Plant Tissue Cult. & Biotech. **17**(2): 173-182, 2007 (December)

Novel ESTs from a Jute (*Corchorus olitorius* L.) cDNA Library

Mohamad W. Wazni, Ahmad S. Islam, J. Matthew Taliaferro, Nabila Anwar and Kanagasabapathi Sathasivan*

Molecular, Cell and Developmental Biology, School of Biological Sciences, The University of Texas at Austin, USA-78712

Key words: Jute cDNA library, *Corchorus olitorius,* Chitinase, Chaperonin, RSH2, RelA/SpoT, V-ATPase, Small subunit RuBisCO, Metallothionein, Actin 7, and HSPs

Abstract

The paper describes the identification of new expressed sequence tags (ESTs) from a cDNA library of *Corchorus olitorius* L. var. O-4 after the initial description of the library construction reported recently. The sequence homology search in The Arabidopsis Information Source (TAIR) using the WU-BLAST tool revealed four complete and ten partial cDNA sequences. The complete cDNA sequences, based on their similarities to those of *Arabidopsis* proteins, encode V-ATPase subunit F, mitochondrial NADH-ubiquinone oxidoreductase, RuBisCO small subunit 1A and heat shock protein (HSP 60). Based on similar homology results the partial cDNAs encode proteins, namely, chaperonin, actin 7, RelA-spoT homology, 60s ribosomal protein L 36a, transport protein, chloroplast inner membrane import protein Tic22, formate dehydrogenase, serine hydroxy-methyltransferase, metallothionein 2B and tansmembrane kinase. These cDNA sequences are new additions to the ESTs that encode chitinase-like protein (Class I) and 60S acidic ribosomal protein reported earlier. All the available cDNA sequences including that for the chitinase-like protein have been registered with GenBank, bearing Accession Nos. EU024510 through EU024520 and EU092254, EU057193-95. The cDNA samples are available for any researchers interested in such ESTs. A Clustal-W study of the amino acid sequences of proteins encoded by cDNA isolated in the present investigation with those of cotton, citrus, *Arabidopsis*, tobacco and other organisms revealed that the homology is maximum between jute and cotton followed by citrus, grapevine, tobacco and *Arabidopsis*.

Introduction

We have embarked on a project on isolating expressed sequence tags (ESTs) of an important fiber crop, jute. Jute fibers are extracted from a number of cultivars of

*Corresponding author. E-mail: sata@mail.utexas.edu

Corchorus olitorius and *C. capsularis,* now transferred to the Malvaceae family. The biodegradable nature of jute fiber and the potential high yield of cellulose biomass per acre have increased the global interest in jute. As a result, jute breeders and biologists are now turning their attention to use molecular tools to increase its fiber quality, yield and improve agronomic traits (Khatun 2007). Furthermore, the textile and paper industry are interested in its potential as an important ingredient for producing paper, fine textiles as well as a renewable source for biofuel. In addition to our laboratory, a few other institutions currently involved in the molecular biology of jute cultivars are the Indian Institute of Technology, Kharagpur (Basu et al. 2004a,b); Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal; and the Molecular Biology Laboratory, Department of Agricultural Botany, Ch. Charan Singh University, Meerut (Mir et al. 2007) in India, and the Department of Biochemistry and Molecular Biology, Dhaka University (Hossain et al. 2002, 2003) in Bangladesh. The latter group has recently registered a gene expressing a putative low density lipoprotein B-like protein in *C. olitorius* (Ashraf et al. 2007).

The South African group at the Department of Botany, Rand Afrikaans South Africa has determined the partial cDNA sequence of tRNA-leu (UAA) and tRNA-Phe (UAA) of 12 wild *Corchorus* species based on nuclear rDNA internal transcribed spacer (ITS) of chloroplasts (Moeaha et al. 2006), while those at Taiwan are using ITS sequence of nuclear ribosomal DNA to identify two locally grown *Corchorus* species, namely, *C. capsularis* and *C. olitorius* (Liu et al. 2005). While some labs are currently pursuing the development of RAPD or SSR markers for assisting in molecular breeding, only two labs are focusing on the isolation of agronomically important genes or cDNAs.

The objective of the present study is to isolate, identify and understand the expressed sequences from the genome of jute, with a long term goal of genetically modifying jute to produce low lignin content, disease resistance, cold tolerance as well as varieties that may help remove arsenic from polluted soils. An earlier paper (Islam et al. 2005) reported several ESTs recovered from this library including the isolation of endochitinase that may help engineer disease resistant jute varieties. Here, we report the isolation of a few more cDNA clones such as metallothionein, RSH2 Rela-SpoT, RuBisCO subunits etc. with their GenBank Accession Nos. Some of these cDNAs have significant potential for genetic improvement of jute. The present investigation also lays particular emphasis on examining inserts that contained the entire open reading frame for specific proteins.

Materials and Methods

Seeds of *Corchorus olitorius* L. var. O-4 obtained from Professor Haseena Khan at the Department of Biochemistry and Molecular Biology, Dhaka University, were germinated in sterile Petri dishes and kept in the dark. RNA was extracted from Seven-day-old etiolated seedlings. The protocol for construction of cDNA library of *C. olitorius* var. O-4 from the total RNA has been described in an earlier paper (Taliaferro et al. 2006). The cDNA library used in this work was constructed using the pBluescript II XR cDNA Library Construction Kit from Stratagene. The cDNA fragments were cloned into pBluescript between the EcoRI and XhoI sites in pBluescript SK+. The recombinant plasmids were then sequenced using T3 and/or T7 primers. The plasmid DNA was isolated using Invitrogen's plasmid DNA Miniprep Kits and PvuII was used to digest the plasmids to identify the recombinant DNA clones. Selected DNA samples were sent for sequencing in the core DNA facility in the Institute of Cell and Molecular Biology at the University of Texas, Austin. Sequence homology searches were done through the WU-BLAST program in TAIR, Resource (www.arabidosis.org). Using the Pearson format explicitly set to protein, Clustal W was used for the alignment of the amino acid sequences of the encoded proteins.

The procedure followed in GenBank to register a full or partial cDNA sequences has been refined, making it more and more difficult to register a gene with NCBI. In order for a gene to be registered with GenBank detailed knowledge about some of the tools such as VecScreen, ORF finder in NCBI, WU blast in TAIR and Ex-Pasy proteomic tools is necessary. In a separate article (Britton et al. 2007), published in this issue (pp. 161-172), this procedure has been described and discussed in detail with examples. For instance, VecScreen and ORF Finder, two very useful data mining tools help a researcher find whether the insert is vector-contaminated and indicate both the initiation- and stop codons in the correct reading frame. TAIR protein blast (Blast X) helps in determining the actual reading frame in usage.

Results and Discussion

Since our last reporting of EST sequences (Taliaferro et al. 2006), we have identified several recombinant cDNA clones and analyzed their DNA sequences. Inserts with significant lengths were used to search for homologies through WU-BLAST in TAIR to known sequences of *Arabidopsis* genome. Table 1 shows a list of inserts that were found similar to those of *Arabidopsis* in the NCBI database. Both full and partial cDNA sequences representing 16 cDNAs, listed in Table 1 have been registered with GenBank, bearing Accession Nos. EU024510 through EU024520 and EU092254, EU057193-95. In addition, 29ESTs have been registered under EST Batch submission section bearing Accession Nos. EV283112- 283119, EV 2831124-31127 and ES 673254-673270.

A survey of GenBank revealed that NCBI database contains 179 ESTs of another jute species, namely, *C. capsularis* var. 321 registered by the IIT group at Kharagpur, India. A comparison of ESTs obtained in our study with those of Sadhukhan et al. (2007) showed that there were five similar sequences between the two species of *Corchorus* in respect of RuBisCO and 60S ribosomal RNA. Since both *C. olitorius* and *C. capsularis* belong to the same genus, such similarities were expected.

Table 1. ESTs showing significant homology to known sequences in *Arabidopsis* or other higher plants.

Insert No.	GenBank Acc. No.	Insert length (bp)	Protein name
423	EF641793	268	60S ribosomal protein L36a/L44 (RPL36aB)
004	ABS72187	360	Acidic 60S ribosomal protein P3* (complete)
552	ABS83240	1199	Actin 7 heat stress protein during seed germination
351	ABS72190	660	Chaperonin
165	ABS72188	798	Chitinase class I*
683	ABS83241	490	Chloroplast inner membrane import protein Tic22
478	ABS72196	216	Formate dehydrogenase
594	ABU63401	638	Heat Shock Protein (HSP) 60 (complete)
554	ABS72197	524	Metallothionein 2b
387	ABS72192	312	Mitochondrial NADH-ubiquinone oxidoreductase (complete)
344	ABS72189	552	Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) small subunit 1A (complete)
404	ABS72194	719	RSH2 (Rel-A SpoT)
470	EU024518	361	Serine hydroxymethyl transferase
536	EU258555	1009	Transmembrane kinase
359	ABS72191	795	Transport protein
399	ABS72193	393	V-ATPase subunit F (complete)

The first column indicates the sample ID No. of the plasmid cDNA library of *C. olitorius* var. O-4. *reported earlier. Complete cDNA sequence is indicated within parenthesis in the last column.

Comparison of DNA sequence homologies between C. olitorius *and some selected species using TAIR WU-BLAST analysis:* Only a summary of DNA sequence homologies is presented in Table 2. In almost all cases, bp sequence homologies between *C. olitorius* and all the three *Gossypium* species, namely, *G. hirsutum, G arboreum, G. raimondii,* were stronger than between *C. olitorius* and other species used for comparison. This was expected as both *Gossypium* and *Corchorus* belong to the family Malvaceae. See the supplemental information posted at (http://www.gnobb.org/Jute_Homology_table.doc) describing the results of Multiple Alignment for All Full Open Reading Frame Protein Sequences obtained in the present investigation.

Table 2. Summary of DNA sequence homologies between *C. olitorius* and some other dicot species using TAIR WU-BLAST analysis.

Protein encoded cDNA from *C. olitorius*	*Gossypium hirsutum*	*G. arboreum* or *raimondii*	*Citrus clementina* or *sinensis*	*Arabidopsis thaliana*	*Vitis* sp.	*Prunus* sp.	Other plants
60 S acidic ribosomal protein P3		76	72	67			
60 S ribosomal protein L36a	82		73	65			*Cacao* 85
Actin 7		*G. raimondii* 87		74			
Chaperonin	88			76			*Helianthus* and *Nicotiana* 80
Chitinase class I		*G. raimondii* 63	75	60		74	
Chloroplast import protein	76		70	68	69		
Formate dehydrogenase		*G. raimondii* 86	*C. sinensis* 86	75	83		
Heat shock protein 60				65	72		
Metallothionein 2b		77		64			*Cacao* 70
NADH-dehydrogenase	81			77		72	*Fragaria* 77
RSH2 RelA-SpoT	91			76	82	80	
RuBisCO 1A	77			79			
Serine hydroxymethyl transferase		*G. raimondii* 73	*C. sinensis* 76	84			*Populus* 74
Transmembrane kinase 1	88		76	74	81		
V-ATPase		84	79	75		80	

Importance of newly identified and selected cDNA sequence in improving agronomic traits in *C. olitorius*.

RelA/SpotT: Literature search revealed that salt tolerance is conferred by RelA/SpotT homologue (Sj-RSH) on *E. coli* transformants. Yamada et al. (2003) reported that its insertion in *E. coli* controls the amount of guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) in the organism, thereby enhancing its salt tolerance capacity. In support of their claim, they added further proof by showing that expression of this gene, driven by GAL1 promoter enhances salt tolerance in yeast also. From our point of view, this gene is very important because enhancement of its expression will pave the way to produce jute cultivars suitable for marginal lands, which currently do not support any standard jute varieties.

Metallothionein 2B: The metallothionein gene may also be exploited for the similar purpose. Overexpression of this gene in jute cultivars may potentially enable them to grow in marginal lands that at present do not support any agricultural crops. Of relevance in this context is the work of White and Rivin (1995) who reported a cDNA sequence in maize similar to Zn^{2+} associated with class II metallothionein in wheat. According to the researchers, the gene is associated with maturation process involving storage proteins and putative desiccation protectants that are synthesized and stored in roots.

Chitinase: Overexpression of cDNA sequence of the chitinase gene reported by Taliaferro et al. (2006) may exhibit enhanced levels of resistance to deadly soil borne fungal disease caused by *Rhizoctonia solani* (cf. María et al. 2006). During some years the attack by this fungus is so severe that the entire jute crop is wiped out from particular regions. Transgenic plants overexpressing chitinases may also delay disease symptoms, when challenged with fungal pathogens. Overexpression of cellulose is especially important if we were to reduce the lignin content and increase the cellulose component.

Actin 7: The *actin7* cDNA may be regarded as one of the house-keeping genes because mutation of that gene results in physiological defects as reported by Gilliland et al. (2003) in their study with the *act7 Arabidopsis* mutant. They showed that homozygous adult plants for the *act7* mutant alleles suffer from undetected physiological deficiencies.

Chaperonin: The jute seedlings were grown for a week in the dark, i.e., under stress condition. Therefore, the presence of chaperonin (cpn 60) in light-starved jute seedlings is expected and is in agreement with the study of Holland et al. (1998) who detected a significant accumulation of cpn 60 in *Nicotiana* seedlings under a brief period of stress condition.

V-ATPase: The vacuolar $H^{(+)}$-ATPases (also known as V-ATPases) are a family of ATP-dependent proton pumps that acidify intracellular compartments of eukaryotic cells. The V-ATPases have several subunits and are composed of two functional domains. The accumulation of V-ATPase in seven-day-old jute

seedlings is expected as have been reported in earlier studies. For instance, based on the results of their study in germinating pumpkin seeds, Maeshima et al. (1994) suggested that there is an accumulation of the above protein along with two other enzymes, namely, V-ATPase and VM23 in the membrane of protein storage vacuole.

Mitochondrial NADH-ubiquinone oxidoreductase: This protein is located in the inner mitochodrial membrane where it catalyzes the transfer of electrons from NADH to coenzyme Q (CoQ). This enzyme (complex I) is the first in the mitochondrial electron transport chain.

RuBisCO: Ribulose-1,5-bisphosphate carboxylase oxygenase, is an enzyme that catalyzes the first major step of carbon fixation in the Calvin Cycle: the atoms of carbon dioxide in the atmosphere are made accessible to organisms as sucrose and other molecules rich in energy. This protein also catalyzes either the carboxylation or oxygenation of ribulose-1,5-bisphosphate (also known as RuBP) with carbon dioxide or oxygen. RuBisCO has a considerable biological influence since it catalyzes the most frequently utilized chemical reaction that allows inorganic carbon to enter the biosphere. It is also a very abundant protein in leaves, and it may actually be the most abundant protein in the world. For these reasons, there are presently efforts to genetically engineer crop plants so that they may contain more effective RuBisCO proteins (Portis and Parry 2007).

Heat shock proteins (HSPs): When a plant is exposed to stress (i.e., elevated temperatures), this protein increases in expression via transcription regulation (Queitsch et al. 2000). Heat shock factor (HSF) is a significant part of HSPs and is a primary inducer of HSP upregulation when a plant is exposed to stress. HSPs are named in relation to their molecular weights. For instance, Hsp60 (the one reported in this article) refers to the group of heat shock proteins that are 60 kilodaltons in size. Hsp60 plays a part in protein folding after it is imported after post-translation to the mitochondrion or chloroplast.

Acidic ribosomal protein: Ribosomes are organelles that catalyze protein synthesis. They typically consist of a small 40S subunit and a large 60S subunit (this subunit is reported in this article). The ribosomal protein L5 (RPL5) gene encodes a ribosomal protein from the L18P family found in the cytoplasm. This L18P protein is a part of the 60S subunit.

Conclusion

Construction of cDNA or/and genomic library of an organism requires teamwork not only of one but many laboratories across the globe. Such worldwide collaboration, if established will hasten the completion of DNA based sequencing as have been demonstrated in crops, animals, microbes in addition to the human genome. A start has been made to construct cDNA libraries of the

two species of jute, namely *C. capsularis* and *C. olitorius* and in two laboratories, namely, in the Indian Institute of Technology, Kharagpur and at the University of Texas, Austin, respectively. The preliminary results reported by a number of laboratories on both construction of cDNA library as well on development of molecular tools may lead to the development of a better jute crop with improved disease resistance, low lignin content, cold tolerance and abiotic stresses. Production of bioengineered jute varieties with low lignin content and disease resistance may lead to the possibilities of growing jute to supply basic raw material to paper mills and ligno-cellulose for biofuels.

Acknowledgement

Part of the work was funded by USDA grant awarded to Professor Haseena Khan, Department of Biochemistry and Molecular Biology, Dhaka University (DU). Part of the funding for this project came from the Undergraduate Research Fellowship by University CoOp at the University of Texas, Austin. The authors sincerely acknowledge the technical help by George Britton, Xuan Yue and Victor Bui at The University of Texas, Austin in isolating plasmid DNA from a number of cDNA clones and submitting them for sequencing. We are grateful to Sohel Shamsuzzaman and Sharmin Jahan at the Department of Biochemistry and Molecular Biology, DU for their help in analyzing the cDNA sequences.

References

Basu A, Ghosh M, Meyer R, Powell W, Basak SL and **Sen SK** (2004) Analysis of Genetic Diversity in Cultivated Jute Determined by Means of SSR Markers and AFLP Profiling. Crop Sci. **44**: 678-685.

Britton GL, Islam AS, Xuan Y and **Sathasivan K** (2007) Procedure to Identify and Submit cDNA Sequences to GenBank. Plant Tissue Cult. & Biotech. **17**(2): 165-176.

Gilliland LU, Pawloski LC, Kandasamy MK and **Meagher RB** (2003) Arabidopsis actin gene ACT7 plays an essential role in germination and root growth Plant J. **33**(2): 319-28.

Holland N, Belkind A. Holland D, Pick U and **Edelman M** (1998) Stress responsive accumulation of plastid chaperonin 60 during seedling development. Plant J. **13**(3): 311-316.

Hossain MB, Haque S and **Khan H** (2002) DNA fingerprinting of jute germplasm by RAPD. J. Biochem and Mol. Biol. **35**(4) : 414-419.

Hossain MB, Haque S and **Khan H** (2003) Distinction between cold sensitive and cold tolerant jute by a combined RAPD and AFLP study. J. Biochem and Mol. Biol. **36**(5): 427-432.

Islam AS, Taliaferro JM, Lee CT, Ingram CI, Montalvo RJ, Van der Ende G, Alam S, Siddiqui J and **Sathasivan K** (2005) Preliminary Progress in Jute (*Corchorus* species) Genome Analysis. Plant Tissue Cult. & Biotech. **15**(2): 145-156.

Khatun Asma (2007) Recent agricultural developments in jute, kenaf and mesta through traditional and biotechnological approaches – a seminar talk held in Myanmar in February 2007 under the auspices of International Jute Study Group

Maeshima M, I. Hara-Nishimura, Takeuchi Y and **Nishimura M** (1994) Accumulation of Vacuolar H^+-Pyrophosphatase and H^+-ATPase during Reformation of the Central Vacuole in Germinating Pumpkin Seeds. Plant Physiol. **106**: 61-69.

María de las Mercedes Dana, José A. Pintor-Toro and **Beatriz Cubero** (2006) Transgenic Tobacco Plants Overexpressing Chitinases of Fungal Origin Show Enhanced Resistance to Biotic and Abiotic Stress Agents. Plant Physiol. **142**: 722-730.

Mir RR, Rustgi S, Sharma S, Ravinder Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Mohit K. Sinha MK, Balyan HS and **Gupta PK** (2007) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. Euphytica (online).

Portis Jr. AR and **Parry MAJ** (2007) *Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective*. Photosynth Res **94**: 121–143

Queitsch C, Hong S-W, Vierling E and **Lindquist S** (2000) Heat Shock Protein 101 Plays a Crucial Role in Thermotolerance in *Arabidopsis*. Plant Cell **12**: 479-492.

Taliaferro JM, Islam AS and **Sathasivan K** (2006) Expressed Sequence Tags (ESTs) from a Jute (*Corchorus olitorius*) cDNA Library. Plant Tissue Cult. & Biotech. **16**(2): 95-104.

White CN and **Rivin CJ** (1995) Characterization and expression of a cDNA encoding a seed-specific metallothionein in maize (Plant Gene Register) Plant Physiol. **108**: 831-832.

Yamada A, Tsutsumi K, Tanimoto S and **Ozeki Y** (2003) Plant RelA/SpoT Homolog Confers Salt Tolerance in *Escherichia coli* and *Saccharomyces cerevisiae*. Plant & Cell Physiol. **44**(1): 3-9.

The references cited below are available from GenBank.

Ashraf N, Awal A, Haque S, Obaid M and **Khan H** (2007) Development of Molecular markers and identification of an expressed sequence linked to low temperature tolerance trait in jute (Accession Number EF208753).

Basu A, Maiti MK and **Sen SK** (2004 a) Cloning and sequencing of mRNA for caffeoyl-CoA-o-methyltransferase from *Corchorus capsularis* (AY500159) complete CDS (908 bp) PLN 17-JAN-2004 from NCBI Genbank (Unpublished).

Basu A, Maiti MK and **Sen SK** (2004 b) Nucleotide sequence of cinnamyl alcohol cDNA isolated from *Corchorus capsularis* (AY504425) complete CD (1243 bp) PLN 13-JAN-2004 from NCBI Genbank, (Unpublished).

Liu S-L, Chang Y-S and **Hsieh C-C** (2005) 18S ribosomal RNA gene from chloroplasts, partial sequence from *Corchorus capsularis, Corchorus aestuans* var. *brevicaulis* and *Corchorus olitorius* PLN. 8-Nov.-2005 from NCBI GenBank (unpublished)

Khan Haseena et al. (2006) The development of microsatellite for genetic analysis and marker-assisted selection of jute. ICGEB funded collaborative research program completed in 2006.

Moeaha MJ, Van der Bank M and **Winter P** (2006) partial tRNA-Leu (UUA) gene, IGS and partial tRNA-Phe gene of *Corchorus asplenifolius* and 15 other species of *Corchorus* chloroplast from NCBI GenBank PLN 22-JUL-2006 (unpublished).

Montalvo R J, Ingram CM, Stone LA, Islam A and **Sathasivan K** (2005) 18S ribosomal RNA gene from chloroplast, partial sequence from *Corchorus olitorius,* partial sequence (DQ151662) PLN 24-AUG-2005 from NCBI Genbank (Unpublished).

Sadhukhan S, Basu A, Maity MK and **Sen SK** (2006) Studies on development of jute fibre through generation of ESTs from NCBI Genbank (Unpublished).

Sadhukhan S, Basu A, Maity MK and **Sen SK** (2007) Jute Plasmid library *Corchorus capsularis* var. 321. 179 entries of ESTs numbering from 00001 to 00179 representing cDNA, mRNA sequences submitted to GenBank on Aug. 11, 2006 and Feb. 22, 2007 (Unpublished).

Stone LA, Ingram CM, Montalvo RJ, Islam A and **Sathasivan K** (2005) Putative Phosphate Transport ATP-Binding Protein gene from *Corchorus capsularis,* partial sequence (DQ151661) PLN 24-AUG-2005 from NCBI Genbank (Unpublished).