Plant Tissue Cult. & Biotech. **19**(2): 207-215, 2009 (December)

PTC&B

Identification and Characterization of Ribosomal Protein S8 Gene of Jute

Salim Ahmed[1], Md. Maksudul Alam, Md. Sazzadul Islam, MD. Shafiuddin, Md. Mosharrof Hossain Mondal[1], Md. Amzad Hossain[1] and Haseena Khan*

Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh

Key words: Identification, Characterization, Ribosomal protein S8, Jute

Abstract

Identification and confirmation of ESTs (expressed sequence tags) corresponding to genomic clones is one of the most important steps in the identification of genes. In this regard, computational approaches such as *ab initio* and homology based searches using NCBI web portal, together with common laboratory approaches - PCR, RT-PCR and DNA sequencing were used to identify jute ESTs from a jute genomic library. Using degenerate primer-based gene walking from a computationally identified and experimentally verified jute EST, this study has led to the identification of the full length sequence of a jute gene, namely ribosomal protein S8. The sequence of this gene was found to be similar to ribosomal protein S8 gene of related species like *Hibiscus macrophyllus*, *Gossypium hirsutum* etc. and that of other species like *Carica papaya, Arabidopsis thaliana* etc. This gene was further characterized for determining its cellular location to the chloroplast.

Introduction

Jute is a natural fiber obtained as an extract from the bark of the jute plant belonging to the species *Corchorus*. Till to date, only 1210 nucleotide sequences of jute are deposited in the GenBank. Molecular research on jute is limited to analysis of genetic diversity of the crop using RAPD (Hossain et al. 2002) and SSR markers (Gupta et al. 1996, Gupta and Varshney 2000, Basu et al. 2004, Akter et al. 2008, Keka et al. 2008), tissue culture (Islam et al. 1992, Saha et al. 1999, Huda et al. 2007), somatic hybridization (Khatun 2007), ESTs (Teliaferro et al. 2006) and genetic transformation (Ghosh et al. 2002, Sajib et al. 2008). In this backdrop, the present study was undertaken to identify and confirm jute ESTs from jute genomic clones and to determine the full length sequences of a jute gene.

*Author of correspondence. <pmblab@gmail.co>. [1]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh.

Our understanding of biology has been greatly improved by studying genome structure and gene function through bioinformatics studies. However, for jute the major obstacle in conducting bioinformatics-based research is the paucity of available sequence information both at the nucleotide and at the amino acid level. The whole genome sequence of jute is yet to be determined but DNA sequences of some genomic clones are now available in the database. These sequences can be a starting point for identification of jute genes. Sequential applications of commonly available bioinformatics tools in the public domain (e.g., Blast searches, ORF finder, VecScreen, Gene identification software, multiple sequence alignment tools, protein structure analysis tool etc.) along with a combination of experimental approaches (e.g., PCR, RT-PCR, DNA sequencing etc.) are mandatory for such research. Although computational gene identification in eukaryotic genomes remains a challenging task (Zhang 2002), available gene prediction softwares like GenmarkHMM (Lomsadze et al. 2005), Genscan (Burge et al. 1998) and GlimmerHMM (Pertea and Salzberg 2002) are adequate in identifying the coding regions to a satisfactory level. To verify results obtained from these softwares, reverse transcription polymerase chain reaction (RT-PCR) is considered a gold standard.

Determination of the full length sequence of a gene from an EST is not straight forward. Bioinformatic approaches can however help resolve this difficulty.. Alignment of the EST sequences using Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) with the available orthologous sequence information stored in databases gives an idea about the candidate gene(s). Based on these sequence similaritiers degenerate primer-based gene walking technique has become a routine procedure for identifying the flanking regions of the ESTs (Parker et al. 1991).

Our previous study on computational analysis of some jute genomic clones (unpublished data) suggested that 20 clones carried putative genes. From these 20, 17 were found to be RT-PCR positive i.e., these sequences were found to be transcribed (data not shown). Because of significant homology with other plant sequences and because of the presence of a large ORF, a jute genomic clone, J 122 was selected in this study from among the 17 experimentally established ESTs to identify the full length sequence of the candidate gene. This clone matched with the ribosomal protein S8 gene of other plants and the full coding sequence was determined using degenerate primer-based PCR technique. A comparative study of this protein with chloroplast ribosomal protein S8 of *Hibiscus macrophyllus* (the top hit in Blastx search) and *Arabidopsis thaliana* (the model plant) was also made using hydrophobicity and hydropathy analyses, amino acid distribution histogram and determination of secondary structure including 3D analysis.

Materials and Methods

From computational and experimentalal approaches 17 genomic clones of *Corchorus oiltorius* Var-O4, including the clone J 122 were identified as jute ESTs (unpublished data). For the homology based analysis, clone J 122 was aligned using Blastx (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Before performing Blastx, J 122 was analyzed by VecScreen (http://www.ncbi.nlm.nih.gov/ Vec-Screen/VecScreen.html) to remove vector sequence contamination from the clone.

In *ab initio* study, the clone was analyzed by three gene prediction software: These are GENSCAN(http://genes.mit.edu/GENSCAN.html), GenMark.HMM (http://exon.gatech.edu/genemark/eukhmm.cgi) and GlimmerHMM (http:// nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/glimmerhmm/). For every analysis *Arabidopsis* sequence was selected as reference data set against which the query sequences were compared. When the clone matched to a satisfactory level with the same gene of other plant species, gene specific primers were designed based on the sequences of the clone with the aid of online available Primer3 tool (http://frodo.wi.mit.edu/).

DNA of *C. olitorius* Var-O-9897 was isolated using CTAB method (Doyle and Doyle 1990) from four-day-old seedlings. Total RNA was also isolated from four-day-old seedlings using TRIZOL reagents (Gibco BRL of Invitrogen, Life Technologies). The purity and integrity of the isolated DNA's and RNA's were checked by spectrophotometeric readings and gel electrophoresis while it's concentration was determined using a Nanodrop (ND-1000).

RT-PCR was conducted using GeNei™ one step M-MuLV RT-PCR kit. Thermal cycler profile followed for the one step RT-PCR was as follows:

Step 1: RT incubation: 50ºC for 30 minutes

Step 2: Initial denaturation at 95ºC for 5 minutes

Step 3: Denaturation at 95ºC for 30 seconds,
Annealing at 58ºC for 45 seconds,
Elongation at 72ºC for 45 seconds

Step 4: Final elongation at 72ºC for 5.0 minutes

Each time a control PCR was carried out using RNA as template to rule out contaminating DNA in isolated RNA. For sequencing, the DNA band(s) were extracted from gel using QIAGEN MinElute Gel Extraction Kit and then sequenced using services provided by Macrogen (Korea).

The sequence of the RT-PCR product(s) was translated using Transeq of EMBL-EBI database (http://www.ebi.ac.uk/Tools/emboss/transeq/index.html). Sequences similar to J122 were downloaded from the NCBI database using Blastx search and the amino acid sequences of these genes along with that of J 122 were used for multiple sequence alignment by ClustalW (http://www.ebi.ac.

uk/Tools/clustalw/). Based on this alignment, degenerate primers (listed in Table 1) were designed from the well conserved regions of amino acid sequences across the species.

Table 1. List of primers (with their sequences) used in this study. These primers were custom synthesized by 1st BASE, Singapore. All primer sequences are written in 5'-> 3' direction.

Primer name	Sequence
J 122 F	CATCGGGAAAGCAACAAATC
J 122 R	AGCCTCCCGATCTGTCATTA
J-122 DEG 1F	ATRWRTAAMNACAMWATTKCT
J-122 DEG R	AASACAMMATTKCTSRKMYRAT
J-122 GSP REV 1	GATTTGTTGCTTTCCCGATG
Chl 3 (F)	ATTTGAGGTGACACGAG
Chl 3 (R)	AGTCCCATTCTACATGTCAATAT

These degenerate primers were combined with the gene specific primers for further PCR amplification to obtain the full length sequence of the gene. Further bioinformatics analyses were performed on the full length sequence to identify the structure and function of the protein. These analyses include PSI-PRED (http://bioinf.cs.ucl.ac.uk/psipred) for secondary structure prediction, PSI-BLAST (http://www.ebi.ac.uk/Tools/psiblast/) for detection of distant evolutionary relationships etc.

The sub cellular localization of the protein was determined computationally by LOCtree (http://www.predictprotein.org/cgi/var/nair/loctree/query) and CELLO (http://cello.life.nctu.edu.tw/). 3D structure of the protein was predicted using Swiss PDB Viewer (spdbv.vital-it.ch) and other statistical analyses of this protein was performed using CLC protein workbench 5.0 (www.clcbio.com/protein).

To experimentally prove that the gene is localized in the chloroplast, as predicted by the bioinformatics analyses, PCR was performed using purified chloroplast DNA (Triboush et al. 1998) and nuclear DNA isolated from mature leaves of jute variety, O-9897. PCR amplification of chloroplast DNA and nuclear DNA was also performed using chloroplast specific primer pair Chl 3 (F) and Chl 3 (R). This primer pair was designed against IGS 2 region of chloroplast genome (Hamzeh and Dayanandan 2004).

Results and Discussion

The homology based study was carried out using Blastx, Blastn and ORF finder from NCBI. In Blastx analysis J 122 matched with plant ribosomal protein S8. This search is more sensitive than nucleotide blast since the comparison is performed at the protein level (NCBI Blast program selection guide). On the

other hand, *ab initio* gene prediction programs rely only on the statistical qualities of exons (Sempel 2000). In this study, GENSCAN, GenmarkHMM and Glimmer were used for *ab initio* gene prediction. All three software were trained with *Arabidopsis* genome sequences. So during analyses *Arabidopsis* was selected as the standard organism. These software predicted the clone J 122 to be part of a gene sequence.

To verify the results of *ab initio* gene prediction, RT-PCR was performed using GeNei™ one step M-MuLV RT-PCR kit from the total RNA of O-9897. The RNA concentration and OD ratio at 260 nm/280 nm was measured by a nanodrop (ND 1000, USA). The 260 nm/280 nm ratio was within the range of 1.87-2.06 which was an indicator of good quality RNA. Distinct bands of 18S and 28S ribosomal RNAs indicated presence of non-degraded RNA in the sample. Exact size of the expected band was observed in the gel. Using the gene specific primers (J 122 F and J 122 R) a single band of RT-PCR product of J 122 clone was found. The sequence was found to match with that of the initial clone, J 122 thus confirming that the insert in the clone was indeed a jute EST.

Next, the full length sequence of J 122 was determined using degenerate primer-based gene walking. Degenerate primers were designed to get the sequences of both 5′ and 3′ ends. Each time degenerate primers were used in combination with gene specific primers. Degenerate primer -based amplification gave bands of the expected sizes. A band of ~380 bp was found using the primer pair, J-122 DEG 1F and J-122 R and a ~250 bp band was found using the J-122 F and J-122 DEG R primer pair and ~170 bp was obtained using J-122 DEG 1F and J-122 GSP REV 1 primer pair (Fig. 1).

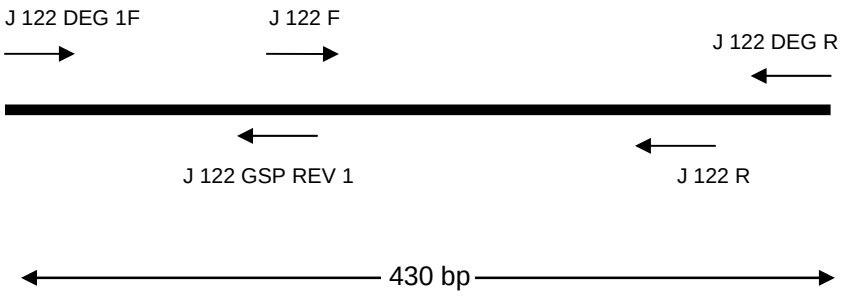

Fig. 1. Schematic diagram of full length ribosomal protein S8 gene derived by using degenerate and gene specific primers as described in results and discussion. J 122 F – J 122 R amplified a region of ~200bp, J 122 DEG 1F – J 122R amplified a region of ~380 bp, J 122F-J 122 DEG R amplified a region of ~250bp and J 122 DEG 1F- J 122 GSP REV 1 amplified a region of ~170 bp.

All the sizes were as expected for the predicted ribosomal protein S8 gene. Sequencing and assembly of the jute gene sequences revealed that this gene is

around 99% identical to the chloroplast ribosomal S8 protein of different plants and the top three hits were for *Hibiscus macrophyllus*, *Gossypium hirsutum* and *Carica papaya*. Predicted 3D structure of the jute ribosomal S8 protein was also similar to those species (Fig. 2).

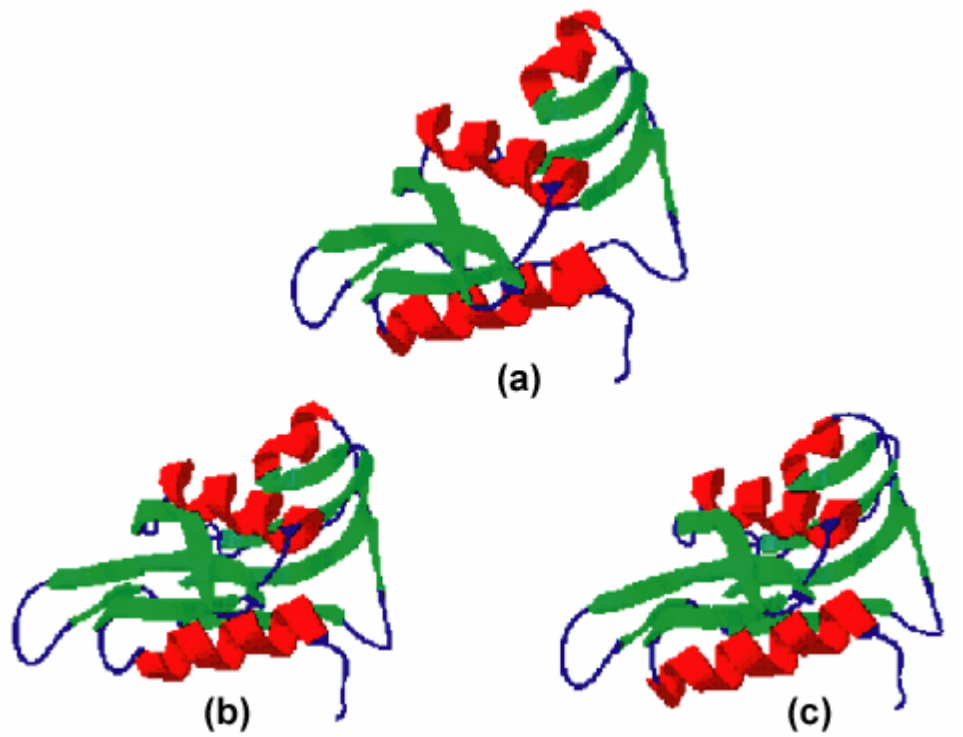

Fig. 2. 3D structure of protein predicted by Swiss PDB Viewer: (a) Jute protein, (b) *Hibiscus macrophyllus* ribosomal protein S8 and (c) *Arabidopsis thaliana* ribosomal protein S8.

As expected, this result strongly suggests that the jute ribosomal S8 protein gene is no different from that of other plant species and is well conserved across the taxa. The sequence has been submitted in the GenBank and the accession number of which is GQ325661.

The complete DNA sequence of predicted jute ribosomal protein S8 was analyzed by LOCtree and CELLO. CELLO suggested the presence of the gene both in the nucleus and the chloroplast while LOCtree analysis revealed it's presence in chloroplast only. LOCtree is a novel system of support vector machines (SVMs) that predicts the sub cellular localization of proteins, and DNA-binding propensity for nuclear proteins, by incorporating a hierarchical ontology of localization classes modeled on to biological processing pathways. Biological similarities are incorporated from the description of cellular components provided by the gene ontology consortium (GO) where GO definitions have been simplified and tailored to the problem of protein sorting.

Technically the ontology has been used using a decision tree with SVMs as the nodes (Nair and Rost 2005). CELLO is a multi-class SVM classification system which uses 4 types of sequence coding schemes: the amino acid composition, the di-peptide composition, the partitioned amino acid composition and the sequence composition based on the physico-chemical properties of amino acids. Votes from these classifiers are combined and use of the jury votes is made to determine the final assignment (Yu et al. 2006). Since LOCtree uses both SVM and GO classifiers, its prediction is better than CELLO which uses only SVM of multi class. But the computational approach alone is not satisfactory.

Both chloroplast and nuclear DNA were amplified using J122 gene specific primers (J122 F and J122 R of Fig. 1) and a band of almost 200 bp was found in gel electrophoresis for both the isolated organelles. Interestingly when both nuclear - and chloroplast DNA were amplified using chloroplast specific Chl3 primer set, around 600bp band was found for both DNAs. Since the primer is highly chloroplast specific, and the intensity of the band of amplified chloroplast DNA was higher than that of nuclear DNA (data not shown), it can be said that the nuclear DNA might be contaminated by a small amount of chloroplast DNA which might well be the reason why the ribosomal protein S8 gene which is supposed to be chloroplast encoded gave a band for nuclear DNA as well. It is well known that in mature leaves the amount of chloroplast DNA exceeds that of nuclear DNA (Jope et al. 1978), and therefore there is always a chance of nuclear DNA contamination by a small amount of chloroplast DNA.

This study has successfully achieved the target of confirming a jute EST from jute genomic clones and through this study the full length sequence of the ribosomal S8 protein gene has been deduced. Attempts have been made in determining the cellular localization of the protein using computational and experimental approaches although further studies are necessary for confirming this.

Acknowledgement

This work was financially supported by US Department of Agriculture (USDA). The authors are grateful to Bangladesh Jute Research Institute (BJRI) for supplying jute seeds.

References

Akter J, Islam MS, Sajib AA, Ashraf N, Haque S and **Khan H** (2008) Microsatellite markers for determining genetic identities and genetic diversity among jute cultivars. Aust. J. Crop Sci. **1**(3): 97-107.

Altschul SF, Gish W, Miller W, Myers EW and **Lipman DJ** (1990) Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.

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 Science. **44**: 678–685.

Burge CB and **Karlin S** (1998) Finding the genes in genomic DNA. Curr. Opin. Struct. Biol. **8**: 346-354.

Doyle JJ and **Doyle JL** Isolation of plant DNA from fresh tissue (1990) Focus. **12**(1): 13-15.

Ghosh M, Saha T, Nayak P and **Sen SK** (2002) Genetic transformation by particle bombardment of cultivated jute, *Corchorus capsularis* L. Plant Cell Rep **20**: 936–942

Gupta PK, Balyan HS, Sharma PC and **Ramesh B** (1996) Microsatellites in plants: a new class of molecular markers. Curr. Sci. **70**: 45–54.

Gupta PK and **Varshney RK** (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. **113**:163-185

Hamzeh M and **Dayanandan S** (2004) Phylogeny of populus (salicaceae) based on nucleotide sequences of chloroplast trnT-trnF region and nuclear rDNA1. Amer. J. Bot. **91**(9): 1398-1408.

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

Huda KMK, Bhuiyan MSR, Kabir MH, Jamal Uddin AFM and **Khatun A** (2007) *In vitro* plant regeneration protocol of Tossa Jute (*Corchorus olitorius*). Intl. J. Integrat. Biol. **1**(2): 96-101

Islam AS, Haque MM, Hoquc Ml and **Seraj ZI** (1992) Tissue Culture and micro propagation of jute (*Corchorus* species) in Biotechnology in Agriculture and Forestry, Vol 1 9, high Tech and Micropropagation Ill. Bajaj YPS (Ed.), Springer Verlag, Berlin, Heidelbcrg pp. 505-526.

Jope CA, Hirai A and Wildman SG (1978) Evidence that the amount of chloroplast DNA exceeds that of nuclear DNA in mature leaves. JCB. **79**: 631-636

Keka SI, Shamsuzzaman M, Pahloan MU, Pervin S, Rahman MM and **Khan H** (2008) Identifying Simple Sequence Repeat (Ssr) Marker Linked To Mite Tolerance In Jute Species. Bangladesh J. Bot. **37**(2): 161-171.

Khatun A (2007) Recent agricultural developments in Jute, Kenaf and Mesta through Traditional and Biotechnological approaches, paper presented in a seminar in Myanmar co-organized by IJSG. www. jute.org/news.htm

Lomsadze A, Ter-Hovhannisyan V, Chernoff Y and **Borodovsky M** (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Research **33**(20): 6494-6506

Nair R and Rost B (2005) Mimicking cellular sorting improves prediction of Subcellular Localization. JMB: 22; **348**(1): 85-100.

Parker JD, Rabinovitch PS and **Burmer GC** (1991) Targeted gene walking polymerase chain reaction. Nucleic Acids Res. **19**: 3055-3060

Pertea M and **Salzberg SL** (2002) Computational gene finding in plants. Plant Molecular Biology **48**(1-2): 39-48

Saha T, Ghosh M and **Sen SK** (1999) Plant regeneration from cotyledonary explants of jute *Corchorus capsularis* L. Plant Cell Rep **18**: 544–548

Sajib AA, Islam MS, Reza MS, Bhowmik A Fatema L and **Khan H** (2008) Tissue culture independent transformation for *Corchorus olitorius*. Plant Cell Tiss and Organ Cult. **95 (3)**: 333-340

Sempel C (2000) Gene prediction: the end of the beginning. Genome Biology. **1**(2): 4012.1-4012.3

Teliaferro 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.

Triboush SO, Danilenko NG and **Davydenko OG** (1998) A Method for Isolation of Chloroplast DNA and Mitochondrial DNA from Sunflower, Plant Molecular Biology Reporter **16**: 183.

Yu CS, Chen YC, Lu CH and **Hwang JK** (2006) Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics **64**: 643-651.

Zhang MQ (2002) Computational prediction of eukaryotic protein-coding genes. Nature Rev. Genet. **3**: 698-710.