UNDER PEER REVIEW

Research paper

Establishment of cDNA library and EST analysis from leaves of *Phyllanthus amarus*

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 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

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

UNDER PEER REVIEW

is subdivided into 10 or 11 subgenera. [33; 3]. *Phyllanthus amarus* Schum & Thonn. popularly known as bhuiamlaki, is found in different parts of India.

Taxonomy

Kingdom: Plantae

Division: Angiospermae

Class : Dicotyledoneae

Order : Tubiflorae

Family : Euphorbiaceae

Genus : *Phyllanthus*

Species: *amarus*

Common names : Phyllanthus amarus in North, Central and South America are black catnip, carry-meseed, chanca piedra, djari-bita, egg woman, fini-bita, flor escondida, gale-of-(the)-wind, hurricane weed, quebra-pedra, quinine creole, quinine weed, seed-under-leaf, stone breaker and yerba de la nina [16].

This plant is a common pantropical weed that grows well in moist, shady and sunny places. It is a small erect, tropical annual herb whose stem has a green capsule and grows up to 10-60 cm tall with smooth cylindrical stem, 1.5 to 2 mm thick and deciduous horizontal branchlets 4 to 12 cm long and about 0.5 cm thick, with 15 to 30 leaves. The plants are monoecious or homogamous. Leaves are simple about 5 to 11 mm long and 3 to 6 mm wide, rounded to slightly pointed at the tip, scarcely oblique on one side at the base; alternate or opposite, on petioles 0.3 to 0.5 mm long, elliptic, oblong or obovate, some are leathery; flowers are very small and diclinous, they cluster in cup-shaped structures, greenish, often with glands. The flowers are alone or usually one male and one (larger) female are in each leaf axil together. The fruit is a three-lobed capsule extending from the cup and commonly the long stalk pendant [14].The seed capsules on stalks are 1 to 2 mm long, round, smooth, 2 mm wide, with 6 seeds. *Phyllanthus 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 [3; 29]. 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. [31; 30; 17; 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

UNDER PEER REVIEW

and wounds. It also has antimutagenic activities [11], antinociceptive activities [12] and antilipidemic potentials [1; 2].The organic compounds of medicinal interest present in this herb include lignans – phyllanthin and hypophyllanthin [24; 25], geraniin and 5 flavanoids -quercetin, astralgin, quercetrin, isoquercetin and rutin [16; 7]. It also contains minor compounds like hydrolysable tannins like phyllanthusiin D [10], amariin [7], amarulone [21], amarinic acid and alkaloids like ent-norsecurinine, sobubbialine, epibubbialine [10]; diarylbutane, nyrphyllin, and a neolignan, phyllnirurin besides lactones, steroids, terpenoids, , and so forth. The plant is bitter, astringent, cooling, diuretic, febrifuge and antiseptic.

In any attempt to study the gene function of an organism, the incorporation of a cDNA-based approach is unavoidable. Indeed, one of the fundamental tools of molecular biology is the enzymatic conversion of poly (A)+ messenger RNA (mRNA) to double stranded (ds) DNA, and the insertion of this DNA into both prokaryotic and eukaryotic vectors [13]. The genetic material of the cell is composed of nucleic acids present in two forms: deoxyribonucleic acid (DNA) which makes up the chromosomes and ribonucleic acid (RNA) which decodes the genetic information encoded in the DNA and produces proteins for the cell. If we reverse this process and artificially build DNA from the RNA, it reveals the original protein-coding information contained in the genes; because this DNA is a copy of RNA, it is called complementary DNA (cDNA). cDNA is thus the reverse transcriptase product of mRNA and represents the coding sequence of all transcribed genes at the time of mRNA isolation [13]. cDNA has two advantages over chromosomal DNA: there are no introns - non-coding sequences that often occur within eukaryotic gene sequences - so it is easier to identify and characterize genes; and cDNA only represents those genes that are being actively used by the cell, since RNA polymerase only transcribes activated genes in the first place.

cDNA library is thus a random collection of cDNA fragments, typically representing the entire mRNA of a target tissue that have been inserted into a cloning vector. If a particular gene sequence is required, it may be easier to find it in a cDNA library rather than a gene library, which is larger. cDNA libraries are simpler to construct, because cDNA fragments, like their parental mRNAs, are already fairly short, so an entire cDNA can be spliced into a single vector. Other advantages include: since there are no introns, so there is no danger of pieces of the target gene being chopped onto separate clones; the library is enriched for a target gene, since instead of one or two copies, as in the genomic library, there are as many copies as the cell could produce mRNA's for that gene [18]. In order to isolate and study a particular eukaryotic gene, transcribed at a certain time or under certain conditions, a cDNA clone is usually isolated. The general

UNDER PEER REVIEW

method involves the construction of a cDNA library representing the mRNA population using poly(A)+ RNA, extracted from the appropriate tissue or cell type, followed by the identification of the cDNA clone of interest.

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 [8]. Various modifications followed over the years to improve the method's versatility [23], but the procedure to synthesise cDNA has not been extensively changed, with the possible exception of the construction of uniform-abundance (normalised) cDNA libraries [20].

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 [19].

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 [6; 22]. Resources pertaining to EST analysis are maintained at http://biolinfo.org/EST/.

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 *Phyllanthus 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

UNDER PEER REVIEW

129 **2.1 Plant material**

130 *Phyllanthus amarus* cultivated in the garden of Indian Institute of Chemical Biology, Kolkata was
131 collected and the fresh young leaves of the plant were used as the experimental material for the
132 construction of a cDNA library of this plant.

133

134 **2.2 RNA isolation and mRNA purification**

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

139

140 **2.3 Construction of cDNA library**

141 1 µg of DNase free total RNA, isolated from young leaves, was used for the synthesis of first
142 strand cDNA according to the protocol of SMARTer™ cDNA Library Construction Kit (Clontech
143 Lab Inc, USA). The first step involved mixing of 1µl of RNA with 1µl of 12µM 3' SMART CDS
144 Primer II A in addition to 2.5 µl of deionized water in a total of 4.5 µl of reaction mixture and
145 incubated at 72^{0}C for 3 minutes with a further reduction of temperature to 42^{0}C for 2 minutes.
146 Meanwhile during the incubation period a master mix was prepared at room temperature to a total
147 volume of 5.5µl using 2µl of 5X first strand buffer, 0.25µl of 100mM DTT, 1µl of 10mM dNTP mix,
148 1µl of 12µM SMARTer II A Oligonucleotide, 0.25µl of RNase inhibitor and 1µl of SMARTScribe™
149 Reverse Transcriptase. This total volume of the master mix was added to the reaction mixture of
150 the previous step and the entire mixture was incubated at 42^{0}C for 90 minutes. Amplification of
151 cDNA and preparation of double stranded cDNA (ds cDNA) was performed by long distance PCR
152 (LD PCR) using 2µl of the total first strand cDNA synthesized storing the remaining first strand
153 cDNA at -20^{0}C. The entire process of LD PCR was done in duplicates i.e. two separate 100µl
154 PCR reactions were set up and each combining 2µl of first strand cDNA, 10µl of 10X Advantage
155 2 PCR Buffer, 2µl of 10mM 50X dNTP Mix, 2µl of 12µM 5' PCR Primer II A, 2µl of 50X Advantage
156 2 Polymerase Mix along with 82µl of deionized water. The reactions were placed in a thermal
157 cycler preheated to 95^{0}C and each reaction was subjected to 15 cycles after which the PCR

UNDER PEER REVIEW

158 program was paused and 30µl from each tube was transferred to a second reaction tube named
159 the "optimization tube" which was further used for the remaining additional PCR reactions. The
160 other reaction tube containing 70µl of the 15th cycle PCR reaction mix named the "experimental
161 tube" was stored at 4^{0}C. 5µl from the 15 cycle PCR reaction (optimization tube) was transferred
162 to another tube for agarose/EtBr gel analysis. To prevent distortion of representation of cDNA
163 molecules within the total cDNA population and to maintain the double-stranded cDNA in the ex-
164 ponential phase of amplification, the optimal number of PCR cycles was determined by sampling
165 the remaining 25µl of PCR mixture of the optimization tubes by running three additional cycles
166 (for a total of 18) and increments of three cycles thereafter, up to 27 cycles. 5µl from each of the
167 PCR mixtures of 21st, 24th, and 27th cycle was aliquoted and kept separately and was
168 electrophoresed along with the 18th cycle PCR aliquot and 1kb DNA marker on a 1.1%
169 agarose/EtBr gel in 1X TAE buffer to visualize the amount and the size distribution of the cDNA
170 (Figure 1). Following the LD PCR, the ds cDNA was purified and size-fractioned using the
171 CHROMA SPIN™ +TE-1000 Column. According to the protocol of the CHROMA SPIN +TE-1000
172 column was prepared and centrifuged at 700g for 5 minutes to purge the equilibration buffer and
173 semi-dry the gel matrix. In the next step the cDNA sample was added to the center of the flat
174 surface of the gel matrix carefully to prevent the flow of the sample into the inner wall of the
175 column and the centrifuged at 700g for 5 minutes. The purified sample of cDNA obtained was
176 pooled into a clean microcentrifuge tube and ethanol precipitated. According to the protocol 1.5µl
177 of Glycogen (20µg/µl), 10µl of 3M Sodium Acetate (pH 4.8), 250µl of ice-cold ethanol (100%)
178 were added one by one and the tube was placed in –20°C freezer for 1 hr to precipitate the
179 cDNA. After centrifugation at 14,000 rpm for 20 minutes and careful removal of the supernatant,
180 the pellet was washed using 100µl 70% ethanol. The tube was again centrifuged at 14,000 rpm
181 for 5 minutes at room temperature and carefully removing the remaining supernatant the pellet
182 was air dried for 10 minutes. The cDNA was then resuspended in 15µl of TE buffer. 2-5µg of ds
183 cDNA obtained was ligated into pGEM-T vector (Promega) using the TA cloning procedure and
184 for the availability of the M13 forward and the reverse primers. Next after 18 hours of ligation at
185 4^{0}C the ligation mixture was transferred into the competent cells of *Escherichia coli* strain JM109
186 by the $CaC1_2$ transformation method. To screen white plaques with insert fragment, the
187 transformed cells were plated on 140 mm Petri dishes containing LB/ampicillin/ IPTG/X-Gal and
188 were placed overnight at 37^{0}C. The colonies were counted and the the ratio of white plaques
189 (recombinants) to blue plaques (non-recombinants) was rapidly estimated as the recombination
190 efficiency.
191

UNDER PEER REVIEW

192 **2.4 Sequencing and EST Analysis**

193 The white bacterial colonies obtained from the library were randomly picked and cultured on
194 selective LB media and grown overnight at 37^0C shaker. The plasmid DNA was isolated from the
195 randomly picked colonies using QIAprep Plasmid DNA isolation kit. The presence of inserts was
196 confirmed by restriction enzyme analysis done with EcoRI in the presence of suitable buffers at
197 37^0C for 1½ hours. The restriction enzyme products were analyzed on a 1.1% agarose/EtBr gel.
198 After clones' analysis, they were sent to TCGA (The Center for Genomic Application, New Delhi,
199 India) for sequencing. Single pass sequenced clones were used for downstream applications.

200

201 **2.5 Editing of ESTs and BLASTN Analysis**

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

209

210 **3. RESULTS AND DISCUSSION**

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

UNDER PEER REVIEW

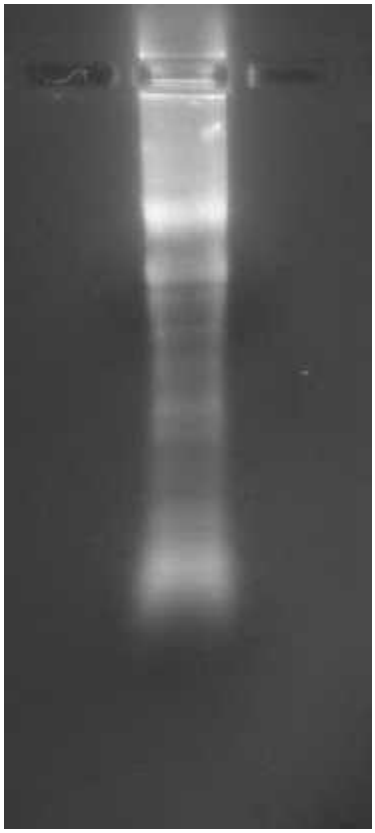

Figure 1: 1.5% agarose-formaldehyde gel showing RNA isolated from fresh young leaves of *Phyllanthus amarus*.

DNA Marker	18 cycles	21 cycles	24 cycles	27 cycles

UNDER PEER REVIEW

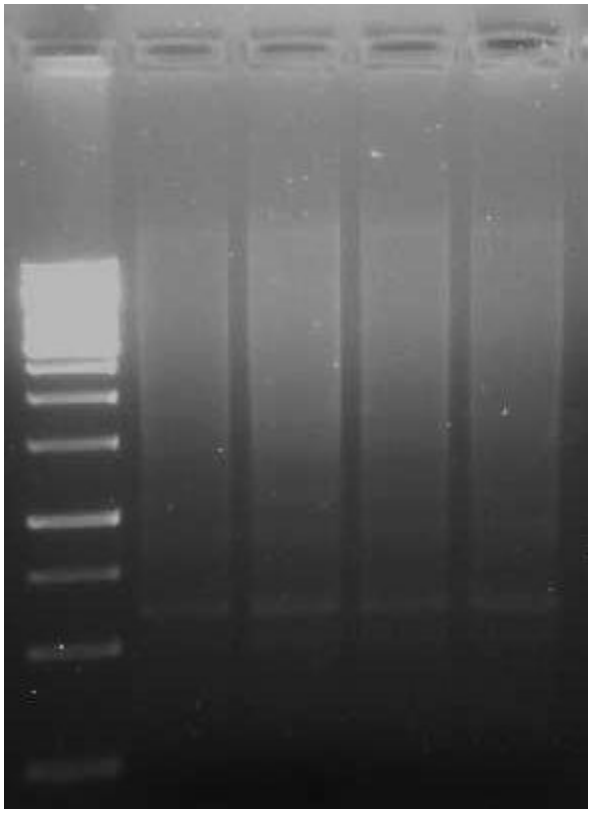

Figure 2 : LD PCR cycles yielding smear ranging in size from 0.1 to 4 Kb.

3.1 Restriction enzyme analysis of random clones obtained from the cDNA library

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

3.2 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 Figure 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.

UNDER PEER REVIEW

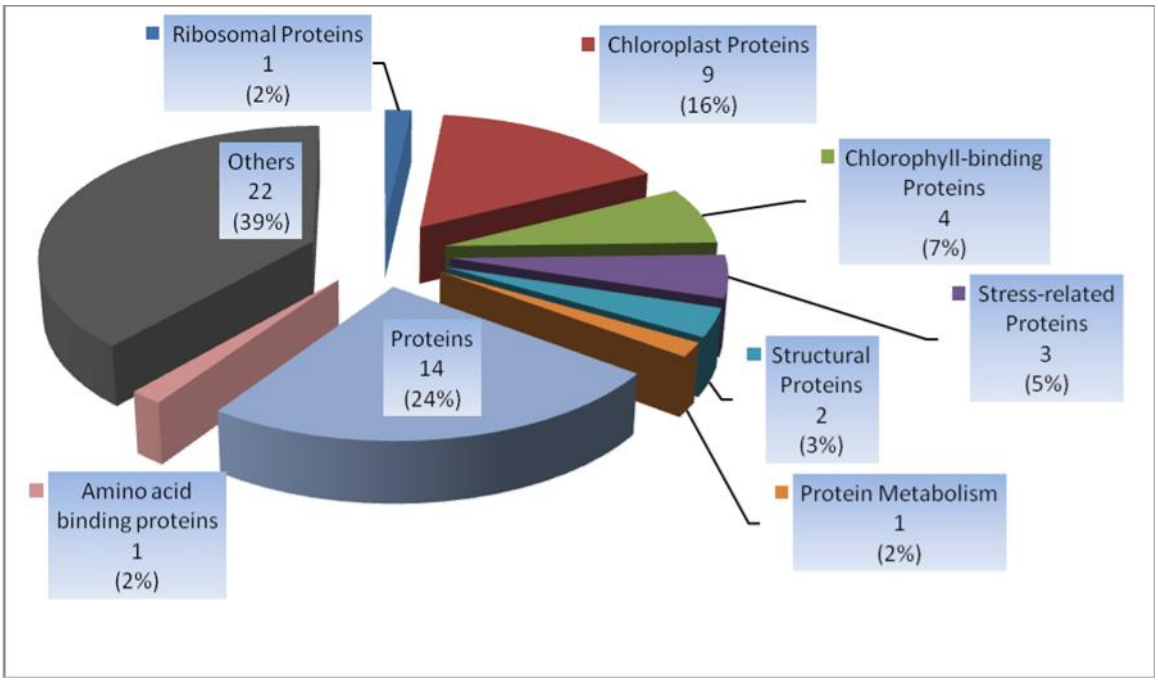

Figure 3: Functional categorization of 57 sequences into nine classes according to BLASTX analysis.

Ribosomal proteins/protein metabolism included S28 and rubber elongation factor protein 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 which is one of the important chloroplastic proteins related to photosynthesis. Besides chloroplast membrane protein ALBINO 3 and chloroplast ferredoxin I (fdn-1) mRNA were also obtained.

Chlorophyll-binding proteins included chlorophyll A/B binding and LHCII type II chlorophyll a/b-binding (CipLhcb2) protein types. Chlorophyll-binding proteins constitute a large family of proteins and have diverse functions in both light-harvesting and photoprotection [5]. Stress-related proteins included methionine sulfoxide reductase type (PtrMsrB2), mRNA and ESTs derived from severe to mild drought stressed leaves of *Populus*. Methionine sulfoxide reductase (MSR) proteins fulfill essential functions in stress tolerance and during ageing in bacterial, yeast, and

UNDER PEER REVIEW

mammal cells. MSRs are key components in the control of oxidative damage associated with the development of disorders and the process of ageing [28].

Structural proteins included leucine-rich repeat-containing and proline-rich proteins. 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 [32]. Proline-rich protein's (PRP) assembly in the plasma membrane and cell wall may increase tolerance to certain stresses [26].

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 endotransglucosylase/hydrolase protein, hexaubiquitin and ubiquitin-like proteins, latex plastidic aldolase-like protein. Xyloglucan endotransglucosylase/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 [15]. Protein metabolism type and amino acid-binding protein included rubber elongation factor protein etc.

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)	Identities (%Identity)	e-value
Ribosomal proteins							
1.	PA lib 96	*Ricinus communis* ribosomal protein S28, putative, mRNA	XM002531546	JK492943	700	83%	2e-53
Chloroplast Proteins							

UNDER PEER REVIEW

1.	PA lib 7	*Eucalyptus globulus* EgRBCS1 mRNA for ribulose-1,5bisphosphate carboxylase/ oxygenase small subunit, complete cds	AB537497	JK492909	1132	79%	3e-111
2.	PA lib21	*Arabidopsis thaliana* chloroplast membrane protein (ALBINO3) (At2g28800) mRNA, complete cds	BT002436	JK492915	1174	75%	2e-76
3.	PA lib 26	*Ricinus communis* Ribulose bisphosphate carboxylase small chain, chloroplast precursor, putative, mRNA	XM002531578	JK492964	743	78%	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%	1e-31
5.	PA lib 42	*Corchorus olitorius* ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 1A mRNA, complete cds	EU024512	JK492923	938	79%	8e-130
6.	PA lib 49	*Citrus reticulata* ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit precursor, gene, partial cds, nuclear gene for	AF312228	JK492926	668	80%	1e-87

UNDER PEER REVIEW

		chloroplast product					
7.	PA lib 99	*Corchorus olitorius* ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit 1A mRNA, complete cds	EU024512	JK492944	547	82%	9e-101
8.	PA lib 100	*Linum usitatissimum* cultivar Nike ribulose bisphosphate carboxylase activase mRNA, partial cds	GU581046	JK492945	587	82%	4e-11
9.	PA lib 57	*Eucalyptus globulus* EgRBCS1 mRNA for ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit, complete cds	AB537497	JK492931	1134	76%	4e-84
Chlorophyll-binding proteins							
1.	PA lib 65	*Jatropha curcas* chlorophyll A/B binding protein mRNA, complete cds	GQ984214	JK492933	1181	82%	0.0

UNDER PEER REVIEW

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%	3e-56
3.	PA lib 74	*Ricinus communis* chlorophyll A/B binding protein, putative, mRNA	XM002524 570	JK492936	820	85%	0.0
4.	PA lib 101	*Gossypium hirsutum* chloroplast photosystem II chlorophyll A/B-binding protein gene, complete cds	L07119	JK492946	1130	84%	0.0
Stress-related proteins							
1.	PA lib 94	*Populus trichocarpa* methionine sulfoxide reductase type (PtrMsrB2), mRNA	XM002316 839	JK492941	676	80%	2e-116
2.	PA lib 35	*Populus* EST from severe drought-stressed leaves	CU226865	JK492948	1109	72%	9e-48
3.	PA lib 39	*Populus* EST from mild drought-stressed leaves	CU232168	JK492949	524	85%	2e-59
Structural Proteins							

UNDER PEER REVIEW

1.	PA lib83	*Ricinus communis* leucine-rich repeat-containing protein, putative, mRNA	XM002522286	JK492938	821	75%	5e-68
2.	PA lib 93	*Gossypium hirsutum* hybrid proline-rich protein (HyPRP4) mRNA, complete cds	HM989876	JK492940	931	74%	8e-35
Protein Metabolism							
1.	PA lib 95	*Ricinus communis* Rubber elongation factor protein, putative, mRNA	XM002512381	JK492942	1087	77%	2e-63
Proteins							
1.	PA lib 32	*Populus trichocarpa* predicted protein, mRNA	XM002303120	JK492947	555	78%	2e-72
2.	PA lib 34	*Ricinus communis* Xyloglucan endotransglucosylase/ hydrolase protein 22 precursor, putative, mRNA	XM002526182	JK492921	1124	83%	0.0
3.	PA lib 47	*Zea mays* uncharacterized protein LOC100382066 (LOC100382066), mRNA	NM001174829	JK492954	1174	100%	5.7
4.	PA lib 41	*Ricinus communis* conserved hypothetical protein, mRNA	XM002528323	JK492951	806	83%	9e-103

UNDER PEER REVIEW

5.	PA lib 43	*Arabidopsis lyrata* subsp. lyrata predicted protein, mRNA	XM002873053	JK492952	1112	77%	3e-105
6.	PA lib 50	*Ricinus communis* Xyloglucan endotransglucosylase/hydrolase protein 22 precursor, putative, mRNA	XM002526182	JK492927	1164	83%	0.0
7.	PA lib 51	*H.annuus* UbB1 mRNA for hexaubiquitin protein	X57004	JK492928	584	68%	3e-12
8.	PA lib 52	*Pisum sativum* ubiquitin-like protein mRNA, partial cds	EF212231	JK492929	704	78%	5e-55
9.	PA lib 59	*Arabidopsis lyrata* subsp. *lyrata* hypothetical protein, mRNA	XM002876737	JK492932	1171	67%	2e-55
10.	PA lib 63	*Ricinus communis* conserved hypothetical protein, mRNA	XM002519900	JK492956	721	81%	8e-78
11.	PA lib 76	*Populus 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

UNDER PEER REVIEW

14.	PA lib 102	*Hevea brasiliensis* latex plastidic aldolase-like protein mRNA, complete cds	AF467803	JK492962	1132	82%	0.0
Amino acid binding proteins							
1.	PA lib 28	*Ricinus communis* amino acid binding protein, putative, mRNA	XM002522978	JK492917	1141	81%	0.0
Others							
1.	PA lib 30	*Jatropha curcas* nucleoside diphosphate kinase B (NDKB) mRNA, complete cds	FJ906846	JK492919	779	87%	1e-151
2.	PA lib 38	*Populus trichocarpa* galactinol synthase 1 (GOLS1) mRNA, complete cds	EU305718	JK492922	1187	80%	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	*Vitis vinifera* contig VV78X157397.15, whole genome shotgun sequence	AM432029	JK492925	922	100%	6.1

UNDER PEER REVIEW

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%	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), nuclear gene encoding mitochondrial protein, mRNA	NM001086 864	JK492939	348	100%	5.4
10.	PA lib 103	*Solanum lycopersicum* cDNA, clone: LEFL1030AH04, HTC in leaf	AK247080	JK492963	900	73%	3e-28
11.	PA lib 4	*Populus trichocarpa* lysine/histidine transporter, mRNA	XM002298 561	JK492908	907	78%	2e-142
12.	PA lib 9	*Ricinus communis* glyceraldehyde 3-phosphate dehydrogenase, putative, mRNA	XM002512 744	JK492910	1113	81%	0.0
13.	PA lib 11	*Drosophila erecta* GG12460 (Dere\GG12460), mRNA	XM001982 143	JK492911	125	84%	1.6
14.	PA lib 12	*Populus trichocarpa* clone ACSB2121-B05, complete sequence	AC210380	JK492912	1170	74%	3e-155
15.	PA lib 15	*Populus trichocarpa* clone WS0127_L10	EF148116	JK492914	427	85%	1e-59

UNDER PEER REVIEW

		unknown mRNA					
16..	PA lib 25	TSA: *Jatropha 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	*Ricinus communis* ATP-dependent Clp protease proteolytic subunit, putative, mRNA	XM002515 227	JK492930	1175	82%	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	*Medicago sativa* class IV chitinase mRNA, complete cds	FJ487629	JK492957	913	89%	0.008
22.	PA lib 66	*Populus tomentosa* calmodulin mRNA, complete cds	AY836672	JK492958	657	88%	1e-169

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.

UNDER PEER REVIEW

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. Adeneye A.A, Benebo A.S, Agbaje E.O. Protective effect of the aqueous leaf and seed extract of Phyllantthus amarus on alcohol-induced hepatotoxicity in rats. West African Journal of Pharmacology and Drug Research. 2006; 23: 42-50.

2. Adeneye A.A, Amole O.O, Adeneye A.K. The hypoglycemic and hypocholesterolemic activities of the aqueous leaf and seed extracts of Phyllanthus amarus in mice. Fitoterapia. 2006; 77:511-514.

3. Calixto JB , Santos AR , Cechinel Filho V , Yunes RA. A review of the plants of the genus *Phyllanthus*: Their chemistry, pharmacology, and therapeutic potential. Medicinal Research Reviews. 1998; 18 (4): 225-258.

4. Chevallier, A. Encyclopedia of Herbal Medicine: Natural Health. Dorling Kindersley Book. USA. 2000; Second edition : 336.

5. 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 Evolutionary Biology. 2010; 10: 365.

6. Dong Q, Kroiss L, Oakley FD, Wang BB, Brendel V.Comparative EST analyses in plant systems. Methods in Enzymology. 2005; 395: 400–418.

UNDER PEER REVIEW

7. Foo L.Y. Amarulone, a novel cyclic hydrolysable tannin from *Phyllanthus amarus*. Natural Product Letters. 1993; 3: 45-52.

8. Gubler U, Hoffman B.J. A simple and very effective method for generating cDNA libraries. *Gene*. 1983; 25: 263-269.

9. Harish R, Shivanandappa T. Antioxidant activity and hepatoprotective potential of *Phyllanthus niruri.* Food Chemistry. 2006; 95 (2):180–185.

10. Houghton P.J, Woldemariam T.Z, O'Shea S, Thyagarajan S.P. Two securinega-type alkaloids from *Phyllanthus amarus*. Phytochemistry. 1996; 43 (3): 715-717(3).

11. Joy, K.L, R. Kuttan. Inhibition by Phyllanthus amarus of hepatocarcinogenesis induced by N-nitrosodiethylamine. J.Bioch.Nutr. 1998; 24: 133-139.

12. Kassuya C.A, A.A.Silerstre, V. Rehder, Calixto J.B. Anti allodynic and antiedematogeni properties of the lignin from *Phyllanthus amarus* in models of persistent inflammatory and neuropathic pain. Eur.J.Pharm. 2003; 478: 145-153.

13.Kimmel, A.R, Berger, S.L. Preparation of cDNA and the generation of cDNA libraries. Methods in Enzymology. 1987; 152: 307-316.

14. Lewis, W.H, Elvin-Lewis, P.F Medical Botany Plants Affecting Man's Health. A Wiley-Interscience Publication. John Wiley & Sons, New York- London-Sydney-Toronto. 1977; 515 .

15. 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. Journal of Experimental Botany, 2009; 60 (13): 3959–3972.

16. Morton, J.F. Atlas of Medicinal Plants of Middle America. Library of Congress cataloging in Publication Data. Thomas books. 1981; 1420.

UNDER PEER REVIEW

17. 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.

18. Old R.W., Primrose S.B., Principals of gene manipulation: an introduction to genetic engineering. Chapter 6, 5th edition, Blackwell science Ltd. 1994.

19. Parkinson J., Blaxter M., Expressed Sequence Tags: an overview. Methods in Molecular Biology. 2009; 533 : 1-12.

20. Patanjali, S, Parimoo, S. Weissman, S.M. Construction of a uniform-abundance (normalized) cDNA library. Proceedings of the National Academy of Sciences. USA. 1991; 88: 1943-1947.

21. Rao G.S, R. Bramley. Hypophyllanthin.Tetrahedron Letters. 1971; 34: 3175-3178.

22. Rudd S, Expressed sequence tags: alternative or complement to whole genome sequences? Trends in Plant Science. 2003; 8 (7): 321–329.

23. Rutledge R.G, Seligy, V.L, Cote, M.J, Dimock, K, Lewin L.L, Tenniswood M.P. Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors. Gene 68. 1988; 151-158.

24. Sharma A, K.Lal, S.S. Handa. Estimation of phyllanrthin and hypophyllanthin by high performance liquid chromatography in *Phyllanthus amarus*. Phytochemical analysis. 1993; 4: 226 - 229.

25. Somanabandhu A, S.Nitayangkura, C. Mahidol, S. Ruchirawat, K.Likhitwitayawuid. 1H- and 13C- nmr assignments of phyllanthin and hypophyllanthin: Lignans that enhance cytotoxic responses with cultured multidrug resistant cells. Journal of Natural Products. 1993; 56: 223 - 239.

26. 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

UNDER PEER REVIEW

manipulations of proline biosynthesis and degradation in plants. Plant Science 181. 2011; 2: 140-150.

27. Sterck L, Rombauts S, Vandepoele K., Rouzé P., Van de Peer Y.. How many genes are there in plants (...and why are they there?). Current Opinion in Plant Biology 2007; 10(2) :199-203.

28. Tarrago L, Laugier E, Rey P. Protein-repairing methionine sulfoxide reductases in photosynthetic organisms: gene organization, reduction mechanisms, and physiological roles. Molecular Plant. 2009; 2(2): 202–217.

29. The Wealth of India; A dictionary of Indian raw materials and industrial product. CSIR, New Delhi, India. 2003; 34–36.

30. Thyagarajan SP, Jayaram S, Gopalakrishnan V, Hari R, Jeyakumar P, Sripathi MS,. Herbal medicines for liver diseases in India. Journal of Gastroenterology and Hepatology. 2002; 17:370-376.

31. 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.

32. Torii KU. Leucine-rich repeat receptor kinases in plants: structure,function, and signal transduction pathways. International Review of Cytology. 2004; 234: 1-46.

33. Unander, D.W, Webster, G.L Blumberg, B.S. Usage and bioassays in *Phyllanthus* (Euphorbiaceae). IV. Clustering of antiviral uses and other effects. Journal of Ethnopharmacology. 1995; 45: 1-18.

Abbreviations

EST: Expressed Sequence Tag

NCBI: National Centre for Biotechnology Information

BLAST: Basic Local Alignment Search Tool