Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 10, pp.21063-21073, October, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

IDENTIFICATION OF FUNCTIONAL RELEVANCE OF SSRS DEVELOPED FROM ESTS THROUGH DATA MINING IN CORCHORUS SP.

***Udayan Bhattacharyya, Disharee Nath and Tapash Dasgupta**

[1]Department of Biotechnology, Bengal Institute of Technology, West Bengal University of Technology, Kolkata West Bengal, India
[2]Department of Genetics and Plant Breeding, University of Calcutta, Kolkata, West Bengal, India

ARTICLE INFO

Article History:
Received 19th July, 2015
Received in revised form
26th August, 2015
Accepted 21st September, 2015
Published online 20th October, 2015

Key words:

Expressed Sequence Tags (ESTs),
Functional Domain Marker (FDM),
Simple Sequence Repeats (SSRs)

ABSTRACT

Jute (Corchorus spp.) is the second most important fibre producing plant in India. The study on jute at the molecular level till date is very limited. Simple sequence repeats (SSRs) are useful tools as molecular markers in genetic studies. Recently with rapid advancement of sequencing and documentation in public data base has enabled scientists in hassle free data mining. Development of high throughput method for detection of SSRs has given a new dimension in their use as molecular markers. In the present study, 826 EST sequences of Corchorus capsularis and 30 EST sequences of Corchorus olitorius, downloaded from National Center for Biotechnology Information (NCBI) and were analyzed through bioinformatic tools like UniVec, trimest, CAP3, and MISA. Several microsatellite markers were identified for these two species of jute which would be further helpful for genetic mapping. The study of functional domain markers (FDM) can provide information of functional property of microsatellite markers and predicted protein domains. The functional domains designed from SSR-FDMs will help to analyze molecular markers that have functional importance and should also facilitate the analysis of genetic diversity in plants.

***Citation*:** Udayan Bhattacharyya, Disharee Nath and Tapash Dasgupta, 2015. "Identification of functional relevance of SSRs developed from ESTs through data mining in Corchorus sp.", *International Journal of Current Research*, 7, (10), 21063-21073.

INTRODUCTION

Jute (*Corchorus* sp) is one of the major fibre crops of the Indian subcontinent apart from cotton (Basu *et al*., 2004). Despite wide range of applications of jute and its derived products (Ahmed and Nizam, 2008), very little investigation has been carried out at molecular level. So far molecular study of jute includes genetic analysis through molecular markers e.g. RAPD, AFLP, SSR, chloroplast- SSR, STMS and ISSR (Hossain *et al*., 2002a, 2003; Qi *et al*., 2003a, 2003b; 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). Genetic, proteomic and biotechnological study on *Corchorus* sp are in progress. Molecular marker plays an important role in modern advanced breeding technology. Among different classes of molecular markers, microsatellite markers or simple sequence repeats (SSRs) are unequivocally the most favored **(**Joshi *et al*., 2011), albeit SNP markers are now on cross road due to availability of high throughput next generation sequencing technology.

****Corresponding author: Udayan Bhattacharyya***
Department of Biotechnology, Bengal Institute of Technology, West Bengal University of Technology, Kolkata West Bengal, India.

SSRs are multi-allelic, reproducible, co-dominant marker, abundant and cover extensive genome (Gupta and Varshney, 2000). Though study on microsatellites in different cereals, oilseeds, medicinal and aromatic plants of commercial importance **(**Tripathi *et al*., 2009) has been started far beyond but in *Corchorus* sp microsatellite markers have been recently developed and deployed for the study of genetic polymorphism at within and between species or genus (Shokeen *et al*., 2005). However developing genomic SSRs are highly arduous, cumbersome and expensive.

Presently abundant availability of biological data extrapolating from whole genome sequence initiatives, concomitant with the use of bioinformatics tools help to determine useful sequences commensurate with generation of efficient markers. Myriad expressed sequence tags (ESTs) available in public database facilitated EST-SSR marker development through data mining and processing usually 200-500nucleotides single pass sequences generating from mRNA or cDNA are designated as genes expressed in a specific tissue (Adams *et al*., 1991). The SSR markers identified from ESTs are also transferable across different species as they are very conserved regions of genome (Joshi *et al*., 2011).

The priority of functional genomic research turned out large dataset of ESTs and by utilizing the data through modern bioinformatics tools, it is now feasible to figure out and develop EST-SSR markers abundantly within a limited span of time and cost-effective manner (Kantety *et al*., 2002**;** Varshney *et al*., 2002). The recent data of GenBank (March 07, 2014) reveals that 30 and 826 EST sequences are available for *Corchorus olitorius* and *Corchorus capsularis* respectively (http://www.ncbi.nlm.nih.gov/dbEST/dbESTsummary.html).

EST based SSRs had already been worked out in several plants species including grape (Scott *et al*., 2000**)**, sugarcane (Cordeiro *et al*., 2001), durum wheat **(**Eujayl *et al*., 2002), rye (Hackauf and Wehling, 2002) and medicinal plant like basil (Gupta *et al*., 2010). Keeping in mind the upshot of findings in other crops and possible advantages of EST-SSR primers, the present study aims at data mining the updated EST libraries of two species of jute namely *Corchorus olitorius* and *Corchorus capsularis* and subsequent searching out SSRs. Simple sequence repeat functional domain marker (SSR-FDM) relies on development of molecular markers for putative functional domains. The FDM represents the SSR-based molecular markers and predicted protein domain (Yu *et al*., 2010). The annotation helps to know the putative functions of the ESTs and to find the important functional domain markers (FDM) related to the SSR-ESTs leading to gene ontology study. The gene ontology in general, covers three domains namely (a) biological process: operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units like cells, tissues, organs, and organisms, (b) cellular component: the parts of a cell or its extracellular environment and (c) molecular function: the elemental activities of a gene product at the molecular level, such as binding or catalysis.

MATERIALS AND METHODS

Data Source

All the EST sequences of *Corchorus capsularis* (826 sequences) and *Corchorus olitorius* (30 sequences) were retrieved from dbEST database of National Center for Biotechnology Information (NCBI). The retrieved sequences were related to different plant tissues e.g. leaves, stem, root, etc. The downloaded sequences were obtained in FASTA format for sequence assembly and SSR analysis.

Processing of EST sequence

EST sequences were searched for vector sequence and removal of the vector sequence was done using Cross_match software**.** The vector sequences were obtained from the UniVec database**.** Then the Trimmest program was used for removing the poly-A and poly-T ends of the EST sequences**.** It allows to trim the poly-A and poly-T ends from the given sequence according to the parameters given.

EST sequence assembly

EST sequences after trimming were assembled using the CAP3 (Contig Assembly Program).

The basic CAP3 tool with default parameters was used for assembly of EST sequences. CAP3 generated Contig files, Singlets files, Qual files, Info files and Out files. For the SSR identification, the contig and singlet sequences were combined to form non-redundant sequence data set.

Detection of SSR motifs

To detect SSRs in the EST sequence data set, the software used MISA (MIcroSAtellite identification tool) was used**.** MISA (misa.pl) is a freely downloadable perl script. Along with this misa.ini was also downloaded that contained search parameters.

Primer designing

Primer sequence designing for SSR-EST sequences was performed through PRIMER3 software (http://frodo.wi.mit.edu/primer3/). The conditions for primer designing were set at default values.

Functional annotation

The Functional Domain Markers were found from the SSR containing sequence using InterProScan of EBI (Quevillon *et al*., 2005)**.** InterProScan provided the platform to analyze the functional domains with the help of the member databases such as BlastProDom, FPrintScan, HAMAP, HMMPIR, HMMPfam, HMMSmart, HMMTigr, HMMPanther, Pattern Scan, ProfileScan, Super Family, Signal PHMM, TMHMM and Gene3D. The SSR ESTs were searched for significant matches using a special type of BLAST program named BLASTx at NCBI against non-redundant protein database entries**.** BLASTx searched protein database using a translated nucleotide query. BLASTx was carried out maintaining identity value parameter > 70%. The SSR-FDM contig sequences found from InterProScan were annotated for Biological process, Cellular component and Molecular function using QuickGO (http://www.ebi.ac.uk/QuickGO) at the EBI server (Binns *et al*., 2009)**.** QuickGO is provided by UniProtKB-GOA group that is used as fast web-based browser for Gene Ontology terms and annotations.

RESULTS

Corchorus capsularis L.

The EST sequences especially non-redundant ones are useful resources for SSR motifs in part of the genome responsible for transcription (Varshney *et al*., 2005). ESTs often stand for redundant cDNA sequences which necessitate the use of CAP3 program for turning out the sequences without overlapping or contigs. Program MISA was used for detection of SSRs (Table 1a). 826 redundant EST sequences obtained from NCBI stand for 274±10.7 $MbpC^{-1}$ of *Corchorus capsularis* genome. SSRs identified from this dataset summed up to 81 and thus are arranged to 1.00 SSR per 3535.39 bases. The pre-processing process involved removal of 20692bp empty vectors, low-quality sequences and Poly A/T tails. Trimming of poly A and poly T tails produced exclusion of original dataset. The remaining sequences were subsumed into a non-redundant dataset of 559 unique gene sequences (122 contigs and 437 singlets).

Table 1a.Results of microsatellite search using MISA in *Corchorus capsularis*

Total number of sequences examined:	559
Total size of examined sequences (bp):	286367
Total number of identified SSRs:	81
Number of SSR containing sequences:	66
Number of sequences containing more than 1 SSR:	11
Number of SSRs present in compound formation:	8

Table 1b. Results of microsatellite search using MISA in *Corchorus olitorius*

Total number of sequences examined:	28
Total size of examined sequences (bp):	7113
Total number of identified SSRs:	2
Number of SSR containing sequences:	2
Number of sequences containing more than 1 SSR:	0
Number of SSRs present in compound formation:	0

Table 2. FDM Analysis

FDM Name	No.
RNA recognition motif domain (nucleic acid binding)	2
Polyprenyl synthetase (isoprenoid biosynthetic process)	1
Photosystem II PsbP, oxygen evolving complex (calcium ion binding)	1
RNA polymerase III transcriptional repressor, MAF1 (negative regulation of transcription from RNA polymerase III promoter)	1
S-adenosylmethionine synthetase (methionine adenosyltransferase activity)	1
Glycoside hydrolase, family 1 (carbohydrate metabolic process)	3
Plant peroxidase (response to oxidative stress)	2
Protein family Cys-rich	1
Mitochondrial brown fat uncoupling protein (mitochondrial transport)	1
Lipoxygenase (metal ion binding)	1
Bet v I domain (Biological Process)	1
ClpP (serine-type endopeptidase activity)	1
Pathogenic type III effector avirulence factor Avr cleavage site	1
Phosphoribosyltransferase domain (nucleoside metabolic process)	1
DNA-binding WRKY	2
Hydroxymethylglutaryl-CoA reductase, class I/II (hydroxymethylglutaryl-CoA reductase (NADPH) activity)	1
Major intrinsic protein (transporter activity)	1
Ribosomal protein L7A/L8 (ribosome biogenesis)	2

Table 3. Percentage distribution of various BLASTx results (*Corchorus capsularis)*

BLASTx Results	No. of results
Significant matches	25 (38.46%)
Low sequence similarity	28 (43.07%)
Unknown/Hypothetical/Predicted Proteins/Gene ontology absent	8 (12.31%)
No Hits	4 (6.15%)

Skimming of class I microsatellites in this gene sequence disclosed 66 unique SSR sequences. The results signified the frequency of 1 SSR per 3.5 kb of *Corchorus capsularis* genome. The dwindling in redundancy of sequence sorted out from non-redundant dataset is presented in Figure 1a. Two microsatellites adjacent to each other were found in 17 cases and two adjoining repeats within 10 bp were observed in 57 ESTs. Presence of adjacent SSRs from non-redundant ESTs was described by Cardle *et al.* (2000) in various crop plants like rice (3.4), soybean (7.4), tomato (11.1), *Arabidopsis* (13.8), poplar (14.0) and cotton (20.0) through data mining and FDM analysis. Following the analytical approach similar to the present study one SSR per 14.6 kb of *Catharanthus* ESTs was reported by Joshi *et al.* (2011) and in sesame it is one SSR per 6625 bases (Bhattacharyya *et al*., 2014). The study revealed that the occurrence of SSR from cDNA is less in Jute compared to rice, soybean, tomato and Arabidopsis. Several factors like origin of crops to evolutionary factors, selection or even artificial selection may play a role in such variation. Trimming of poly A/T tract did not restrict existence of mononucleotide which clearly proved their presence in the genome rather than fag end of mRNA. The dinucleotide CT was more frequent (40%) followed by AT, GA and TA with same abundance (20%) among dimeric SSRs. However, the dinucleotide AT was conspicuously absent in the present investigation.

Table 4a. SSR Primers of *Corchorus capsularis* developed by PRIMER3

Sl. No.	Oligo	Start	Length	tm	GC%	Primer Sequence	Product size	Motif	SSR	SSR Length
1	F	146	21	55.05	42.86	CTCTTCTGTTCTGCTCTTCAA	154	TCT	TCTTCTTCTTCTTCTTCTTCTTCTTCT	30
	R	299	21	55.08	42.86	CATCTTCTTCTTCATCACTGC				
2	F	495	21	55.93	42.86	CTCAACAGTTCAATGGCTATG	159	CCT	CCTCCTCCTCCT	12
	R	653	21	54.88	42.86	GGTTACCAACATACACACGAT				
3	F	11	21	55.16	33.33	TCCCATACACAAAAACAAAAC	150	ACC	ACCACCACCACCACCACCACC	21
	R	160	21	55.08	42.86	GCTGTGAAGATGAAGAAGATG				
4	F	11	21	55.16	33.33	TCCCATACACAAAAACAAAAC	150	CCG	CCGCCGCCGCCG	12
	R	160	21	55.08	42.86	GCTGTGAAGATGAAGAAGATG				
5	F	213	21	55.20	42.86	TATCCTCAGATCCCCAACTAT	143	CAG	CAGCAGCAGCAGCAGCAGCAGCAG	24
	R	355	20	55.50	45.00	ATGGATCATCATCCTGTGAC				
6	F	393	18	54.54	61.11	CTTCTCCACCTCCTCCTC	150	ATC	ATCATCATCATCATC	15
	R	542	21	55.26	38.10	ATGATGATGATGGTTCTGGTA				
7	F	422	18	56.28	61.11	CACCTACCCCGTCCATAG	150	CAT	CATCATCATCAT	12
	R	571	20	55.53	45.00	ACAACAGCCTCAACTTCTTG				
8	F	18	22	54.82	31.82	GCACACTTGAAACAAGAAATTA	138	CAA	CAACAACAACAACAACAA	18
	R	155	21	54.88	38.10	CAAGAAGTTTCTCAGCCATTA				
9	F	25	21	54.57	33.33	ATTTTTGACCAATCTCTCCTT	151	CT	CTCTCTCTCTCT	12
	R	175	20	54.98	35.00	TAACAATGGGCTCTGTTTTT				
10	F	25	21	55.31	33.33	CATCATCAAACAACATCAACA	160	CCAT	CCATCCATCCATCCATCCAT	20
	R	184	21	54.41	33.33	AAATTCCCCTGTTTTCTTTAG				
11	F	29	21	55.25	33.33	ATCAAACAACATCAACACCAT	156	CATA	CATACATACATACATACATA	20
	R	184	21	54.41	33.33	AAATTCCCCTGTTTTCTTTAG				
12	F	22	21	54.41	42.86	TACTACTTGGTTCCCGATATG	131	GAT	GATGATGATGATGAT	15
	R	152	21	55.06	38.10	TCTTCATCTTCTTCACCTTCA				
13	F	106	21	54.61	33.33	ATGGCTTTGTATAAGGGATTT	165	GAT	GATGATGATGATGAT	15
	R	270	22	52.43	31.82	TTACGTAACTTGAAGCATAGAA				
14	F	355	21	54.62	47.62	CATCCCTCTCTTATTCCTCTC	182	GGC	GGCGGCGGCGGCGGCGGCGGC	24
	R	536	19	54.84	42.11	ATTTTCCACCAACCACTTC				

Continue……

15	F	200	21	55.21	33.33	TTGCAGAAGGATCAATTAGAA	149	TA	TATATATATATATA	14
	R	348	22	55.17	31.82	TTGCCCTTATTATACAATTTCC				
16	F	265	20	54.47	35.00	CGCGATTTGTAATTAATGTG	220	GAT	GATGATGATGAT	12
	R	484	20	55.41	45.00	GAAGATGCATAATCCACCAC				
17	F	266	21	55.36	33.33	GCGATTTGTAATTAATGTGGA	219	GTG	GTGGTGGTGGTGGTG	15
	R	484	20	55.41	45.00	GAAGATGCATAATCCACCAC				
18	F	266	21	55.36	33.33	GCGATTTGTAATTAATGTGGA	219	GGT	GGTGGTGGTGGT	12
	R	484	20	55.41	45.00	GAAGATGCATAATCCACCAC				
19	F	277	21	54.19	33.33	TTAATGTGGATGATGATGATG	239	GTG	GTGGTGGTGGTG	12
	R	515	21	55.10	33.33	CAAAGCCTAACATAGCAAAAA				
20	F	10	21	55.03	38.10	TTTTAGGGTTTGAGTGACAGA	152	GA	GAGAGAGAGAGA	12
	R	161	20	55.42	40.00	TACAAGACCCAAAACCACAT				
21	F	7	20	55.52	45.00	AGTCCGTTCTGGACTCTTTT	172	CTT	CTTCTTCTTCTTCTT	15
	R	178	22	55.03	31.82	GTTTTTGATTGACGGAATCTAT				
22	F	240	21	54.9	42.86	CCGTGATGAGAAATCTATGAG	147	GCG	GCGGCGGCGGCGGCGGCG	18
	R	386	18	57.07	55.56	CTGTAGCCACCACCGTTT				
23	F	3	21	55.54	33.33	AATTCAACTTCAACCGTTTCT	151	TCC	TCCTCCTCCTCC	12
	R	153	21	56.16	42.86	GATGGTGATGATGATGATGAG				
24	F	107	20	55.21	50.00	GGTTTCTTACCATCCTCCTC	162	TCA	TCATCATCATCATCA	15
	R	268	21	54.76	38.10	TCCCCTGTTAAAGATGACATA				
25	F	274	21	54.93	42.86	GATGTTAGTCCCAATGGTGTA	177	CAT	CATCATCATCATCATCAT	18
	R	450	21	55.27	38.10	CTTGGTTGGTTCTTTCTTTCT				
26	F	306	21	55.3	42.86	GCTAAGGAGACATCATCATCA	145	CTTT	CTTTCTTTCTTT	12
	R	450	21	55.27	38.10	CTTGGTTGGTTCTTTCTTTCT				
27	F	322	21	54.42	38.10	CATCATCATCATCTCGTTGTA	146	AAAG	AAAGAAAGAAAG	12
	R	467	21	54.49	38.10	ATTCTCAGATTACATGGCTTG				

F- Forward primer R- Reverse primer

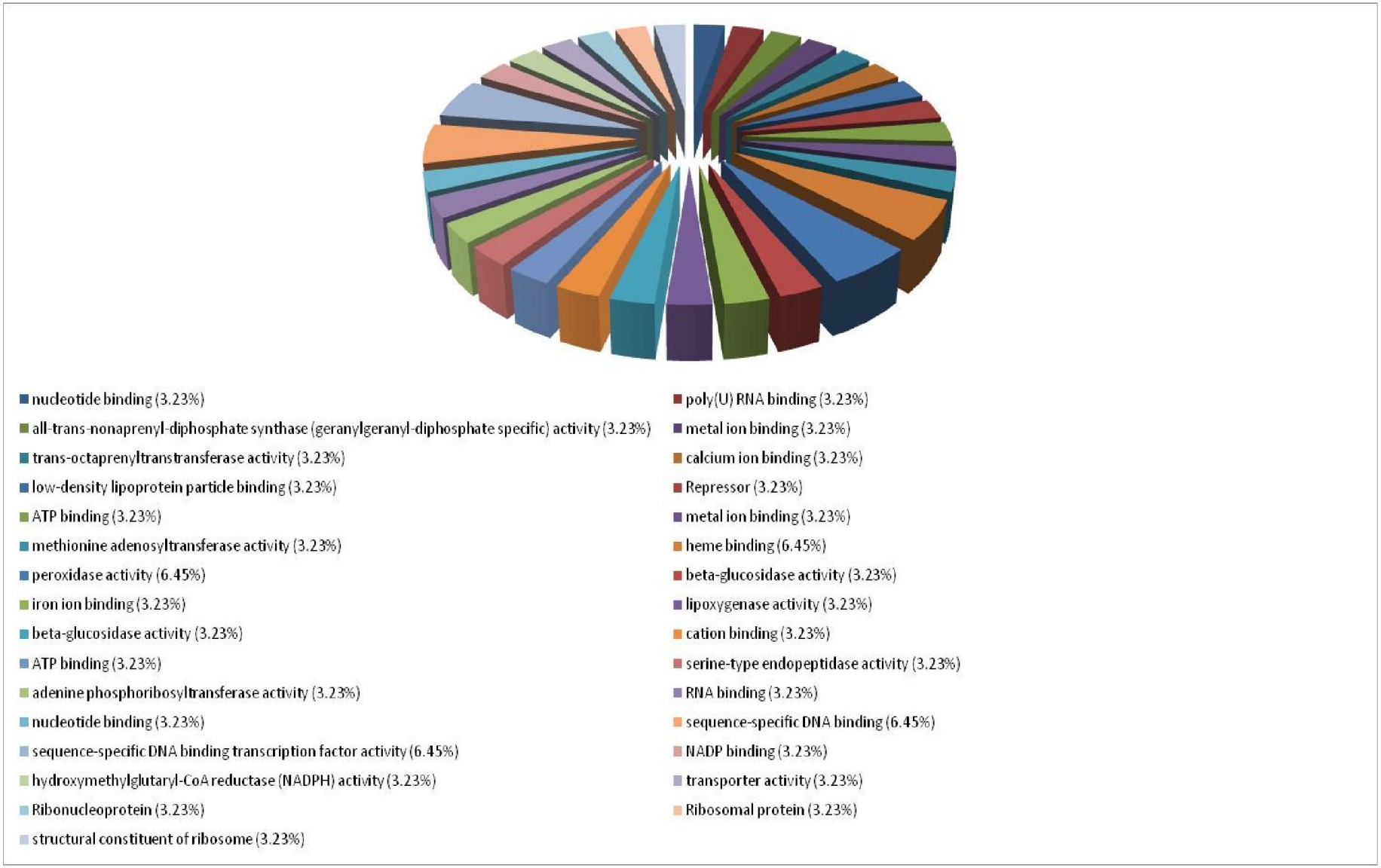

Figure 2a. Gene Ontology (Molecular Function)

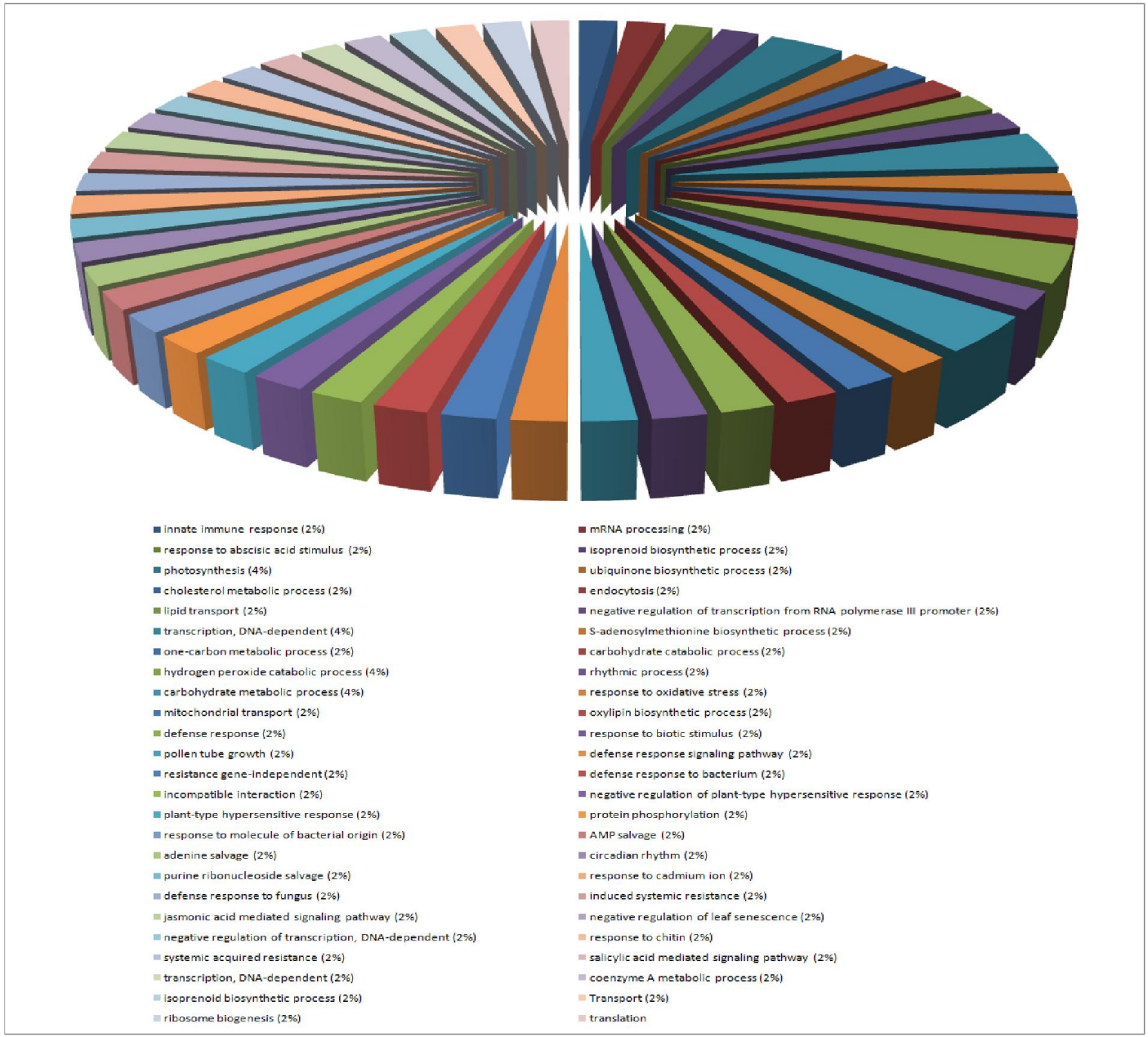

Figure 2b. Gene Ontology (Biological process)

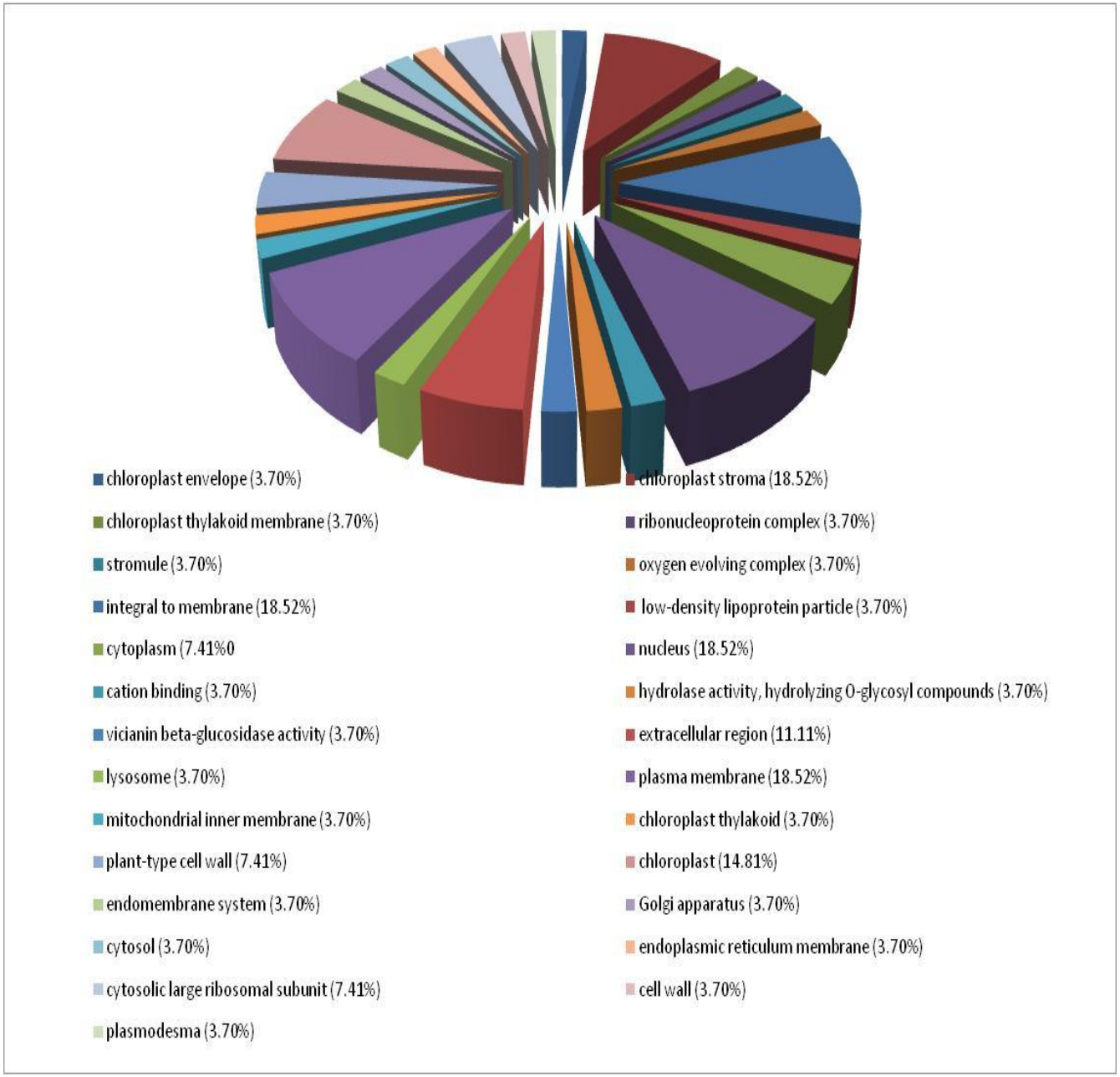

Figure 2c: Gene Ontology (Cellular component)

Table 4b. Primers of *Corchorus olitorious* developed by PRIMER3

Sl. No.	Oligo	Start	Length	tm	GC%	Primer Sequence	Product size	Motif	SSR	SSR Length
1	F	45	21	54	42.86	GATCATGACTTTGGTCTGTGT	152	TA	TATATATATATA	12
	R	196	21	55	38.1	GCACTGAATGTTTCATAAGGA				

F- Forward primer R- Reverse primer

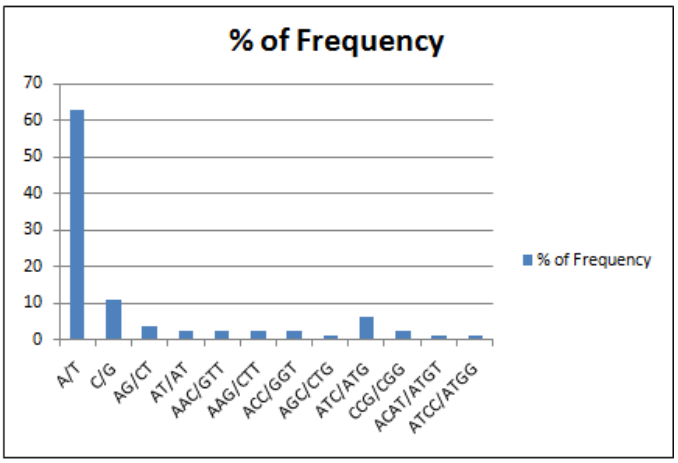

Figure 1a. Distribution of EST-SSRs based on the motifs in *Corchorus capsularis*

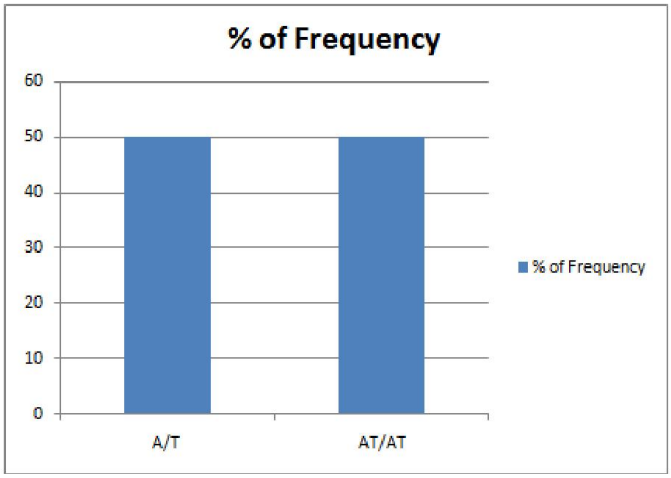

Figure 1b. Distribution of EST-SSRs based on the motifs in *Corchorus olitorious*

Similar observation was recorded in rice (Temnykh *et al*., 2000), Arabidopsis (Cardle *et al.,* 2000) and maize (Chin *et al*., 1996). Moreover, the repeats GC was absent like that of cereals crops (Senthilvel *et al*., 2008). The CT/AG motif stands for GAG, AGA, UCU and CUG in mRNA. Among trinucleotide repeats, GAT was more frequent, while other trimeric nucleotides like ACC, ATC, CAA, CAG, CAT, CTT, GCG, GTG, TCA, TCT and TTG occurred almost equally. However, Gao *et al.* (2003) reported that the microsatellite AAG was more abundant in plants in general, and CCG in monocot genome. Tetrameric microsatellite motifs like ACAT/ATGT and ATCC/ATGG were also observed in present findings. Among all SSR motifs figured out from ESTs, A/T (62.96%) was more abundant followed by C/G (11.11%), ATC/ATG (6.17%). The remaining microsatellite repeats from ESTs represented less than 5%.

Analysis of SSR-FDM

Identified 81 SSR containing sequences were analyzed for Functional Domain Markers (FDM). Through InterProScan, 200 functional domains were examined from InterPro member databases such as pattern scan, SignalPHMM, TMHMM, HMMPanther, and FPrintScan. The functional domains were responsible for RNA recognition motif, Polyprenyl synthetase, Photosystem II PsbP, oxygen evolving complex, and also function as RNA polymerase III transcriptional repressor, MAF1, S-adenosylmethionine synthetase, Glycoside hydrolase, family 1, Plant peroxidase, protein family Cys-rich, Mitochondrial brown fat uncoupling protein, Lipoxygenase, Bet v I domain, ClpP, Pathogenic type III effector avirulence factor Avr cleavage site, Phosphoribosyltransferase domain, DNA-binding WRKY, Hydroxymethylglutaryl-CoA reductase, class I/II hydroxymethylglutaryl-CoA reductase, Major intrinsic protein and Ribosomal protein L7A/L8. All together 25 SSR-FDMs were identified those provide information regarding transcribed genetic markers having putative functions (Table 2).

Analysis of BLASTx results

BLASTx was performed on SSR-ESTs to search proteins with significant match to translated SSR-EST nucleotide sequence against non-redundant GenBank database. Out of 81 unique SSR-ESTs, 25 had significant match to proteins (Table 3).

Functional annotation of significant matches

Functional annotation of the SSR-ESTs was performed using Swiss-Prot database in terms of gene ontology. The result was tabulated with the gene ontology term and corresponding number of SSR-ESTs. Also, specific primers were designed for such sequences (Table 4a and Ttable 4b). These primers amplify a gene of interest which produces a known protein product. The number of SSR-ESTs that produced no hit was 4 (6.15%). This indicates presence of sequences encoding proteins which are specific to *Corchorus capsularis* or proteins which are present in other plant/animal systems but are still not reported (Table 3).

Biological process

A biological process is a series of events accomplished by one or more ordered assemblies of molecular functions. In a range of biological processes corresponding to SSR-ESTs, the most frequent was ‘Photosynthesis’ (4 SSR-ESTs), ‘hydrogen peroxide catabolic process’ (4 SSR-ESTs), and ‘carbohydrate metabolic process’ (4 SSR-ESTs) (Figure 2a).

Molecular function

Molecular function describes catalytic or binding activities that occur at the molecular level. In an extent of molecular

functions, the most frequent were 'heme binding' (2 SSR-ESTs), peroxidase activity (2 SSR-ESTs) and sequence-specific DNA binding (2 SSR-ESTs) (Figure 2b).

Cellular component

The cellular component of a cell is part of some larger object that may be an anatomical structure or a gene product group. In a gamut of cellular components housing putative proteins, the most frequent were 'chloroplast stroma' (5 SSR-ESTs), 'nucleus' (5 SSR-ESTs) and 'Plasma membrane' (15 SSR-ESTs) (Figure 2c).

***Corchorus olitorius* L.**

A meager total of 30 redundant EST sequences were available in NCBI database representing about 324 Mb of *Corchorus olitorius* genome. Similar to C*orchorus capsularis*, steps like screening of vectors, removal of poly A/T tails were also followed in *Corchorus olitorius* during pre-processing. 28 unique sequences were obtained after redundant EST sequences analysis and total sequences added upto 7113 basepairs which were further being analyzed for detection of class I hypervariable. The program MISA was used for detection of SSRs and only two hypervariable SSR loci was detected having 1 and 2 nucleotide repeat motifs (Table 1b). This clearly emphasizes lacking of transcript data of the plant *Corchorus olitorius*. Among the two detected SSR motifs, one was simple type and another was compound type. It was also identified that the mononucleotide A/T and dinucleotide AT/CT occurred at equal frequency among the SSR repeat motifs (Figure 1b).

EST-SSR markers are useful resources for studying genetic diversity or QTL mapping, but sufficient transcript data are still not available for *Corchorus olitorius*.

DISCUSSION

Microsatellite markers work for multifarious roles in plant genomics. EST databases impart a valuable resource for development of microsatellite markers which are genic in nature. Engendering SSR markers, simply by determining analysis through bioinformatic tools put away both cost and time, provided sufficient EST sequences are available. In *Corchorus capsularis* 81 hypervariable microsatellites were identified from EST database of NCBI following a set of bioinformatics tools like Cross_Match, TrimEST, CAP3 and MISA. Contrary to *capsularis* sp. of jute, in *olitorius* sp. insufficient data were available and only two microsatellites were detected from EST sequences. Since one of the features of EST-SSR is transferability among species and so identified EST-SSRs would provide effective result in both species of jute. The frequency analysis described the study could be used to design informative SSR primers that can be applied in studies of divergence analysis, linkage map construction, comparative genomics and locating of genes on chromosomes. Out of 81 SSR-ESTs, successful functional annotation of 65 SSR-ESTs was performed using Gene Ontology terms. These 65 SSR-ESTs were subjected to primer designing which yielded a total of 54 primer sets for *Corchorus capsularis.* The sequences having both SSRs and FDMs signified that functional domains provided predicted functions to the molecular markers.

REFERENCES

Adams, M.D., Kelle, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B. and Moreno, R.F. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651-1656.

Ahmed, Z. and Nizam, S.A. 2008. Jute - Microbiological and Biochemical Research. *Plant Tiss. Cult. Biotech.,* 18: 197-220.

Akter, J., Islam, M.S., Saji, A.A., Ashraf, N., Haque, S., Khan, H. 2008. Microsatellite markers for determining genetic identities and genetic diversity among jute cultivars. *Aust. J. Crop Sci.,* 1: 97-107.

Basu, A., Ghosh, M., Meyer, R., Powell, W., Basak, S.L. and Sen, S.K. 2004. Analysis of genetic diversity in cultivated jute determined by means of SSR markers and AFLP profiling. *Crop Sci.,* 44: 678-685.

Bhattacharyya, U., Pandey, S.K. and Dasgupta, T. 2014. Identification of EST-SSRs and FDM in sesame (Sesamum indicum L.) through data mining. *Scholarly J. Agr. Sci.,* 4: 60-69.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C. and Apweiler, R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics,* 25: 3045-3046.

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

Chin, E.C.L., Shu, H., Smith, J.S.C. and Senior, M.L. 1996. Maize simple repetitive DNA sequences: abundance and allele variation. *Genome,* 39: 866-873.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M. and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.,* 160: 1115-1123.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P. and Powell, W. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor. Appl. Genet.,* 104: 399-407.

Gao, L., Tang, J., Li, H. and Jia, J. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breeding,* 12: 245-261.

Gupta, P.K. and Varshney, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica,* 113: 163-185.

Gupta, S., Shukla, R., Roy, S., Sen, N. and Sharma, A. 2010. In silico SSR and FDM analysis through EST sequences in Ocimum basilicum. *Plant Omics. J.,* 3: 121-128.

Hackauf, B. and Wehling, P. 2002. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breeding,* 121: 17-25.

Hossain, M.B., Haque, S. and Khan, H. 2002a. DNA fingerprinting of jute germplasm by RAPD. *J. Biochem. Mol. Biol.,* 35: 414-419.

Hossain, M.B., Awal, A., Rahman, M.A., Haque, S. and Khan, H. 2003. Distinction between cold sensitive and cold tolerant jute by DNA polymorphisms. *J. Biochem. Mol. Biol.,* 36: 427—432.

Islam, A.S., Taliaferro, M., Lee, C.T., Ingram, C., Montalvo, R.J., Ende, G., Alam, S., Siddiqui, J. and Sathasivan, K. 2005. Preliminary progress in Jute (*Corchorus* species) genome analysis. *Plant Tiss. Cult. Biotech.,* 15: 145-156.

Joshi, R.K., Kar, B. and Nayak, S. 2011. Exploiting EST databases for the mining and characterization of short sequence repeat (SSR) markers in *Catharanthus roseus* L. *Bioinformation,* 5: 378-381.

Kantety, R.V., La Rota, M., Matthews, D.E. and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.,* 48: 501-510.

Mir, R.R., Rustgi, S., Sharma, S., Singh, R., Goyal, A., Kumar, J., Gaur, A., Tyagi, A.K., Khan, H., Sinha, M.K., Balyan, H.S. and Gupta, P.K. 2008. A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. Euphytica., 161: 413-427.

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

Qi, J.M., Zhou, D.X., Wu, W.R., Lin, L.H., Fang, P.P., Wu, J.M. 2003. The application of RAPD technology in genetic diversity detection of Jute. Yi Chuan Xue Bao 30: 926–932.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.,* 33: W116-W120.

Roy, A., Bandyopadhyay, A., Mahapatra, A.K., Ghosh, S.K., Singh, N.K., Bansal, K.C., Koundal, K.R. and Mohapatra, T. 2006. Evaluation of genetic diversity in jute (*Corchorus* species) using STMS, ISSR and RAPD markers. *Plant Breeding,* 125: 292-297.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S. and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.,* 100: 723-726.

Senthilvel, S., Jayashree, B., Mahalakshmi, V., Kumar, P.S., Nakka, S., Nepolean, T. and Hash, C. 2008. Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. *BMC Plant Biol.,* 8: 119-128.

Shokeen, B., Sethy, N.K., Choudhary, S. and Bhatia, S. 2005. Development of STMS markers from the medicinal plant Madagascar periwinkle [*Catharanthus roseus* (L.) G. Don.]. *Mol. Ecol. Notes,* 5: 818-820.

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T., McCouch, S.R. 2000. Mapping and genome organization of microsatellite sequences in rice (Oryza sativa L.). *Theor. Appl. Genet.,* 100: 697-712.

Tripathi, K.P., Roy, S., Maheshwari, N., Khan, F., Meena, A. and Sharma, A. 2009. SSR polymorphism in Artemisia annua: Recognition of hotspots for dynamics mutation. *Plant Omics J.,* 2: 228-237.

Varshney, R.K., Graner, A. and Sorrells, M.E. 2005. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.,* 23: 48-55.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P. and Graner, A. 2002. In Silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol. Biol. Lett.,* 7: 537-546.

Wazni, M.W., Islam, A.S., Taliaferro, J.M., Anwar, N. and Sathasivan, K. 2007. Novel ESTs from a Jute (*Corchorus olitorius* L.) cDNA library. *Plant Tiss. Cult. Biotech.,* 17: 173-182.

Yu, J.K., Paik, H., Choi, J.P., Han, J.H., Choe, J.K. and Hur, C.G. 2010. Functional domain marker (FDM): an in silico demonstration in Solanaceae using simple sequence repeats (SSRs). *Plant Mol. Biol., Rep* 28: 352-356.

****