
Chloroplast Proteins								
1.	PA lib 7	*Eucalyptus globulus* EgRBCS1 mRNA for ribulose-1,5bisphosphate carboxylase/ oxygenase small subunit, complete cds	AB537497	JK492909	1132	79%	46%	3e-111
2.	PA lib21	*Arabidopsis thaliana* chloroplast membrane protein (ALBINO3) (At2g28800) mRNA, complete cds	BT002436	JK492915	1174	75%	26%	2e-76
3.	PA lib 26	*R.communis* Ribulose bisphosphate carboxylase small chain, chloroplast precursor, putative, mRNA	XM002531578	JK492964	743	78%	66%	2e-84
4.	PA lib 31	*Nicotiana tabacum* chloroplast ferredoxin I (fdn-1) mRNA, complete cds; nuclear gene for chloroplast product	AY864890	JK492920	916	72%	20%	1e-31
5.	PA lib 42	*Corchorus olitorius* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 1A mRNA, complete cds	EU024512	JK492923	938	79%	47%	8e-130
6.	PA lib 49	*Citrus reticulata* ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit precursor, gene, partial cds, nuclear gene for chloroplast product	AF312228	JK492926	668	80%	53%	1e-87
7.	PA lib 99	*C. olitorius* ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit 1A mRNA, complete cds	EU024512	JK492944	547	82%	50%	9e-101
8.	PA lib 100	*Linum usitatissimum* cultivar Nike ribulose bisphosphate carboxylase activase mRNA, partial cds	GU581046	JK492945	587	82%	48%	4e-11
9.	PA lib 57	*E. globulus* EgRBCS1 mRNA for ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit, complete cds	AB537497	JK492931	1134	76%	63%	4e-84
Chlorophyll-binding proteins								
1.	PA lib 65	*Jatropha curcas* chlorophyll A/B binding protein	GQ984214	JK492933	1181	82%	20%	0.0

		mRNA, complete cds						
2.	PA lib 68	*Vigna radiata* LHCII type II chlorophyll a/b-binding protein (CipLhcb2) mRNA, complete cds; nuclear gene for chloroplast product	AF279248	JK492935	471	80%	43%	3e-56
3.	PA lib 74	*R.communis* chlorophyll A/B binding protein, putative, mRNA	XM002524570	JK492936	820	85%	61%	0.0
4.	PA lib 101	*Gossypium hirsutum* chloroplast photosystem II chlorophyll A/B-binding protein gene, complete cds	L07119	JK492946	1130	84%	64%	0.0
Stress-related proteins								
1.	PA lib 94	*Populus trichocarpa* methionine sulfoxide reductase type (PtrMsrB2), mRNA	XM002316839	JK492941	676	80%	47%	2e-116
2.	PA lib 35	*Populus* EST from severe drought-stressed leaves	CU226865	JK492948	1109	72%	4%	9e-48
3.	PA lib 39	*Populus* EST from mild drought-stressed leaves	CU232168	JK492949	524	85%	42%	2e-59
Structural Proteins								
1.	PA lib83	*R. communis* leucine-rich repeat-containing protein, putative, mRNA	XM002522286	JK492938	821	75%	1%	5e-68
2.	PA lib 93	*G. hirsutum* hybrid proline-rich protein (HyPRP4) mRNA, complete cds	HM989876	JK492940	931	74%	15%	8e-35
Protein Metabolism								
1.	PA lib 95	*R. communis* Rubber elongation factor protein, putative, mRNA	XM002512381	JK492942	1087	77%	11%	2e-63
Proteins								
1.	PA lib 32	*P. trichocarpa* predicted protein, mRNA	XM002303120	JK492947	555	78%	-	2e-72
2.	PA lib 34	*R. communis* Xyloglucan endotransglucosylase/ hydrolase protein 22 precursor, putative, mRNA	XM002526182	JK492921	1124	83%	56%	0.0
3.	PA lib 47	*Zea mays* uncharacterized protein LOC100382066 (LOC100382066), mRNA	NM001174829	JK492954	1174	100%	-	5.7
4.	PA lib 41	*R. communis* conserved hypothetical protein, mRNA	XM002528323	JK492951	806	83%	-	9e-103
5.	PA lib 43	*A. lyrata* subsp. lyrata predicted protein, mRNA	XM002873053	JK492952	1112	77%	-	3e-105
6.	PA lib 50	*R. communis* Xyloglucan endotransglucosylase/hydrolase protein 22 precursor, putative, mRNA	XM002526182	JK492927	1164	83%	56%	0.0
7.	PA lib 51	*H.annuus* UbB1 mRNA for hexaubiquitin protein	X57004	JK492928	584	68%	57%	3e-12
8.	PA lib 52	*Pisum sativum* ubiquitin-like protein mRNA, partial cds	EF212231	JK492929	704	78%	29%	5e-55

9.	PA lib 59	*A. lyrata* subsp. *lyrata* hypothetical protein, mRNA	XM002876737	JK492932	1171	67%	-	2e-55
10.	PA lib 63	*R. communis* conserved hypothetical protein, mRNA	XM002519900	JK492956	721	81%	-	8e-78
11.	PA lib 76	*P. trichocarpa* predicted protein, mRNA	XM002329098	JK492960	673	85%	-	5e-61
12.	PA lib 81	*Glycine max* uncharacterized protein LOC100500600 (LOC100500600), mRNA	NM001249691	JK492937	1005	78%	-	7e-112
13.	PA lib 82	PREDICTED: *Vitis vinifera* hypothetical protein LOC100249177 (LOC100249177), mRNA	XM002280936	JK492961	1200	71%	-	0.0
14.	PA lib 102	*Hevea brasiliensis* latex plastidic aldolase-like protein mRNA, complete cds	AF467803	JK492962	1132	82%	8%	0.0
Amino acid binding proteins								
1.	PA lib 28	*R. communis* amino acid binding protein, putative, mRNA	XM002522978	JK492917	1141	81%	2%	0.0
Others								
1.	PA lib 30	*J. curcas* nucleoside diphosphate kinase B (NDKB) mRNA, complete cds	FJ906846	JK492919	779	87%	26%	1e-151
2.	PA lib 38	*P. trichocarpa* galactinol synthase 1 (GOLS1) mRNA, complete cds	EU305718	JK492922	1187	80%	28%	0.0
3.	PA lib 40	*Schistosoma japonicum* isolate Anhui clone BAC C108_76D19, complete sequence	FN293032	JK492950	388	100%	-	1.7
4.	PA lib 44	*Onchocerca volvulus* glutathione S-transferase Ia gene, complete cds	AF265556	JK492924	470	81%	-	7.5
5.	PA lib 46	*V. vinifera* contig VV78X157397.15, whole genome shotgun sequence	AM432029	JK492925	922	100%	-	6.1
6.	PA lib 45	*Medicago sativa* class IV chitinase mRNA, complete cds	FJ487629	JK492953	393	95%	-	1e-07
7.	PA lib 67	*Triticum aestivum* O-methyltransferase 1 gene, 5' UTR82%	EU831287	JK492934	1031	82%	1%	0.40
8.	PA lib 75	*Lotus japonicus* genomic DNA, chromosome 3, clone: LjT34E16, TM1822, complete sequence	AP009717	JK492959	1057	88%	-	1.4
9.	PA lib 87	*Xenopus laevis* chloride intracellular channel 6 (clic6),	NM001086864	JK492939	348	100%	1%	5.4

		nuclear gene encoding mitochondrial protein, mRNA						
10.	PA lib 103	*Solanum lycopersicum* cDNA, clone: LEFL1030AH04, HTC in leaf	AK247080	JK492963	900	73%	-	3e-28
11.	PA lib 4	*P. trichocarpa* lysine/histidine transporter, mRNA	XM002298561	JK492908	907	78%	52%	2e-142
12.	PA lib 9	*R. communis* glyceraldehyde 3-phosphate dehydrogenase, putative, mRNA	XM002512744	JK492910	1113	81%	29%	0.0
13.	PA lib 11	*Drosophila erecta* GG12460 (Dere\GG12460), mRNA	XM001982143	JK492911	125	84%	-	1.6
14.	PA lib 12	*P. trichocarpa* clone ACSB2121-B05, complete sequence	AC210380	JK492912	1170	74%	-	3e-155
15.	PA lib 15	*P. trichocarpa* clone WS0127_L10 un.own mRNA	EF148116	JK492914	427	85%	-	1e-59
16..	PA lib 25	TSA: *J. curcas* Contig11805, mRNA sequence	EZ419086	JK492916	549	73%	-	9e-25
17.	PA lib 29	*Cryptococcus neoformans* var. *neoformans* JEC21 chromosome 7, complete sequence	AE017347	JK492918	97	100%	-	1.1
18.	PA lib 54	*R. communis* ATP-dependent Clp protease proteolytic subunit, putative, mRNA	XM002515227	JK492930	1175	82%	21%	0.0
19.	PA lib 13	Soybean clone JCVI-FLGm-8L10 unknown mRNA	BT091818	JK492913	1142	72%	-	3e-67
20.	PA lib 48	*Populus* EST from leave	CU226205	JK492955	667	85%	-	2e-59
21.	PA lib 64	*M. sativa* class IV chitinase mRNA, complete cds	FJ487629	JK492957	913	89%	-	0.008
22.	PA lib 66	*P. tomentosa* calmodulin mRNA, complete cds	AY836672	JK492958	657	88%	41%	1e-169

Stress-related proteins included methionine sulfoxide reductase type (PtrMsrB2) (JK492941), mRNA and ESTs derived from severe to mild drought stressed leaves of Populus (JK492948, JK492949). Methionine sulfoxide reductase (MSR) proteins fulfill essential functions in stress tolerance and during ageing in bacterial, yeast, and mammal cells. MSRs are key components in the control of oxidative damage associated with the development of disorders and the process of ageing [38].

Structural proteins included leucine-rich repeat-containing (JK492938) and proline-rich proteins (JK492940). Leucine-rich repeat receptor kinases (LRR-RKs) comprise the largest subfamily of transmembrane receptor-like kinases in plants, with over 200 members in Arabidopsis. LRR-RKs regulate a wide variety of developmental and defense-related processes including cell proliferation, stem cell maintenance, hormone perception, host-specific as well as non-host-specific defense response, wounding response, and symbiosis [39]. Proline-rich protein's (PRP) assembly in the plasma membrane and cell wall may increase tolerance to certain stresses [40].

The other two categories viz. 'proteins' and 'others' included the various types of proteins as mentioned in the table out of which some were the xyloglucan endotrans glucosylase /hydrolase protein (JK492921), hexaubiquitin (JK492928) and ubiquitin-like proteins (JK492929), latex plastidic aldolase-like protein. Xyloglucan endotrans glucosylase/hydrolases (XTHs) are enzymes involved in the modification of load-bearing cell wall components. They cleave xyloglucan chains and, often, re-form bonds to the non-reducing ends of available xyloglucan molecules in plant primary cell walls [41]. Protein metabolism type included rubber elongation factor protein (JK492917) etc.

Out of the total 57 ESTs, 34 ESTs i.e. almost 60% of the ESTs are with known putative functions. The percentages of the ESTs of each protein group having putative matches with the known cDNAs have been mentioned in the Table 1.

4. CONCLUSION

The EST's isolated from the cDNA library of young leaves of *P. amarus* can be used as a first-hand reference at the molecular level to identify the genes of phyllanthin and hypophyllanthin biosynthetic pathway. Further, the cDNA library that we have constructed will be useful for downstream studies of this therapeutically important plant.

ACKNOWLEDGEMENTS

This work received financial support from Department of Science & Technology (DST) and Council of Scientific and Industrial Research, Government of India. ABMG is thankful to DST, New Delhi for Junior Research Fellowship (DST Inspire fellowship).

COMPETING INTERESTS

The authors declare that they have no conflict of interests. We received all the chemicals, kits, etc from the store section of CSIR-IICB. As the authors of the manuscript we do not have any direct or indirect financial relation with the commercial identities.

REFERENCES

1. Calixto JB, Santos AR, Cechinel Filho V, Yunes RA. A review of the plants of the genus *Phyllanthus*: Their chemistry, pharmacology, and therapeutic potential. Med Res Rev. 1998;18(4):225-258.
2. Unander DW, Webster GL, Blumberg BS. Usage and bioassays in *Phyllanthus* (Euphorbiaceae). IV. Clustering of antiviral uses and other effects. J Ethnopharmacol. 1995;45:1-18.
3. The Wealth of India; A dictionary of Indian raw materials and industrial product. CSIR, New Delhi, India. 2003;34–36.
4. Morton JF. Atlas of Medicinal Plants of Middle America. Library of Congress cataloging in Publication Data. Thomas books. 1981;1420.
5. Harish R, Shivanandappa T. Antioxidant activity and hepatoprotective potential of *Phyllanthus niruri.* Food Chem. 2006;95(2):180–185.
6. Notka F, Meier G, Wagner R. Concerted inhibitory activities of *Phyllanthus amarus* on HIV replication in vitro and ex vivo. Antiviral Research. 2004;64(2):93–102.
7. Thyagarajan SP, Jayaram S, Gopalakrishnan V, Hari R, Jeyakumar P, Sripathi MS. Herbal medicines for liver diseases in India. J Gastroen Hepatol. 2002;17:370-376.
8. Thyagarajan SP, Subramanian S, Thirunalasundari T, Venkateswaran PS, Blumberg BS. Effect of *Phyllanthus amarus* on chronic carriers of hepatitis B virus. The Lancet. 332. 1988;(8614):764-766.
9. Chevallier, A. Encyclopedia of Herbal Medicine: Natural Health. Dorling Kindersley Book. USA. 2000; Second edition: 336.
10. Joy KL, Kuttan R. Inhibition by *Phyllanthus amarus* of hepatocarcinogenesis induced by N- nitrosodiethylamine. J Clin Biochem Nutr. 1998;24:133-139.
11. Kassuya CA, Silerstre AA, Rehder V, Calixto JB. Anti-allodynic and anti-oedematogenic properties of the extract and lignans from *Phyllanthus amarus* in models of persistent inflammatory and neuropathic pain. Eur J Pharmacol. 2003;478:145-153.
12. Adeneye AA, Amole OO, Adeneye AK. The hypoglycemic and hypocholesterolemic activities of the aqueous leaf and seed extracts of *Phyllanthus amarus* in mice. Fitoterapia. 2006;77:511-514.
13. Adeneye AA, Benebo AS, Agbaje EO. Protective effect of the aqueous leaf and seed extract of *Phyllanthus amarus* on alcohol-induced hepatotoxicity in rats. West African J of Pharmacol and Drug Res. 2006;23:42-50.
14. Sharma A, Singh RT, Handa SS. Estimation of phyllanrthin and hypophyllanthin by high performance liquid chromatography in *Phyllanthus amarus*. Phytochem analysis. 1993;4(5):226-229.
15. Somanabandhu A, Nitayangkura S, Mahidol C, Ruchirawat S, Likhitwitayawuid K. 1H- and 13C- nmr assignments of phyllanthin and hypophyllanthin: Lignans that enhance cytotoxic responses with cultured multidrug resistant cells. J Nat Prod. 1993;56:223 - 239.
16. Foo LY. Amarulone, a novel cyclic hydrolysable tannin from *Phyllanthus amarus*. Nat Prod Lett. 1993;3:45-52.
17. Houghton PJ, Woldemariam TZ, O'Shea S, Thyagarajan SP. Two securinega-type alkaloids from *Phyllanthus amarus*. Phytochemistry. 1996;43(3):715-717.
18. Rao GS, Bramley R. Hypophyllanthin.Tetrahedron Letters. 1971;34:3175-3178.
19. Shao ZT, Cong X, Yuan JD, Yang G-W, Chen Y, Pan J, An L. Construction and characterization of a cDNA library from head kidney of Japanese sea bass (*Lateolabrax japonicus*). Mol Biol Rep. 2009;36(7):2031-7.

20. Gubler U, Hoffman BJ. A simple and very effective method for generating cDNA libraries. Gene. 1983;25:263-269.
21. Rutledge RG, Seligy, VL, Cote, M-J, Dimock K, Lewin LL, Tenniswood MP. Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors. Gene. 1988;68(1):151-158.
22. Patanjali S, Parimoo S, Weissman SM. Construction of a uniform-abundance (normalized) cDNA library. PNAS USA. 1991;88:1943-1947.
23. Parkinson J, Blaxter M. Expressed Sequence Tags: an overview. Methods Mol Biol. 2009;533:1-12.
24. Dong Q, Kroiss L, Oakley FD, Wang BB, Brendel V. Comparative EST analyses in plant systems. Method Enzymol. 2005;395:400–418.
25. Rudd S. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 2003;8(7):321–329.
26. Jung JD, Park H-W, Hahn Y, Hur C-G, In DS, Chung H-J, Liu JR, Choi D-W. Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. Plant Cell Rep. 2003;22(3):224-230.
27. Kim MK, Lee BS, In JG., Sun H, Yoon JH, Yang DC. Comparative analysis of expressed sequence tags (ESTs) of ginseng leaf. Plant Cell Rep. 2006;25(6):599-606.
28. Choi H-Il, Kim N-H, Lee J, Choi BS, Kim KD, Park JY, Lee SC, Yang T-J. Evolutionary relationship of *Panax ginseng* and *P. quinquefolius* inferred from sequencing and comparative analysis of expressed sequence tags. Genet Resour Crop Ev. 2013;60(4):1377-1387**.**
29. Ha WY, Shaw PC, Liu J, Yau FC, Wang J. Authentication of *Panax ginseng* and *Panax quinquefolius* using Amplified Fragment Length Polymorphism (AFLP) and Directed Amplification of minisatellite region DNA (DAMD). J Agr Food Chem. 2002;50(7):1871–1875.
30. Sticher O. Biochemical, Pharmaceutical, and Medical Perspectives of Ginseng. Phytomedicines of Europe. 1998;Chapter 16:221–240.
31. D'Agostino N, Pizzichini D, Chiusano ML, Giuliano G. An EST database from saffron stigmas. BMC Plant Biol. 2007;7:53.
32. Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe BA, Wang Y. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. BMC Genomics. 2010;11:94
33. Sui C, Wei JH, Chen S-L, Chen H-Q, Dong LM, Yang C-M. Construction of a full-length enriched cDNA library and analysis of 3111 ESTs from roots of *Bupleurum chinense* DC. Bot Stud. 2010;51:7-16.
34. Hao G, Shi R, Wang J, Qi B. Construction of full-length cDNA library of white flower *Salvia miltiorrhiza bge f.alba* root and partial EST sequence analysis. Afr J Biotechnol. 2009;8(10):2027-2035.
35. Jiao P, Chang Q, Yang MF, Zhang F, Zhao XM, Xia ZL. Effects of *Salvia miltiorrhiza Bge f.alba* root preparation on the proliferation and apoptosis of gastric cancer cells. World J Gastroentero. 2007;15(8):820-823.
36. Zhou L, Zuo Z, Chow MSS. Danshen.: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use. J Clin Pharmacol. 2005;45(12):1345-1359.
37. Dittami Simon M, Michel G, Collẻn J, Boyen C, Tonon T. Chlorophyll-binding proteins revisited – a multigenic family of light- harvesting and stress proteins from a brown algal perspective. BMC Evol Biol. 2010;10:365.
38. Tarrago L, Laugier E, Rey P. Protein-repairing methionine sulfoxide reductases in photosynthetic organisms: gene organization, reduction mechanisms, and physiological roles. Mol Plant. 2009;2(2):202–217.

39. Torii KU. Leucine-rich repeat receptor kinases in plants: structure,function, and signal transduction pathways. Int Rev Cytol. 2004;234:1-46.
40. Stein H, Honig A, Miller G, Erster O, Eilenberg H, Csonka LN, Szabados L, Koncz C, Zilberstein A. Elevation of free proline and proline-rich protein levels by simultaneous manipulations of proline biosynthesis and degradation in plants. Plant Sci. 181. 2011;2:140-150.
41. Maris A, Suslov D, Fry Stephen C, Verbelen Jean - Pierre, Vissenberg K. Enzymic characterization of two recombinant xyloglucan endotransglucosylase/ hydrolase (XTH) proteins of *Arabidopsis* and their effect on root growth and cell wall extension. J Exp Bot. 2009;60(13):3959–3972.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=279&id=3&aid=2217