Plant Omics Journal POJ

POJ 3(2):57-65(2010) ISSN:1836-3644

In silico analysis of jute SSR library and experimental verification of assembly

Rozalynne Samira[1, δ], Mahdi Muhammad Moosa[2, δ], Md. Maksudul Alam[1], Shamima Islam Keka[1], Haseena Khan[1]*

[1] Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh
[2] Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh
[δ]Contributed equally to the study.

***Corresponding author: pmblab@gmail.com; haseena@univdhaka.edu**

Abstract

Jute (*Corchorus* spp.) is the second most important fibre producing plant. Till date the study of jute at the molecular level has been very brief. Here we report the result of bioinformatics analyses of two jute Simple Sequence Repeat (SSR) libraries. The analysis reveals presence of nuclear, chloroplast and mitochondrial genes related to signalling pathway, metabolic and functional processes of jute biology. We have also verified experimentally some of the assemblies predicted by bioinformatics analysis.

Keywords: Jute, *Corchorus olitorius,* SSR library, Bioinformatics analysis, Gene prediction

Introduction

Jute (*Corchorus* spp.), classified in the Magnoliophyta division, Magnolopsida class, Malvales order and Malvaceae family, is a major fiber of Bangladesh (Hossain et al., 2003) and is one of the major fibre crops of the Indian subcontinent alongside cotton (Basuet al., 2004). Despite wide range of applications of jute and jute derived products (Ahmed and Nizam, 2008), study of jute at molecular level has been very brief. Molecular study of jute carried out till date includes genetic analysis of jute using different molecular markers e.g. RAPD, AFLP, chloroplast-SSR, SSR, STMS and ISSR (Hossain et al., 2002a,b; Qi et al., 2003 a,b; Basu et al., 2004; Roy et al., 2006; Akter et al., 2008; Mir et al., 2008) and construction of genomic and cDNA library followed by subsequent sequencing of randomly selected clones (Islam et al., 2005; Wazni et al., 2007).

The number of genes or portion of genes identified so far in jute is very limited. Islam *et al.* (2005) reported sequences of 15 jute genomic and cDNA clones which had significant similarity to *Arabidopsis* genes. Wazni and co-workers (2007) reported 16 ESTs showing significant similarity to *Arabidopsis* or other higher plant genes. Both the studies took the approach of homology based gene prediction and revealed genes which might be involved in a number of metabolic and stress related pathways. A computational and experimental approach for developing jute ESTs from genomic clones have been proposed recently (Ahmed et al., 2009).

The two widely cultivated self pollinating species of jute, *Corchorus capsularis* L. and *C. olitorius* L., (Sarker et al., 2007) have very limited genetic diversity (Basu et al., 2004) and the production, quality and yield of these economically important crops is effected by several biotic e.g. fungi, pest, insect, nematode, virus, mite (Ghosh, 1983; BBS, 2004; Keka et al., 2008) and abiotic factors e.g. salinity, submergence and low temperature (Hossain et al., 2003b; Sarker et al., 2008). To determine molecular markers which are linked to different biotic and abiotic stresses, two different jute SSR libraries were developed from *C.olitorius* variety O-4. The library constructed by Khan and her group (Akter et al., 2008) contained 157 clones, while the library prepared by Gupta's group (Mir et al., 2007) had 208 clones. All these clones were subsequently sequenced. Although the main objective of the SSR library development was potential use as a starting point for identification of markers linked to different desirable traits which could be economically important, presence of the sequence data of 365 clones from the SSR libraries inspired us to carry out bioinformatics analyses to determine the plausible functional or regulatory sequences. We searched for protein coding genes (using both *ab initio* and *homology* based prediction methods), microRNA, CpG island. Our analyses revealed 103 (double positive in *ab initio* gene prediction and double negative *ab initio* gene prediction but positive in blastn or blastx or EST blast protein coding (Table 1), one rRNA, genes. No microRNA was found. The analyses also revealed seven CpG islands.

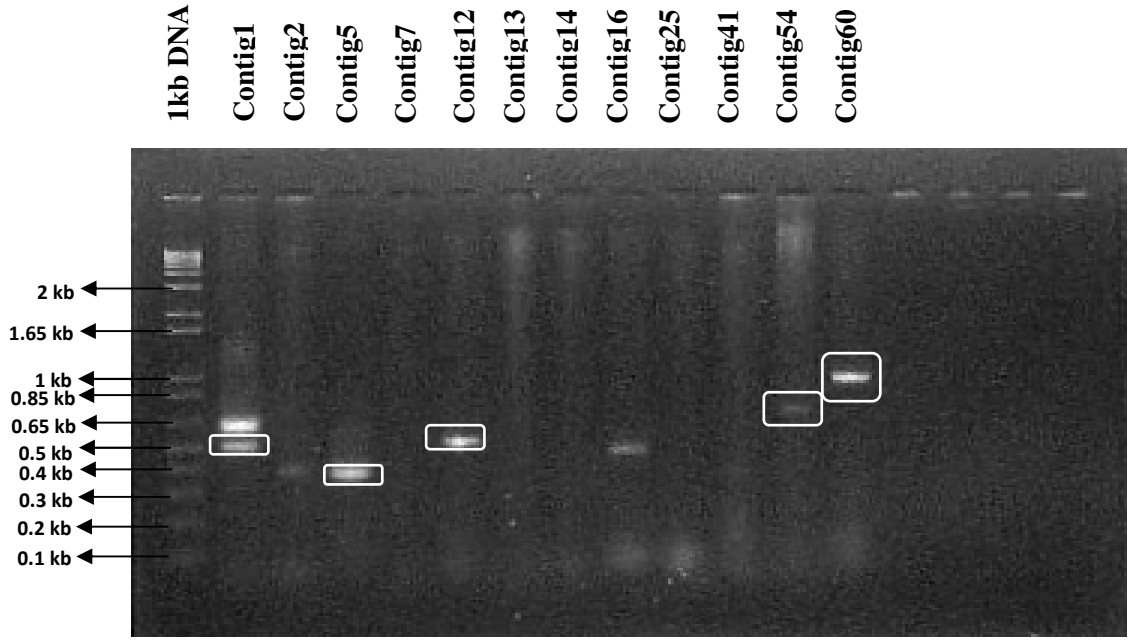

Fig 1. Gel run of Contigs 1, 2, 5, 7, 12, 13, 14, 16, 25, 41, 54, and 60. Amplified bands of Conigs 1, 5, 12, 54 and 60 showed desired sizes, confirming assembly. Boxed bands are of expected size.

Materials and Methods

Assembly of raw sequences: The sequences of two different SSR libraries were cleaned of vector sequences using VecScreen (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html), assembled based on their overlaps using CAP3 (Huang and Madan, 1999). All the assembled contigs and singletons were analyzed for the presence of protein coding genes, rRNA, tRNA, snoRNA genes, miRNAs, LTRs, CpG islands. Unless otherwise mentioned, the tools/software were used with default parameters.

Gene prediction: Both software based and homology based gene prediction methods were employed. Sequences were analyzed by GENESCAN (Burge and Karlin 1997; Burge and Karlin 1998) and geneid (Parra et al., 2000; Blanco et al., 2002) using *Arabidopsis* as reference organism for gene prediction. Homology based gene prediction was done using nucleotide blast, blastx and EST blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Blast hits were used in functional annotation of the predicted genes.

CpG island search: CpG island search was carried out using CpG finder from SoftBerry (http://linux1.softberry.com/all.htm).

LTR Search: LTR search was carried out using tool developed by Xu *et al*.(Xu and Wang, 2007) .

miRNA prediction: miRNA precursors were predicted using findMiRNA from SoftBerry. Candidate miRNA homologues from these predicted precursors were identified using microHarvester (Dezulian et al., 2006).

Assembly verification: Assembly verification was carried out by designing assembly verification primers from regions flanking the overlapping segments. The designing was carried out using Primer3Plus (Untergasser et al., 2007). Presence of desired sized bands upon PCR of genomic DNA using the desired primers confirmed the assembly. The genomic DNA was extracted from *C.olitorius* O-4 seeds by CTAB method (Murray and Thompson, 1980). 35 cycles of amplification was carried out using standard PCR procedure. PCR products were run in 1.5% agarosge gel, ethidium bromide stained and visualized under UV-transilluminator.

Table 1. Summary of assembly and gene prediction

Total number of sequences	365
Contigs	71
Singletons	118
Genescan positive	80
Geneid positive	138
Both positive	80
blastn positive	83
blastx positive	92
EST blast positive	37
Both positive but no blast match	25
Both negative but blast match	23
Only EST	7

Results

The two sequenced SSR libraries had a total of 365 sequences. It was hypothesized that some of these SSR clones were overlaps which could be assembled to get SSR contigs. The sequences of the clones were masked for vector sequences by VecScreen (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) and assembled to 71 contigs (SSR C1 to SSR C71) using CAP3 (Huang and Madan, 1999) from Mobyle portal (http://mobyle.pasteur.fr).

Table 2. Function prediction of identified genes. Double positive in gene prediction and at least two blast positive are only listed in the table. N/A stands for *No significant match was found.*

Sequence	Blastn	Blastx	Blast EST	Molecular Function	Biological Process	Organism
SSR C1	26S rRNA	Cytochrome P450/ Cytochrome P450 monooxygenase	healthy *Prunus persica* cDNA, mRNA sequence	Unspecific monooxygenase (EC 1.14.14.1)	Metabolism (Xenobiotic) (Sugiura et al., 1996)	*Arabidopsis thaliana* *Nicotina tabacum*
SSR C5	Jute SSR	Disease resistance protein	N/A	Disease resistance (Bakker et al., 2006)	Signalling (Slater et al., 2003)	*Arabidopsis thaliana*
SSR C7	Jute SSR	Amidohydrolase family enzyme	Amidohydrolase family enzyme	Carboxypeptidase (Harris et al., 2004)	Proteolysis (Harris et al., 2004)	*Robiginitalea biformata* (bacterial)
SSR C12	Jute SSR	kelch repeat-containing serine/threonine phosphoesterase family protein	AGN_PNL231df1_f8.trimmed. seq AGN_PNL *Nicotiana tabacum* cDNA 5', mRNA sequence.	Ser/Thr phosphatase activity (Harris et al., 2004)	Plant development (Swarbreck et al., 2008) Signalling (Harris et al., 2004)	*Arabidopsis thaliana*
SSR C13	Jute SSR	GrpE protein	N/A	Chaparone, nucleotide exchange factor in prokaryotes, homologue present in At mitochondria (Padidam et al., 1999)		*Flavobacterium johnsoniae* UW101
SSR C14	Jute SSR	hypothetical protein amb4270	N/A	N/A		*Magnetospirillum magneticum* AMB-1 (bacterial)
SSR C16	Jute SSR	Phosphomanno-mutase	N/A	N/A		Pedobacter sp. BAL39 (Bacterial) Presence of PMM is reported in *Arabidopsis thaliana*, *Nicotina tabacum* (Weiqiang Qian, 2007)
SSR C21	Jute SSR	CCHC-type integrase	N/A	RNA binding or single-stranded DNA binding (Clay and Nelson, 2005)	Signaling of cell cycle (Clay and Nelson, 2005)	*Populus trichocarpa*
SSR C25	*Vitis vinifera* shotgun	aspartate transaminase	N/A	Role in *Erwinia amylovora* Infection (Unpublished, from NCBI accession)	Probably member of R family genes, signalling	*Malus* x *domestica*
SSR C30	N/A	Serine/threonine protein phosphatase 2A	GH_TS_02-01-15R_D05_InvR_10Feb04_041 _F Stem - 3-5 weeks after planting Gossypium hirsutum cDNA, mRNA	Ser/Thr phosphatase activity (Harris et al., 2004)	signal transduction, cell cycle progression, and hormonal regulation in plants (Smith and Walker, 1996) Uniprot	*Oryza sativa*

SSR C31	N/A	IQD5 (IQ-domain 5); calmodulin binding	GA__Eb0015L20f *Gossypium arboreum* 7-10 dpa fiber library, *Gossypium arboreum* cDNA clone GA__Eb0015L20f, mRNA sequence	Calmodulin binding (Harris et al., 2004)	Signaling (Ca^{2+} sensor) (Zielinski, 1998; Golovkin and Reddy, 2003;)	*Arabidopsis thaliana*
SSR C41	contig VV78X15997 0.6, whole genome shotgun sequence *Vitis vinifera*	DNA-binding protein phosphatase 2C	H04_F3_032 Cotton 1-14 day post anthesis Lambda Zap Express Library *Gossypium hirsutum* cDNA, mRNA sequence	Serine/threonine specific protein phosphatase	ABA signaling, stress response (Meyer et al., 1994; Knight and Knight, 2001)	*Ricinus communis*
SSR C42	Phosphorenol pyruvate carboxylase	N/A	H_H11_H07_15 H Rosa hybrid cultivar cDNA clone H11H07, mRNA sequence	Phosphoenol pyruvate carboxylase activity	Photosynthesis (C_4 & CAM Plants) (Chollet et al., 1996; Izui et al., 2004) Tri-Carboxylic Acid cycle (Harris et al., 2004)	*Ricinus communis*
SSR C54	*Vitis vinifera* WGS contig	Translation initiation factor IF-2	N/A	Translation initiation factor activity	Protein synthesis (Lee and Young, 2000)	*Arabidopsis thaliana*
SSR C60	*Vitis vinifera* WGS conting	DEGP9 (DEGP PROTEASE 9); serine-type peptidase/ trypsin		Serine type peptidase (Harris et al., 2004)	Proteolysis (Harris et al., 2004)	*Arabidopsis thaliana*
SSR C63	N/A	retroelement pol polyprotein-like	N/A	Viral proteins		*Arabidopsis thaliana*
j-53-M13F	*Vitis vinifera* contig VV78X06912 2.9	Nucleoporin interacting component; Protein prenyltransferase	"Methyl Jasmonate-Elicited Root Cell Suspension Culture *Medicago truncatula* cDNA"	Transport	Nuclear transport (Allen et al., 2001) Protein-Protein Interaction (Protein Prenyltransferase Activity), Protein Membrane Interaction (Maurer-Stroh et al., 2003; Sinensky, 2000), Signaling (Chakrabarti et al., 1998; Chakrabarti et al., 2002)	*Medicago truncatula*
j-62-M13F		ATP synthase CF0 C subunit	"3 weeks after planting Gossypium hirsutum cDNA"	Hydrogen Ion Transporting ATP Synthase activity (Harris et al., 2004)	ATP synthesis coupled proton transport (mitochondrial) (Harris et al., 2004)	*Nicotina tabacum*
j-122-M13F		ribosomal protein S8	Cotton fiber 0-10 day post anthesis *Gossypium hirsutum* cDNA	Chloroplast ribosomal protein (structural constituent) (Shaw et al., 2007)	Translation (Harris et al., 2004)	*Hibiscus macrophyllus*

Table 3. Assembly verification primers and expected size of PCR product.

Name	Sequence	TM (°C)	Size
Contig1_RX26.5.09	5'- GCC GAA ACT CCC ACT TAT ACT A – 3'	62.7	494 bp
Contig1_FX26.5.09	5'-ATC AGG TCT CCA AGG TGA ACA-3'	60.6	
Contig2_RX26.5.09	5'-CTA GTT TCC CTG CGT CTG CTA T-3'	62.7	813 bp
Contig2_FX26.5.09	5'-TTT CCT TCA AAC CTC AGC ATT T-3'	57.1	
Contig5_RX26.5.09	5'-AGA AAC ACA ATG GAA GGC ACA T-3'	58.9	351 bp
Contig5_FX26.5.09	5'-ATG GAA GCA CTT GGA AAG CTG-3'	60.6	
Contig7_RX26.5.09	5'-GCA TTT TAT GTT GGG CAT GG-3'	58.4	686 bp
Contig7_FX26.5.09	5'-CTG ATT GGA TTC ACC AGG AAA-3'	58.7	
Contig12_RX26.5.09	5'-GGT AGC AGC CAA CCT GAT ATT T-3'	60.8	464 bp
Contig12_FX26,5.09	5'-TTG CGA TAA AAC TTA ATG GGT CT-3'	57.4	
Contig13_RX26.5.09	5'-AAT TTC GGC AGC ATC AGC AT-3'	58.4	325 bp
Contig13_FX26.5.09	5'-AGA ATT AAG TGT TGA GGA GCA ATT AG-3'	59.9	
Contig14_RX26.5.09	5'-AAC AGT GCC GTT TGA ATT GAGT-3'	58.9	784 bp
Contig14_FX26.5.09	5'-ATA TTG TCG ATG GCT GTG C-3'	58.0	
Contig16_RX26.5.09	5'-CGT GAT GTA ACG AAA GCC ATA AC-3'	61.0	261 bp
Contig16_FX26.5.09	5'-GAG ATG TAG TAT TCC GGG TAG GTG-3'	64.6	
Contig25_RX26.5.09	5'-CTA AAC TAT TTG GAG CAT TCT GTGC-3	61.3	670 bp
Contig25_FX26.5.09	5'-CTC GAA TCG GAT GAC'CCT TAT-3'	60.6	
Contig41_RX26.5.09	5'-TGG CCT CTA CCA CTT TAT CTA AC-3'	61.0	688 bp
Contig41_FX26.5.09	5'-GTG ATG GAA TAT GGG ATG TTT-3'	56.7	
Contig54_RX26.5.09	5'-TTA AAT GTT TCA CCA AAG ACAT-3'	53.4	656 bp
Contig54_FX26.5.09	5'-CTA CTA ACC CAG AAA GAG TCA TGC-3'	62.9	
Contig60_RX26.5.09	5'-GGG AGA AGA GAA ATC CTA ACATCC-3'	62.9	957 bp
Contig60_FX26.5.09	5'-GCT ACC TGA TTC TAC ATT GGT CTA CA-3'	63.0	

142 sequences (singletons) were not included in the assembly since they did not have significant overlap with other sequences. These sequences were further analyzed for functional annotation using bioinformatics tools.

Gene prediction: Both GENESCAN and geneid predicted the presence of genes in 80 of 213 sequences (contigs and singletons). Of these, 19 had at least two blast matches (any two of blastn, blastx or blastEST). These blast matches were used to assign their functions (Table 2).

CpG island: Of 213 sequences, seven sequences (SSR C1, SSR C14, j-28-M13F, j-31-M13F, j-38-M13F, j-44-M13F, j-125-M13F) had CpG islands and all of them were positive for gene prediction tools (Genescan, geneid) with at least single blast positives.

LTR Search: No LTR was detected.

miRNA search: No microRNA was detected.

Assembly verification: Based on the presence of genes within the assembled contigs, twelve of them were selected. A list of the selected contigs, designed primers and their respective size of bands expected are given in *Table 3.* Upon PCR of these 12 contigs using assembly verification primers, five (Conigs 1, 5, 12, 54 and 60) showed desired size bands, thus verifying the assemblies (Figure 1).

Discussion

Although gene density within SSR is significantly low compared to other genomic regions (Akagi et al., 1996; Cardle et al., 2000), a number of studies successfully developed SSR markers from EST libraries of different multicellular organisms (Tóth et al., 2000; Rohrer 2002; Chagné et al., 2004; Gao et al., 2004; Slate et al., 2007;). Studies on human and other mammalian genome revealed that signalling and communication genes, transcription factors and protein kinases are rich in microsattelites (Hancock and Simon 2005; Sharma et al., 2005). Although the size of the sequences analyzed were statistically too small to draw specific conclusion, presence of eight signalling associated proteins among fourteen predicted nuclear proteins were suggestive of a similar pattern in jute (Table 2). Yet the analysis of completely sequenced dicot genome (*Arabidopsis*) failed to reveal statistically significant enrichment of genes to specific GO category (Lawson and Zhang, 2006). Thus it is possible that the observed high frequency of signalling proteins might be a random occurrence.

Of the nineteen predicted genes, five showed similarity to bacterial genes. Endosymbiotic origin theory of cell organalles predicts the presence of genes similar to prokaryotes in both mitochondria (Bonen and Calixte, 2006) and chloroplast (Gould et al., 2008). In fact, two of the five proteins showing similarity to

bacterial proteins gave significant matches with dicot mitochondrial genes (SSR C13 and j-62-M13F). However, it cannot be ruled out that these sequences might be contaminations arising from symbiotic or other forms of associated bacteria which might have been present in the jute seed coat (source of DNA for the SSR library construction).

One interesting finding of this study is that jute SSR sequences are more likely to give positive hits with *Vitis* whole genome shotgun sequence compared to *Arabidopsis* in blastn. Phylogenetic analysis of different rosids reveals that jute (Malvales) is more related to *Arabidopsis* (Brassicales) than to *Vitis* (Vitales) (Group, 2003; Judd and Olmstead 2004; Wang et al., 2009). Genomes of both *Arabidopsis thaliana* (Initiative, 2000) and *V.vinifera* (Jaillon et al., 2007) have been sequenced and are publicly available. Thus based on phylogenetic relationship it is expected that in blastn, *Arabidopsis* sequence should appear more commonly than *Vitis* sequence. One possible explanation can be the significantly lower rate of change in the Malvales clade than in its Brassicales counterpart. If the rate of change is also lower in Vitales clade, it is possible that the observed similarity is due to both being more similar to the latest common ancestor of the rosid group. Although, to our knowledge, there has been no extensive genome wide analysis of lineage specific mutation rate heterogeneity till date, theoretical analysis supports the possibility of different optimal mutation rate across different taxa (Kondrashov 1995a,b; Baer et al., 2007). It is to be noted that mutation rate heterogeneity across taxa in plant mitochondrial genome (Palmer et al., 2000; Parkinson et al., 2005; Sloan et al., 2008), chloroplast genes (Gaut et al., 1992; Clegg et al., 1994) and selected plant nuclear genes (Clegg et al., 1997) has been reported. Soltis *et al.*'s (2002) integration of fossil data with molecular data also suggests rate heterogeneity among higher plants. An alternate hypothesis is that the large grapevine genome (487 Mb) (Jaillon et al., 2007) gave more random matches than Arabidopsis (125 Mb) (Initiative, 2000). But this cannot explain higher frequency of grapevine hits in predicted protein coding regions (Table 2). Larger eukaryote genomes contain greater amount of repetitive DNA sequences (Flavell et al., 1974; SanMiguel et al., 1996; Hake and Walbot, 1980). Ramsay *et al.* (1999) found intimate association between certain regions of retrotransposon and repetitive elements. Since both jute (1250 Mb) (Mir et al., 2008) and *Vitis* have relatively large genome size compared to *Arabidopsis* genome, it is possible that the observed similarity was due to repetitive regions of the two genomes. Again, this cannot explain single retroelement associated protein among the nineteen predicted proteins (Table 2).

GENSCAN and geneid predicted presence of gene in each of the seven CpG island containing sequences. SSR C1 and C14 had at least two different blast type matches (Table 2). Other five CpG island positive sequences also had single blast type match (data not shown). It has been known that CpG islands play vital role in regulation of gene expression in different plants (Gutiérrez-Marcos et al., 2006; Feil and Berger, 2007). Thus, it's possible that these CpG islands play regulatory roles in expression of these genes.

Of the twelve selected contigs, five had bands of expected sizes after amplification. It is possible that the contigs which did not give desired sized bands actually do not exist, i.e. the reads assembled into contigs were not actually part of that contig. Alternatively it is also possible that the primer pairs failed to amplify desired segment at the reaction condition.

The size of the study was very limited to draw any definite conclusion. Yet the study gave us an insight of jute at the molecular level. Further study of jute genome is required to unravel the molecular story of this commercially important fibre producing plant.

Acknowledgements

This study was financially supported by US Department of Agriculture (USDA). MMM and RS were supported by NSICT Fellowships from Ministry of Science and Information & Communication Technology, Government of Bangladesh. The authors are grateful to Bangladesh Jute Research Institute (BJRI) for supplying jute seeds.

Notes

Sequences of Contigs and the detailed bioinformatics analysis of individual contigs, singletons are available upon request.

References

Ahmed S, Nabi MZ, Alam MM, Islam MS, Samira R, Moosa MM, Khan H (2009) A computational and experimental approach for developing jute ESTs from genomic clones. Aust J Crop Sci 3:322-328

Ahmed Z, Nizam SA (2008) Jute-microbiological and biochemical research. Plant Tissue Cult Biotechnol 18:197-220

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93:1071-1077

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

Allen NPC, Huang L, Burlingame A, Rexach M (2001) Proteomic Analysis of Nucleoporin Interacting Proteins. J Biol Chem 276:29268-29274

Baer C, Miyamoto M, Denver D (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet 8:619-631

Bakker E, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in Arabidopsis. Plant Cell 18:1803-1818

Basu A, Ghosh M, Meyer R, Powell W, Basak SL, Sen SK (2004) Analysis of genetic diversity in

cultivated jute determined by means of SSR markers and AFLP profiling. Crop Sci 44:678-685

BBS (2004) Statistical pocket book of Bangladesh statistics division, Bangladesh Bureau of Statistics, Ministry of Planning, the People's Republic of Bangladesh.

Blanco E, Parra G, Guigó R (eds) (2002) Using geneid to identify genes. John Wiley & Sons Inc., New York

Bonen L, Calixte S (2006) Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. Mol Biol Evol 23:701-712

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78-94

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

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847-854

Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera M, Vendramin G, Garcia V, Frigerio J, Echt C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109:1204-1214

Chakrabarti D, Azam T, DelVecchio C, Qiu L, Park Y, Allen C (1998) Protein prenyl transferase activities of *Plasmodium falciparum*. Mol Biochem Parasitol 94:175-184

Chakrabarti D, Da Silva T, Barger J, Paquette S, Patel H, Patterson S, Allen CM (2002) Protein Farnesyltransferase and Protein Prenylation in Plasmodium falciparum. J Biol Chem 277:42066-42073

Chollet R, Vidal J, O'Leary M (1996) Phospho enol pyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. Annu Rev Plant Biol 47:273-298

Clay N, Nelson T (2005) The recessive epigenetic swellmap mutation affects the expression of two step II splicing factors required for the transcription of the cell proliferation gene STRUWWELPETER and for the timing of cell cycle arrest in the Arabidopsis leaf. Plant Cell 17:1994-2008

Clegg M, Cummings M, Durbin M (1997) The evolution of plant nuclear genes. Proc Natl Acad Sci U S A 94:7791-7798

Clegg M, Gaut B, Learn Jr G, Morton B (1994) Rates and patterns of chloroplast DNA evolution. Proc Natl Acad Sci U S A 91:6795-6801

Dezulian T, Remmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. Bioinformatics 22:359-360

Feil R, Berger F (2007) Convergent evolution of genomic imprinting in plants and mammals. Trends Genet 23:192-199

Flavell R, Bennett M, Smith J, Smith D (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257-269

Gao L, Jing R, Huo N, Li Y, Li X, Zhou R, Chang X, Tang J, Ma Z, Jia J (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392-1400

Gaut B, Muse S, Clark W, Clegg M (1992) Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. J Mol Evol 35:292-303

Ghosh T (1983) Hand book on jute. Food and Agriculture Organization of the United Nations (FAO) 115-116

Golovkin M, Reddy A (2003) A calmodulin-binding protein from Arabidopsis has an essential role in pollen germination. Proc Natl Acad Sci U S A 100:10558-10563

Gould S, Waller R, McFadden G (2008) Plastid evolution. Annu Rev Plant Biol 59:491-517

Group TAP (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot J Linn Soc 141:399-436

Gutiérrez-Marcos J, Costa L, Dal Prà M, Scholten S, Kranz E, Perez P, Dickinson H (2006) Epigenetic asymmetry of imprinted genes in plant gametes. Nat Genet 38:876-878

Hake S, Walbot V (1980) The genome of Zea mays, its organization and homology to related grasses. Chromosoma 79:251-270

Hancock J, Simon M (2005) Simple sequence repeats in proteins and their significance for network evolution. Gene 345:113-118

Harris M, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258

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

Hossain MB, Awal A, Rahman MA, Haque S, Khan H (2003b) Distinction between cold-sensitive and -tolerant Jute by DNA polymorphism. J Biochem Mol Biol 36:427-432

Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Res 9:868-877

Initiative TAG (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796-815

Islam AS, Taliaferro M, Lee CT, Ingram C, Montalvo RJ, Ende Gvd, Alam S, Siddiqui J, Sathasivan K (2005) Preliminary progress in Jute (*Corchorus* species) genome analysis. Plant Tissue Cult Biotechnol 15:145-156

Izui K, Matsumura H, Furumoto T, Kai Y (2004) Phosphoenolpyruvate carboxylase: a new era of structural biology. Annu Rev Plant Biol 55:69

Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467

Judd W, Olmstead R (2004) A survey of tricolpate (eudicot) phylogenetic relationships. Am J Bot 91:1627-1644

Keka SI, Shamsuzzaman M, Pahloan MU, Pervin S, Rahman MM, Khan H (2008) Identifying simple sequence repeat (SSR) marker linked to mite tolerance in jute species. Bangladesh J Bot 37:161-171

Knight H, Knight M (2001) Abiotic stress signalling pathways: specificity and cross-talk. Trends in Plant Sci 6:262-267

Kondrashov A (1995a) Dynamics of unconditionally deleterious mutations: Gaussian approximation and soft selection. Genet Res 65:113-121

Kondrashov A (1995b) Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. Genet Res 66:53-69

Lawson M, Zhang L (2006) Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. Genome Biology 7:R14

Lee T, Young R (2000) Transcription of eukaryotic protein coding genes. Annual Reviews of Genetics 34:77-137

Maurer-Stroh S, Washietl S, Eisenhaber F (2003) Protein prenyltransferases. Genome Biology 4:212

Meyer K, Leube M, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in Arabidopsis thaliana. Science 264:1452-1455

Mir RR, Rustagi S, Sharma S, Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Sinha MK, Balyan HS, Gupta PK (2008) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. Euphytica 161:413-427

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4326

Padidam M, Reddy V, Beachy R, Fauquet C (1999) Molecular characterization of a plant mitochondrial chaperone GrpE. Plant Mol Biol 39:871-881

Palmer J, Adams K, Cho Y, Parkinson C, Qiu Y, Song K (2000) Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A 97:6960-6966

Parkinson C, Mower J, Qiu Y-L, Shirk A, Song K, Young N, dePamphilis C, Palmer J (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol 5:73

Parra G, Blanco E, Guigó R (2000) GeneID in Drosophila. Genome Res10:511-515

Qi J, Zhou D, Wu W, Lin L, Wu J (2003a) The application of RAPD technology in genetic diversity detection of Jute. Yi Chuan Xue Bao 30:926-932

Qi J, Zhou D, Wu W, Lin L, Wu J, Fang P (2003b) Applucation of ISSR technology in genetic diversity detection of jute. Ying Yong Sheng Tai Xue Bao 14:1473-1477

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant Cell 17:415-425

Rohrer GA, Fahrenkrug SC, Nonneman D, Tao N, Warren WC (2002) Mapping microsatellite markers identified in porcine EST sequences. Anim Genet 33:372-376

Roy A, Bandyopadhyay A, Mahaptra AK, Ghosh SK, Singh NK, Bansal KC, Koundal KR, Mohapatra T (2006) Evaluation of genetic diversity in jute (*Corchorus* species) using STMS, ISSR and RAPD markers. Plant Breed 125:292-297

SanMiguel P, Tikhonov A, Jin Y, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer P, Edwards K, Lee M, Avramova Z (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-768

Sarker RH, Al-Amin GM, Hassan F, Hoque MI (2008) *Agrobacterium*-mediated genetic transformation of two varieties of jute (*Corchorus capsularis* L.). Plant Tissue Cult Biotechnol 18:7-16

Sarker RH, Al-Amin GM, Hoque M (2007) In vitro regeneration in three varieties of white jute (*Corchorus capsularis* L.). Plant Tissue Cult Biotechnol 17:11-18

Sharma V, Brahmachari S, Ramachandran S (2005) (TG/CA)n repeats in human gene families: abundance and selective patterns of distribution according to function and gene length. BMC Genomics 6:83

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275-288

Sinensky M (2000) Functional aspects of polyisoprenoid protein substituents: roles in protein-protein interaction and trafficking. Biochimica et Biophysica Acta 1529:203 - 209

Slate J, Hale M, Birkhead T (2007) Simple sequence repeats in zebra finch (Taeniopygia guttata) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. BMC Genomics 8:52

Slater A, Scott NW, Fowler MR (2003) Plant disease resistance. In: Plant Biotechnology. Oxford University Press Inc., New York, pp 157-178

Sloan DB, Barr CM, Olson MS, Keller SR, Taylor DR (2008) Evolutionary Rate Variation at Multiple Levels of Biological Organization in Plant Mitochondrial DNA. Mol Biol Evol 25:243-246

Smith R, Walker J (1996) Plant protein phosphatases. Annu Rev Plant Biol 47:101-125

Soltis P, Soltis D, Savolainen V, Crane P, Barraclough T (2002) Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci U S A 99:4430-4435

Sugiura M, Sakaki T, Yabusaki Y, Ohkawa H (1996) Cloning and expression in Escherichia coli and

Saccharomyces cerevisiae of a novel tobacco cytochrome P-450-like cDNA. Biochim Biophys Acta 1308:231-240

Swarbreck D, Wilks C, Lamesch P, Berardini T, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009-1014

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res10:967-981

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71-74

Wang H, Moore M, Soltis P, Bell C, Brockington S, Alexandre R, Davis C, Latvis M, Manchester S, Soltis D (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences of the USA 106:3853-3558

Wazni MW, Islam AS, Taliaferro JM, Anwar N, Sathasivan K (2007) Novel ESTs from a Jute (*Corchorus olitorius* L.) cDNA library. Plant Tissue Cult Biotechnol 17:173-182

Weiqiang Qian CY, Huanju Qin, Xin Liu, Aimin Zhang, Ida Elisabeth Johansen, Daowen Wang, (2007) Molecular and functional analysis of phosphomannomutase (PMM) from higher plants and genetic evidence for the involvement of PMM in ascorbic acid biosynthesis in Arabidopsis and *Nicotiana benthamiana*. Plant Cell 49:399-413

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res W265-W286

Zielinski R (1998) Calmodulin and calmodulin-binding proteins in plants. Annu Rev Plant Biol 49:697-725