Available online at www.sciencedirect.com

ScienceDirect

Development of a set of SSR markers for genetic polymorphism detection and interspecific hybrid jute breeding☆

Dipnarayan Saha, Rajeev Singh Rana, Sukla Chakraborty, Subhojit Datta, Arroju Anil Kumar, Asim Kumar Chakraborty, Pran Gobinda Karmakar*

Division of Crop Improvement, Indian Council of Agricultural Research (ICAR)-Central Research Institute for Jute and Allied Fibers, Barrackpore, Kolkata 700121, India

ARTICLE INFO

Article history:
Received 7 December 2016
Received in revised form
27 February 2017
Accepted 16 March 2017
Available online 29 April 2017

Keywords:
Database
Genic-SSR
Jute
Interspecific hybrid identification
Unigene sequences

ABSTRACT

Corchorus capsularis (white jute) and *C. olitorius* (dark jute) are the two principal cultivated species of jute that produce natural bast fiber of commercial importance. We have identified 4509 simple sequence repeat (SSR) loci from 34,163 unigene sequences of *C. capsularis* to develop a non-redundant set of 2079 flanking primer pairs. Among the SSRs, trinucleotide repeats were most frequent (60%) followed by dinucleotide repeats (37.6%). Annotation of the SSR-containing unigenes revealed their putative functions in various biological and molecular processes, including responses to biotic and abiotic signals. Eighteen expressed gene-derived SSR (eSSR) markers were successfully mapped to the existing single-nucleotide polymorphism (SNP) linkage map of jute, providing additional anchor points. Amplification of 72% of the 74 randomly selected primer pairs was successful in a panel of 24 jute accessions, comprising five and twelve accessions of *C. capsularis* and *C. olitorius*, respectively, and seven wild jute species. Forty-three primer pairs produced an average of 2.7 alleles and 58.1% polymorphism in a panel of 24 jute accessions. The mean PIC value was 0.34 but some markers showed PIC values higher than 0.5, suggesting that these markers can efficiently measure genetic diversity and serve for mapping of quantitative trait loci (QTLs) in jute. A primer polymorphism survey with parents of a wide-hybridized population between a cultivated jute and its wild relative revealed their efficacy for interspecific hybrid identification. For ready accessibility of jute eSSR primers, we compiled all information in a user-friendly web database, JuteMarkerdb (http://jutemarkerdb.icar.gov.in/) for the first time in jute. This eSSR resource in jute is expected to be of use in characterization of germplasm, interspecific hybrid and variety identification, and marker-assisted breeding of superior-quality jute.

1. Introduction

The *Corchorus* genus of the newly classified *Sparrmanniaceae* family (earlier *Tiliaceae*) comprises more than 100 species. However, cultivated jute belongs to just two species, *C. capsularis* L. (white jute) and *C. olitorius* L. (dark jute) [1]. These two plant species are exploited for natural bast or stem fiber and are second only to cotton in volume of global fiber production [2].

☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding author.*
E-mail address: dipnarayan.saha@icar.gov.in (D. Saha).

http://dx.doi.org/10.1016/j.cj.2017.02.006

The lignocellulosic bast fiber of jute is an important source of biodegradable commercial products including ropes, twines, hessian sacks, and carpet backing cloth. Its potential applications in diversified jute products, such as in geotextiles and the automobile and agricultural industries, are gaining popularity [3,4]. Among jute-growing countries, India (1.94 million t) and Bangladesh (1.39 million t) contribute 98% of total world jute production (FAOSTAT 2014, http://www.fao.org/faostat/en/#data/QC).

A growing preference for ecofriendly natural fibers over synthetics has resulted in an intensification of scientific jute cultivation and processing. Thus, there is an immense need to accelerate development of genetic and genomic resources to raise the production of superior-quality jute fiber. To date, only limited information is available in basic jute biology, genetics, and germplasm resources [5,6]. Molecular markers, such as genomic simple sequence repeats (SSRs) [7], sequence-related amplified polymorphism (SRAP), random amplified polymorphic DNA (RAPD), and inter simple sequence repeat (ISSR) [8], restriction-site-associated DNA (RAD) derived single nucleotide polymorphism (SNP) markers [9], and expressed sequence tag (EST)-SNP markers [10] have been developed in jute to construct genetic linkage maps and characterize genetic diversity [11]. Despite this upsurge in molecular markers, identification of simple to assay functional markers in jute for genetic characterization of genotypes is lagging behind. The recent publication of bast tissue-derived transcriptome resources [12,13] has promised the discovery of functional molecular markers, such as expressed SSRs or genic SSRs. Consolidating the previously developed SSR markers [7] and the EST-SSR markers from a Chinese cultivar of *C. capsularis* recently reported by Zhang et al. [14], the total number of SSR markers reported in jute is about 3600. Only a small fraction of these markers have been validated in jute, resulting in the availability of a few hundred markers for use in breeding or mapping. Even under the assumption that the SSRs are uniformly distributed, they are at least nine megabases (Mb) apart in the jute genome of 350–400 Mb distributed over seven chromosomes, a density that is insufficient for the construction of a high-density linkage map for QTL identification and breeding applications. A compact linkage map usually requires a marker interval of <2 cM (2.12 Mb in jute) necessitating development of several thousand additional markers to map the complete jute genome to a reasonable mapping density. Here we report the development of an additional set of unigene-derived SSR markers from the *C. capsularis* cv. JRC-212, which are mostly nonredundant with respect to the previously reported set. We also performed cross-species validation of a subset of eSSR markers sampled in different *Corchorus* species and in parents of a mapping population. Finally, we developed a functional marker database for ready access to this resource and for use in jute genetic improvement.

2. Materials and methods

2.1. Plant materials and genomic DNA extraction

Twenty-four *Corchorus* accessions were chosen for validation of eSSR primer pairs (Table 1). Of the 24 accessions, five were from *C. capsularis*, twelve accessions represented *C. olitorius*, and seven accessions were from wild jute species. In addition, two parents of an interspecific hybrid population, consisting of a cultivated *C. olitorius* accession (OIJ-248) and a wild species *C. aestuans* (WCIN-136) were used for a parental polymorphism survey. Seeds were collected from the gene bank of the ICAR-Central Research Institute for Jute and Allied Fiber. Plants were grown in pots for genomic DNA extraction from fresh young leaves following Kundu et al. [15] with minor modifications. After incubation at 60 °C for 1 h, the sample lysate was treated with RNase A (20 μg mL^{-1}) and incubated at 37 °C for 15 min. The lysate was extracted using chloroform: isoamyl alcohol (24:1, v/v) prior to DNA precipitation. The integrity of genomic DNA was checked on 1.0% agarose gel and its quantity measured by OD_{280}/OD_{260} nm in a UV spectrophotometer (Biophotometer, Eppendorf AG, Germany) before use in PCR reactions.

2.2. SSR mining from *C. capsularis* unigenes

Bast fiber unigene sequences of white jute, *C. capsularis*, cv. JRC-212 were retrieved from the NCBI Transcriptome Shotgun Assembly (TSA) database (GenBank accessions GBSD01000001 to GBSD01034163) [12]. To identify SSR loci in the 34,163 unigenes, the MIcroSAtellite identification (MISA) tool (http://pgrc.ipk-gatersleben.de/misa/) [16] was used. For dinucleotide repeats, a stretch of nucleotide sequences with less than six repeats was excluded. Similarly, for tri-, tetra-, penta-, and hexanucleotides, a minimum of five repeats was adopted. Mononucleotide repeats were ignored. For perfect SSR identification, the cutoff value of 100 bp was chosen as the maximum length of interruptions between two SSRs. Unigene sequences containing perfect SSRs were further used to design primers from SSR locus flanking sequence with SSRLocator v1.0 [17]. Only di-, tri-, and tetranucleotide SSR loci were used for primer design. The virtual PCR module of SSRLocator was used to eliminate any redundant primers. Positions of SSRs in unigene sequences, such as in coding sequence (cds), untranslated regions (UTRs), or introns or outside genes were predicted ab initio using Augustus 2.5.5 [18] and *Arabidopsis* genes as training model.

2.3. Annotation of unigenes containing SSRs and anchoring to RAD linkage map

Functional annotation of the jute unigene sequences containing SSRs was performed using the program Blast2GO Basic v3.1.3 [19]. The unigenes were searched against the NCBI's nonredundant (NR) protein database using BLASTX. The E-value cutoff was set at 1e−50. The "Plant slim" option of the GO-Slim tool of Blast2GO was used for gene annotations, such as by biological process, molecular function, and cellular component. In order to map the eSSR markers on the existing Restriction-site Associated DNA (RAD) linkage map of jute [9], the SSR-containing unigenes were searched with BLASTN against the 503 RAD marker sequences using an E-value cutoff of 1e−10. The top-hit unigene sequences searched with eSSR markers were manually checked for the position of the SSR repeat, followed by estimation of their locations

Table 1 – Details of plant genotypes of *Chorchorus* species used in the study.

Sl. no.	Species	Genotypes/ varieties (popular name)	National accession numbers	Origin/pedigree	More about genotype
Panel for genetic diversity analysis					
1	*Corchorus capsularis*	JRC-212 (Sabuj sona)	IC0430404	Selection from indigenous genotype (unknown)	Developed at ICAR-CRIJAF, India, year of release 1954, popular variety, premature flowering tolerant variety when sown mid-March
2	*C. capsularis*	JRC-412 (CIM 010)	–	India	Improved white jute cultivar, Susceptible to *Macrophomina phaseolina* stem rot disease of jute
3	*C. capsularis*	CIM 036	–	India	Improved white jute germplasm line, Resistant genotype to *M. phaseolina* stem rot disease of jute
4	*C. capsularis*	JRC-321	–	Selection from indigenous germplasm "Hewti"	Developed during 1954 from ICAR-CRIJAF
5	*C. capsularis*	Bidhan pat 3	–	Selection from D 154 × D 18	Developed at Bidhan Chanda Krishi Viswavidyalaya, West Bengal, India during 2001, Photoperiod insensitive line
6	*C. olitorius*	JRO-2407 (Samapti)	IC0587431	KEN/SM/024 × JRO 524	Stem leaf petiole, stipules and pod red colored, good quality TD3 grade fibers, resistant to pests and diseases and premature flowering when sown in early March
7	*C. olitorius*	Bidhan rupali	–	Mutant of JRO-632	Developed at Bidhan Chandra Krishi Viswavidyalaya, generated through mutation breeding, pale green leaf and stem color
8	*C. olitorius*	JR0–8432 (Shakti, CO-32)	IC0430388	IC15901 × Tanganyika-1	Developed at ICAR-CRIJAF, India, year of release 1999, resistant to flowering on mid-March sowing
9	*C. olitorius*	JRO-204 (Suren)	–	IDN/SU/053 × KEN/DS/ 060	Developed at ICAR-CRIJAF, India, year of release 2007, resistant to flowering, tall plant, lodging tolerant, high fiber yield
10	*C. olitorius*	AAU-OJ-1 (Tarun)	IC0550168	Tanganyika-1 × JRO-640	Developed at Assam Agricultural University during 2007, premature flowering tolerant, tolerant to pests and pathogen, good quality fiber with TD2 grade
11	*C. olitorius*	JRO-524 (Navin)	IC0430385	Sudan green × JRO-632	Developed at ICAR-CRIJAF, India, year of release 1977
12	*C. olitorius*	Tanganyika-1	–	Tanzania	Direct introduction from Tanzania, Africa, Premature flowering tolerant, parental source to many released varieties and cultivars
13	*C. olitorius*	S-19 (Subala)	IC0426944	(JRO-620 × Sudan green) × Tanganyika-1	Developed at ICAR-CRIJAF, India, year of release 2005, premature flowering tolerant on mid-March sowing, finer fiber quality with low lignin, fiber yield high
14	*C. olitorius*	Sudan green	–	Sudan	Land race introduction from Sudan, Africa, and parental source for many varieties
15	*C. olitorius*	OIN-10	IC0503321	India	Wild olitorius, dwarf type
16	*C. olitorius*	OIJ-231	IC0503547	Tanzania	Landrace introduced from Tanzania, Africa
17	*C. olitorius*	OMU-02, DR-2	–	India	Mutant of *C. olitorius*
18	*C. fascicularis*	WCIJ-28	–	Kenya	Wild species of jute, resistant to stemrot disease
19	*C. aestuans*	WCIN0009	IC0610385	Orissa, India	Wild jute, resistant to Bihar hairy caterpillar
20	*C. pseudo olitorius*	WCIJ-034	–	Kenya	Wild species of jute, resistant to fungal stem and root rot disease
21	*C. urticaefolius*	WCIJ-006	–	Kenya	Wild species of jute, resistant to fungal stem and root rot disease but soft rot
22	*C. tridens*	WCIN-012	–	Rajasthan, India	Wild species of jute, fine fiber
23	*C. pseudo-capsularis*	WCIJ007	EC396204	Kenya	Wild species of jute, finest fiber, resistant to fungal stem and root rot disease but anthracnose
24	*C. trilocularis*	WCIN-083	IC333575	Tamil Nadu, India	wild species of jute, water-logging tolerant, partly compatible to interspecific crosses
Panel for parental polymorphism survey					
25	*C. olitorius*	OIJ-248	–	Nepal	Cultivated jute highly susceptible to *M. phaseolina*
26	*C. aestuans*	WCIN-136	–	West Bengal, India	Wild species of jute, resistant to *M. phaseolina* and insect pests

–, Information not available.

on the map by conversion of basepair to cM length using the genome-wide corrected cM/Mb recombination rate of 1.058 [9]. Finally, the eSSR markers anchored to adjacent RAD markers on the linkage map were represented using the Graphical GenoTypes 2 (GGT2) software [20].

2.4. ORF and functional domain predictions of unigenes containing SSRs

OrfPredictor (http://proteomics.ysu.edu/tools/OrfPredictor.html) [21] was used to predict ORFs in SSR-containing

unigenes. Both strands were considered in the prediction and an E-value cutoff of 1e−30 was used in BLASTX-based prediction. ORFs encoding more than 100 amino acids were searched for functional protein domains and motifs. Default parameters of the InterProScan software (Iprscan 5.9–50.0) of the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI) suite [22] were used for functional domain identification.

2.5. Development of eSSR database for *Corchorus sp.*

To compile eSSR primers into a database, the Jute Marker database (JuteMarkerdb) was developed using Microsoft SQL Server® Enterprise edition, 2012 for database management. Microsoft-based ASP·NET v4.5 and C sharp (C#) were used as a programming language. The database is provided through a web-based client–server application, using an object-oriented model having a hierarchical organization of classes, subclasses, and solution classes. Dynamic forms, tables, and frames were incorporated into the database for enhanced data security and fast, easy user interaction. It uses Structured Query Language (SQL) queries and stored procedures to extract eSSR primer data along with other necessary information in both customized query mode and batch mode. Static pages were integrated to provide an introduction to jute and recent publications from various sources.

2.6. SSR-PCR amplification

Seventy-four randomly selected eSSR primer pairs were used for validation in a panel of 24 jute accessions and in two parents of an interspecific hybrid population. Each 20 μL SSR-PCR reaction mixture consisted of 1 U of *Taq* DNA polymerase with 1× reaction buffer (Promega, USA), 2.0 mmol L^{-1} $MgCl_2$, 40 μmol L^{-1} each dNTP mix, 0.5 μmol L^{-1} each of forward and reverse primer, and 50 ng DNA template. Thermal cycling was performed in a Mastercycler pro (Eppendorf Inc., Germany) instrument under the following conditions: initial denaturation at 94 °C for 5 min followed by 36 cycles of denaturation at 94 °C for 50 s, annealing at 52–56 °C for 1 min, and extension at 72 °C for 1 min, with final extension at 72 °C for 7 min. PCR products along with ready-to-use 100 bp GeneRuler (Thermo Fisher Scientific, USA) were separated on 8% non-denaturing polyacrylamide gel (Sigma Aldrich) in 0.5× Tris borate EDTA (44.5 mmol L^{-1} Tris, 44.5 mmol L^{-1} boric acid, and 1.0 mmol L^{-1} EDTA, pH 8.0) buffer. All electrophoresis was performed at 120 V for 3 h in a vertical gel electrophoresis unit (Hoefer, USA) and fragments were stained with 0.5 μg mL^{-1} ethidium bromide solution for 5 min before image capture with a gel documentation system (AlphaImager HP, USA).

2.7. Data acquisition and statistical analysis

Only prominent bands with expected allele sizes, excluding stutter bands, were scored manually. For more than two alleles, as observed with few markers, only alleles corresponding to expected sizes were considered for scoring, thus eliminating undesirable alleles from the analysis. Statistical analysis of allele number (N_A), gene diversity (H_e), heterozygosity (H_o), and polymorphism information content (PIC) was performed with PowerMarker 3.25 [23]. An unrooted Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram was generated using a proportion of the shared alleles and viewed with MEGA 6.06 software [24] to assess genetic polymorphism among the 24 jute accessions.

3. Results

3.1. Distribution of SSRs

The primary resource for eSSR mining was 34,163 jute unigenes, consisting of 49.2 million nucleotides, which were retrieved from the NCBI-TSA accession numbers GBSD01000001 to GBSD01034163 [12]. The Perl script-based MISA tool identified 3827 (11.2%) unigene sequences harboring 4509 (13.2%) putative SSR loci. The mining of SSR loci included di- to hexanucleotides but ignored mononucleotide repeats (Table 2). More than one SSR was observed in 576 sequences, and 197 unigene sequences contained compound SSRs, with more than two different motif types present within <100 bases apart. On average, every 10.9 kb of jute unigene sequence contained one SSR repeat.

Analysis of the distribution pattern of different SSR repeat units revealed that trinucleotide is the most abundant repeat type, with 2527 loci (56.0%), followed by 1889 loci (41.9%) of dinucleotide repeats (Fig. 1-a). Tetra-, penta-, and hexanucleotide repeats were present in low numbers, totaling 93 (2.0%). The maximum frequency 1018 (22.6%) of classified SSR repeat motifs was found as dinucleotide $(AG/CT)_n$, followed by 948 (21.0%) trinucleotide $(AAG/CTT)_n$ and 790 (17.5%) dinucleotide $(AT/AT)_n$ repeat types. With respect to the number of iterations of SSR motifs, the maximum of 12 iterations was present in nine unigenes (Fig. 1-b). Based on number of iterations, SSR loci were placed into two classes: class I contained ≥20 nucleotides also known as hypervariable SSRs, whereas class II, containing ≥12 but <20 nucleotides, contained potentially variable SSRs. Of 2079 jute eSSRs, 492 SSR loci were assigned to class I (10.9%) and the rest to class II. Based on ab initio gene prediction of eSSR-containing unigenes, the positions of 769 SSR loci were located in cds (36.9%), those of 619 in 5′ and 3′ UTRs (29.8%) and those of 61 within introns (2.9%). SSR loci lying outside predicted genes numbered 364 (17.5%), whereas 266 (12.8%) SSR loci-containing unigene sequences failed to correspond to any genes (were noncoding), using the *Arabidopsis* genes as a model system.

3.2. Primers designed from SSR repeats and anchored to a linkage map

The MISA output for 3671 unigene sequences was used as an input file to SSRLocator v.1.0, which produced 3032 primer sequences for di-, tri-, and tetranucleotide repeats. Further, the virtual PCR module of the SSRLocator program and manual inspection generated a nonredundant set of 2079 forward and reverse primer sequences (NCBI Probe accession numbers Pr032754357 to Pr032756435) from 1909 unigene sequences (Table 3). The primer sequences and related data, such as primer melting temperature (T_m), product size in base

Table 2 – Summary statistics of unigene sequence data mining and SSR loci identification using MISA software.

Feature	Value
Input unigene sequences to MISA program (numbers)	34,163
Total input sequence size (in bp)	49,263,066
Mean length of unigene sequences (in bp)	1442
Total number of identified SSRs (di-, tri-, tetra-, penta-, and hexa)	4509
Number of SSRs containing unigene sequences	3827
Number of sequences containing more than 1 SSR	576
Number of SSRs present in compound formation	197
Dinucleotide SSR loci	**1889**
AC/GT	81
AG/CT	1018
AT/AT	790
Trinucleotide SSR loci	**2527**
AAC/GTT	141
AAG/CTT	948
AAT/ATT	232
ACC/GGT	179
ACG/CGT	42
ACT/AGT	39
AGC/CTG	312
AGG/CCT	164
ATC/ATG	389
CCG/CGG	81
Tetranucleotide SSR loci	**80**
Penta- and hexanucleotide SSR loci	**13**

pairs and start and end position of the repeat sequences are presented in Table S1. A total of 1254 (60.3%) primer pairs were associated with trinucleotide repeats, followed by 781 (37.6%) with dinucleotide and 44 (2.1%) with tetranucleotide repeats. Of 2079 primer pairs, 173 primers (8.3%) will amplify sequences with class I repeats, which may be hypervariable in nature.

We further attempted to anchor the eSSR markers on the existing RAD linkage map of jute, consisting of 503 RAD markers (SNPs), using BLASTN search. At an *E*-value cutoff of 1e−10, we could identify top hits of RAD markers for 18 eSSR-containing unigenes with >90% identity. Subsequently, based on the actual position of the SSR repeats, we could successfully map them adjacent to RAD markers on five of the seven linkage groups (LG1–5) (Fig. S1). LG1 harbored the maximum of six eSSR markers, followed by LG2 with five. Three, three, and one eSSR marker were mapped on LG3, LG4, and LG15, respectively. Analysis of the recently published [25] draft genome sequences of jute (*C. capsularis* and *C. olitorious*) showed that 1715 (82.5%) and 512 (24.6%) of the eSSR markers from the present study possesses primer-binding complementary sites, respectively. These markers thus hold promise for integration into the physical map of jute.

3.3. Gene ontology (GO) annotation of unigenes containing eSSRs

Functional annotations of the 1909 unigenes containing eSSRs were analyzed with Blast2GO software for putative functional categorization (Table 4). A total of 1847 (96.8) unigene sequences were matched by BLASTX against the NCBI-NR database at an *E*-value cutoff of 1e−50. Among these, 1282 (67.2%) were fully annotated using the GO-Slim (Plant slim) module for putative protein-coding functions, whereas the rest 797 (41.7%) could not be annotated, owing either to an absence of characterization information or lack of significant homology to known genes. Top-hit species distribution revealed that the 76.2% jute unigene sequences carrying eSSRs bear similarity to *Theobroma cacao* genes. Using the Kyoto Encyclopedia of Genes and Genomes (KEGG), analysis of the Blast2GO-annotated 567 unigenes showed 248 of these matching to known Enzyme Commission (EC) codes. On the basis of GO terms, the annotated eSSR-containing unigene sequences were further categorized into cellular components, molecular functions, and biological processes. According to the cellular component GO, 540 sequences were associated with cellular localization, 351 to organelles, and 301 to the membrane. Categorization of GO terms associated with molecular functions revealed the highest activity of these unigenes for heterocyclic and organic cyclic compound binding followed by transferase activity and small-molecule binding. At the biological process level, the most GO terms were found to be associated with various metabolic processes, followed by biosynthetic processes and cellular component organization. A total of 134 and 104 unigene sequences containing eSSRs were categorized in “response to stress” and “response to abiotic stimuli”, respectively. Forty-two unigenes containing eSSRs fell into the category “response to biotic stimuli”. A total of 117 unigene sequences with eSSRs were classified into the anatomical structure development process. A comprehensive GO categorization of unigenes with eSSRs is graphically depicted in Fig. 2.

3.4. Nature of ORFs and functional domains in unigenes with eSSRs

For functional categorization of the eSSR markers, we further analyzed the nature and location of predicted ORFs and the

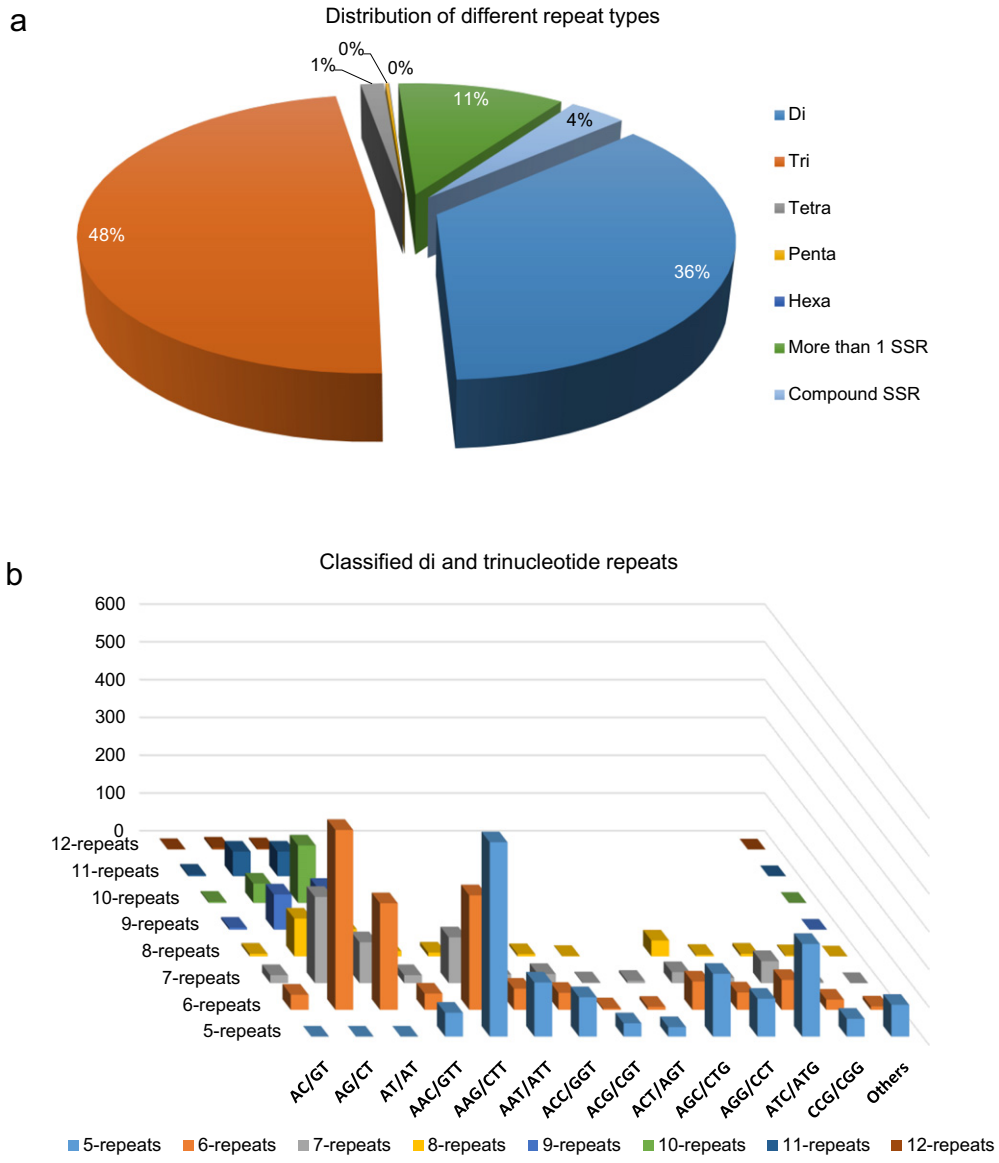

Fig. 1 – Composition of eSSR motifs in jute unigenes. (a) Percent distribution of 4509 eSSR repeat types from jute unigenes as calculated by the MISA program. (b) Frequency distribution of classified di- and trinucleotide repeats in jute unigenes. The vertical axis represents abundance of eSSR motifs with different repeat numbers (from 5 to 12). See color figure online for legends.

functional domains. Of the 1909 unigene sequences consisting of eSSRs, all but two of the unigene sequences were found to encode ORFs ranging from 22 to 3205 amino acids in length, with an average of 371.6 residues per ORF (Table 5). We further filtered the 1907 ORF-coding unigene sequences with eSSRs at a cutoff value of ≥200 amino acid

Table 3 – Details of primer designing from SSR-containing sequences of jute, using SSRLocator.

Feature	Value
Input SSR loci containing nonredundant sequences (numbers)	3671
Total number of SSRs identified by SSRlocator (di-, tri-, and tetra-)	4189
Total number of nonredundant sequences for primer design	2079
Dinucleotide repeat sequences for primer design	781
Trinucleotide repeat sequences for primer design	1254
Tetranucleotide repeat sequences for primer design	43
Total primers for hypervariable class I SSR (i.e., ≥ 20 repeat bases)	173

Table 4 – Functional annotation of eSSR-containing unigene sequences of jute.

Feature	Value	Percentage (%)
Total number of sequences analyzed with Blast2GO[a]	1909	–
With Gene Ontology (GO)-Slim (Plant) annotation	1282	67.2
With BLASTX hits	371	19.4
Without BLASTX hits	232	12.2
With mapping	180	9.4
With only InterProScan	14	0.7
Top Hit species -*Theobroma cacao*	1454	76.2
KEGG annotation	567	29.7
Annotation with known enzyme code (EC)	248	13.0

[a] BLASTX E-value cutoff ≥1e−50.

residues, leaving 1616 unigene sequences. This filter eliminated possible long noncoding RNA sequences. The 1616 unigene sequences with eSSRs were scanned with InterProScan to identify functional domains encoded by the unigenes. A total of 907 (43.6%) unigene sequences were predicted to harbor eSSR loci within the predicted ORF sequence. Among the 907, 139 encoded major transcription factors (TFs), 70 (R-genes) encoded disease-resistance proteins, and 65 encoded various repeat proteins, such as tetratricopeptide repeats (TPR), pentatricopeptide repeats

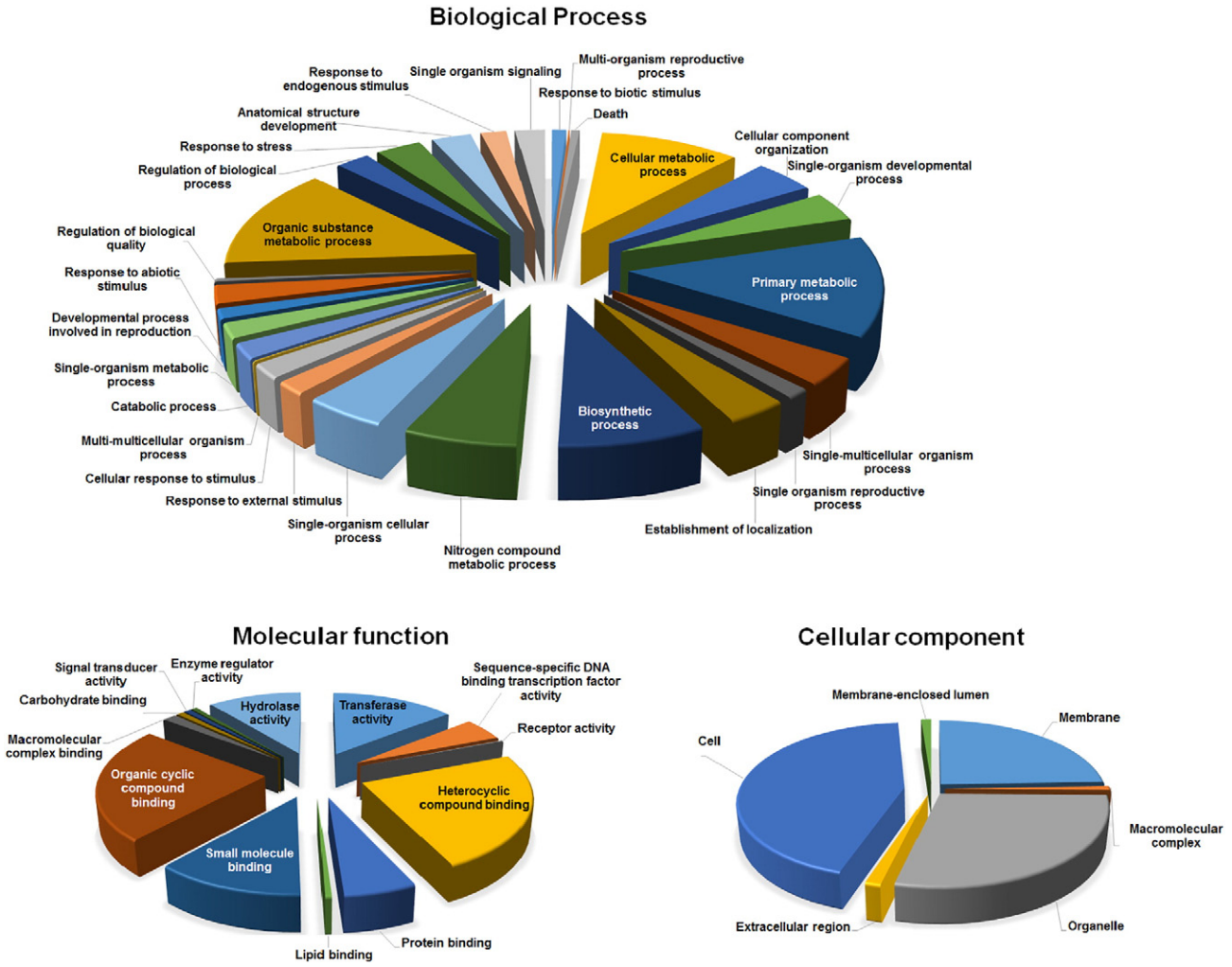

Fig. 2 – Gene ontology (GO) annotation of 1909 jute unigene sequences harboring eSSR loci. **The GO biological process, molecular function and cellular component pie graphs depict level 3 annotations, respectively. Color figure available online.**

Table 5 – ORF and functional domain predictions in jute unigenes with eSSRs.

Feature	Value
Total number of nonredundant eSSRs	2079
Unigenes containing eSSRs	1909
Unigenes with eSSRs with predicted open reading frame (ORF) length of ≥200 amino acid residues	1616
Unigenes with eSSR located within ORFs	907
eSSRs from major transcription factors (TF) encoded by unigenes	139
eSSRs from resistance (R)-genes encoded by unigenes	70
eSSRs from repeat domain proteins encoded by unigenes	65

(PPR), leucine-rich repeats (LRR), and WD40 proteins. Information on different functional domains in the eSSR containing jute unigenes indicate they can be exploited as potential functional domain markers (FDMs) and provide valuable insights into polymorphism associated with a putative trait.

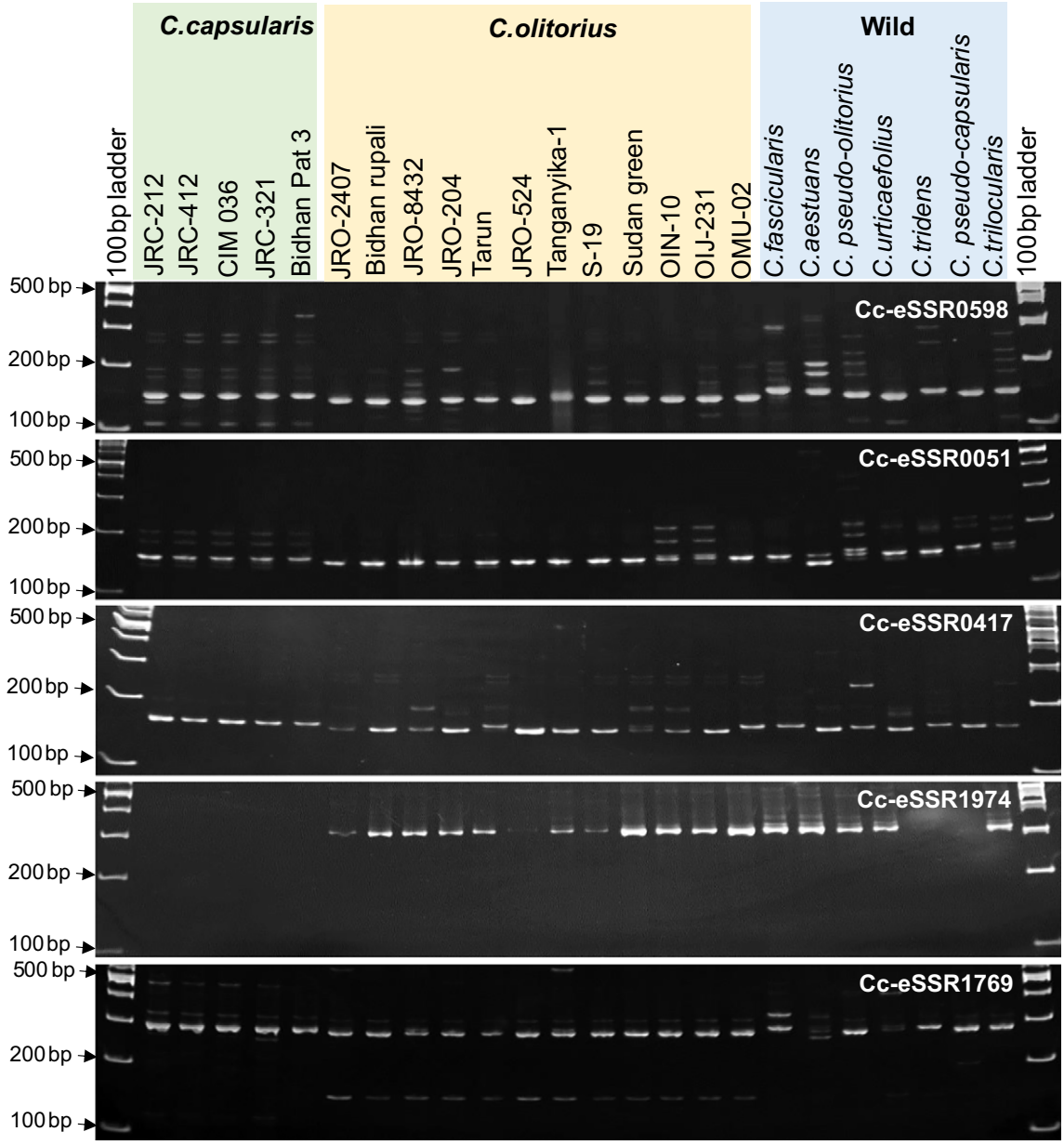

Fig. 3 – Representation of allelic variations in a panel of 24 jute accessions produced by five eSSR markers on 8% non-denaturing polyacrylamide gels stained with ethidium bromide. Major bands of DNA size standards are depicted with arrowheads on the left.

Table 6 – Details of primer efficiency assay using PowerMarker v.3.25.

Sl. no.	Marker	Repeat type	SSR location	Allele number (N_A)	Gene diversity (H_e)	Heterozygosity (H_o)	Polymorphism information content (PIC)
1	Cc-eSSR1265	(TCT)5	Outside gene	2	0.04	0.04	0.04
2	Cc-eSSR0598	(CTT)5	Cds	2	0.46	0.04	0.35
3	Cc-eSSR0698	(ATA)6	Cds	3	0.49	0.21	0.43
4	Cc-eSSR0757	(ACA)6	5′ UTR	2	0.44	0	0.35
5	Cc-eSSR1364	(TA)8	3′ UTR	4	0.62	0	0.57
6	Cc-eSSR1678	(GAA)5	Cds	2	0.22	0	0.19
7	Cc-eSSR1057	(GCA)6	Non-coding	3	0.29	0	0.26
8	Cc-eSSR0051	(TTG)5	Cds	3	0.44	0.13	0.36
9	Cc-eSSR0324	(TA)6	3′ UTR	3	0.38	0.04	0.32
10	Cc-eSSR0417	(AAC)5	Cds	2	0.48	0.04	0.36
11	Cc-eSSR1616	(AG)6	5′ UTR	4	0.57	0	0.50
12	Cc-eSSR1275	(TGA)7	Cds	2	0.22	0	0.19
13	Cc-eSSR0009	(TCT)6	5′ UTR	2	0.22	0	0.19
14	Cc-eSSR0440	(GGA)6	cds	2	0.08	0	0.08
15	Cc-eSSR1092	(CCT)5	Non-coding	3	0.45	0.04	0.38
16	Cc-eSSR0776	(GGA)7	Cds	3	0.51	0.04	0.41
17	Cc-eSSR0800	(CAT)5	Cds	3	0.66	0.17	0.59
18	Cc-eSSR0311	(AT)10	3′ UTR	4	0.61	0.29	0.55
19	Cc-eSSR0430	(CAA)5	Cds	2	0.08	0	0.08
20	Cc-eSSR1129	(CT)9	Non-coding	3	0.38	0.13	0.34
21	Cc-eSSR0505	(ATC)7	cds	2	0.47	0	0.36
22	Cc-eSSR0574	(TGA)7	Cds	2	0.47	0	0.36
23	Cc-eSSR2053	(TGC)8	Cds	3	0.19	0.04	0.18
24	Cc-eSSR0432	(TA)6	5′ UTR	3	0.48	0.21	0.43
25	Cc-eSSR1124	(AT)6	Non-coding	4	0.53	0.21	0.49
26	Cc-eSSR2038	(CAT)7	cds	4	0.63	0.13	0.58
27	Cc-eSSR1904	(AG)7	3′ UTR	2	0.44	0	0.35
28	Cc-eSSR0641	(TG)6	3′ UTR	3	0.5	0	0.41
29	Cc-eSSR1974	(AT)6	Cds	2	0.41	0	0.33
30	Cc-eSSR1801	(GGT)5	Cds	2	0.28	0	0.24
31	Cc-eSSR1334	(ACT)6	Cds	3	0.32	0.21	0.29
32	Cc-eSSR1026	(TGC)7	Non-coding	2	0.5	0.92	0.37
33	Cc-eSSR0427	(CAA)6	Cds	3	0.5	0	0.41
34	Cc-eSSR2021	(TGA)8	Cds	3	0.29	0	0.27
35	Cc-eSSR1172	(CAC)6	Cds	2	0.49	0	0.37
36	Cc-eSSR1879	(AT)10	Non-coding	5	0.64	0.21	0.60
37	Cc-eSSR0225	(GAC)7	Cds	2	0.15	0	0.14
38	Cc-eSSR1487	(AT)8	3′ UTR	3	0.43	0.08	0.37
39	Cc-eSSR1462	(AT)7	Non-coding	2	0.22	0	0.19
40	Cc-eSSR0107	(GA)9	Outside gene	2	0.5	0.96	0.37
41	Cc-eSSR1769	(AAC)6	Cds	3	0.08	0.04	0.08
42	Cc-eSSR0270	(TGA)7	5′ UTR	4	0.57	0.21	0.52
43	Cc-eSSR1048	(GA)7	Non-coding	3	0.16	0	0.15
	Mean			2.74	0.39	0.1	0.34

Cds, coding sequence; noncoding, having no gene predictions or encoding fewer than 100 amino acids; UTR, untranslated region.

3.5. *Cross-species validation and genetic polymorphism analysis*

Cross-species transferability is one of the major advantages of expressed gene-derived SSR markers. Accordingly, we checked the cross-species transferability of 74 randomly selected primer pairs from the 2079 identified jute eSSR primer pairs (Table S2). Of the 74, 53 (71.6%) were successfully amplified in a panel of 24 jute accessions. Forty-three (58.1%) and ten (13.5%) primer pairs produced polymorphic and monomorphic patterns, respectively. Four (5.4%) of the eSSR primers produced no amplifications, whereas 17 (23.0%) produced multiple bands, which were not included in the statistical analysis of transferability, primer efficiency, and genetic similarity. We estimated the percentage of primer transferability on the basis of amplification in at least one *C. olitorius* and seven wild species of jute. The percentage of transferability among related jute species ranged from 37.5 to 100.0 with an average of 88.7. Overall, the allele sizes in *C. capsularis* and *C. olitorius* differed for the majority of the primer pairs assessed. Apart from the expected size of the eSSR loci, distinct alleles were found in wild species of jute, indicating a greater range of polymorphism at the species level than in the

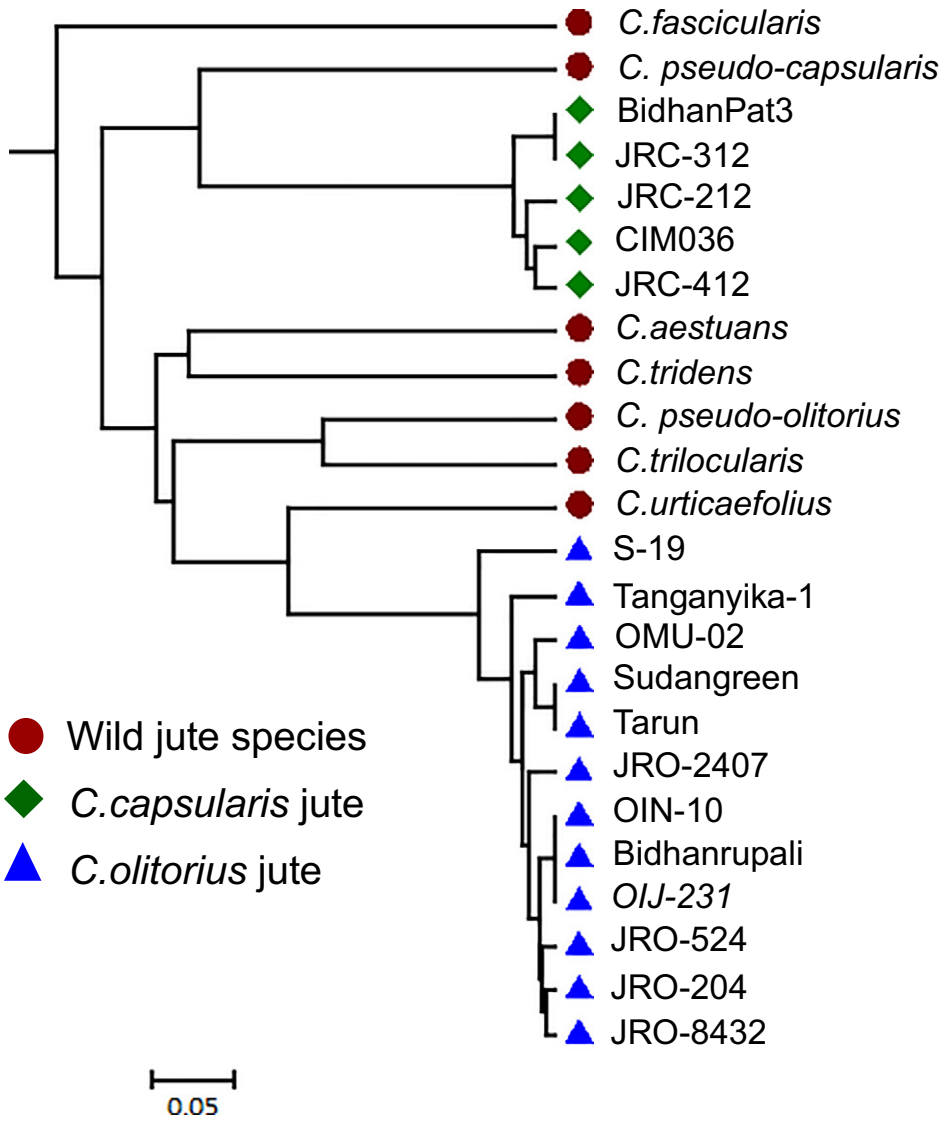

Fig. 4 – UPGMA dendrogram showing genetic relationships among 24 jute accessions, derived from proportion of shared alleles of 43 Cc-eSSR primers by PowerMarker v.3.25 and viewed using MEGA 6.06 software.

cultivated jute species alone (Fig. 3-a). In an investigation of polymorphism in parental genotypes of an interspecies cross OIJ-248 (cultivated *C. olitorius* as female parent) × WCIN-136 (wild *C. aestuans* as male parent), thirteen and three primers showed variations in allele size and presence/absence of alleles, respectively, between them, suggesting these primers may be used in interspecific hybrid identification (Fig. S2).

Codominant allele data were used to calculate the number of observed alleles (N_A), gene diversity (H_e), heterozygosity (H_o) and polymorphism information content (PIC) using PowerMarker software (Table 6). The N_A, H_e, and H_o ranged from 2.00 to 5.00 (mean N_A 2.74), 0.04 to 0.66 (mean H_e 0.39), and 0 to 0.96 (mean H_o 0.10), respectively. Values of PIC ranged from 0.04 to 0.60 with a mean of 0.34. Seven (9.5%) eSSR primer pairs yielded PIC values >0.50, suggesting the potential of these primers to measure high genetic diversity. All but one of the eSSR markers with high (>5.0) PIC values and high (>3) allele numbers are related to di- and trinucleotide repeats, with an average high repeat unit count of 7.6 (Table 6). Twenty-six (35.1%) eSSR primer pairs yielded PIC values ranging from 0.25 to 0.50, showing their potential to measure a moderate level of genetic diversity. The rest of the eSSR primers yielded low PIC values below 0.25. The codominant marker data were further employed for the construction of a UPGMA dendrogram showing genetic relationships among the 24 jute accessions (Fig. 4). Clusters of *C. capsularis* and *C. olitorius* jute accessions were differentiated distinctly from each other and from the wild jute species. According to the dendrogram, the wild jute species *C. fascicularis* was the most diverse among the 24 jute accessions used in the study. The majority of the wild jute species showed greater genetic similarity to *C. olitorius* than to *C. capsularis* accessions, indicating the possibility of their interspecies crosses. Overall, these results reveal the

potential utility of the identified eSSR primers for cross-species genetic diversity analysis, interspecific hybrid or variety identification, and marker-assisted breeding applications.

3.6. Web access of jute eSSR markers via JuteMarkerdb

To facilitate the use of jute eSSR markers for genetic polymorphism and linkage mapping studies, we constructed a web-based, downloadable and searchable database of jute markers (JuteMarkerdb) accessed via URL http://jutemarkerdb.icar.gov.in/ (Fig. S3). The current version of JuteMarkerdb contains basic data for the SSR loci derived from *Corchorus* and the details of the primers designed from SSR flanking unigene sequences. Users may browse through all the SSR markers using a "Browse all" option or may batch-download all the primers or groups of eSSR primers based on repeat motifs (di-, tri-, or tetranucleotide), length of repeat nucleotides, or annotations. These options are provided as a dropdown list. It is also possible to perform search operations for specific eSSR types in query access mode, for either repeat-unit type, number of repeats, or SSR motif-containing unigene annotations. JuteMarkerdb offers the option to update with other functional markers from jute and allied fibers in the future. Pages introducing the database and jute and a hyperlinked up-to-date set of literature references for jute genomics and biotechnology are provided for rapid access to information.

4. Discussion

4.1. Large-scale functional marker resource development in jute

Development of molecular markers in the cultivated jute species *C. capsularis* and *C. olitorius* aims to speed up genetic studies and breeding for quality enhancement and production of commercially important bast fiber. Several molecular markers, such as random markers [26,27], genomic SSRs [7,11,28], and SNPs [10,12] have been reported in jute for use in genetic diversity studies and linkage map construction. Random molecular markers, though easy to generate and use in genetic diversity assessment, are often inefficient in breeding applications owing to low genetic information content and lack of repeatability [29,30]. SNP markers are high-throughput in nature and are highly abundant in genomes, but their generation and use in breeding applications remains skill-intensive and expensive. For these reasons, SNP markers cannot be afforded by the majority of laboratories with low to medium funding support. Thus, SSR markers, which are simple to assay and analyze, are highly rated as a marker of choice for numerous breeding applications, such as genome analysis, QTL mapping, genome-wide association, and marker-assisted selection. Between genomic and expressed gene-derived SSR markers, the latter have emerged as highly preferred markers owing to their simplicity and ease of large-scale development from next-generation sequencing data or public sequence databases. Polymorphism information in the eSSR markers is typically derived from the functional units or transcribed region of genes and they are more amenable to cross-species transfer for analysis of genetic polymorphism and population structure of germplasm [31,32]. In jute, transcriptome sequence of bast fiber tissue from Indian *C. capsularis* variety JRC-212 is available with 34,163 assembled unigene sequences (NCBI-TSA accessions GBSD01000001 to GBSD01034163) [12]. This data served as a primary resource for SSR data mining and primer development in the present study. In a contemporary study, Zhang et al. [14] reported development of 1906 EST-SSR markers from the bast fiber transcriptome of Chinese *C. capsularis* variety Huangma 179 [33]. The eSSR primers developed in the present study are nonredundant with respect to the above set (containing more than 85% unique SSRs).

4.2. Distribution of SSR loci in jute unigenes for marker development

Analysis of the frequency distribution of SSR loci in transcriptome sequence is important for developing genome-wide markers. In the present study, we discovered 4509 SSR loci, resulting in a chance of locating one SSR in every 10.9 kb of jute unigene sequence. As per our observation, SSR frequency distribution value in jute unigene is high, 13.2% compared to the SSR distribution of 3.9% in the transcriptome data of Chinese jute variety Huangma 179. The percent of SSR distribution is also higher compared to reports from other bast fiber crops, such as kenaf (12.3%) [34], ramie (4.3%) [35], and flax (3.5%) [36]. The above results indicate that the abundance of SSR loci in jute unigenes are potential in developing a large-number of nonredundant SSR loci-flanking primer pairs.

Analysis of the distribution of different repeat motifs showed that trinucleotide repeats were the most abundant (56.0%) type, followed by dinucleotide repeats (41.9%), with the two together comprising 98.0% of the eSSR repeats. This finding is important for developing a large number of effective eSSR primers in jute, as the markers associated with smaller SSR motif lengths (di- and tri-) are more variable than other motif lengths [37]. The abundance of trinucleotide repeats in the transcriptome of the Chinese jute variety Huangma 179 was similar, but that transcriptome differed in the abundance of dinucleotide and other repeat motifs [14]. A similar high abundance of trinucleotide repeats was reported in flax [36] and kenaf ESTs [34], but an almost equal distribution of di- and trinucleotide repeats was reported in ramie [35]. The repeat motifs AG/CT and AAG/CTT were the most prevalent form of SSR motif, in agreement with findings from the Chinese jute variety Huangma 179 [14], kenaf [34], ramie [35] and other dicot plant species [38]. The lowest number of iterations of SSR motifs in jute unigenes was ≤12 units, a result likely due to higher selection pressure in coding than in noncoding regions of the genome [31]. While attempting to map the eSSR markers on the existing RAD linkage map, we successfully anchored 18 of them in close proximity to SNP markers. The low number of eSSR-containing unigenes with identity to RAD markers could be explained by the fact that the SNP markers were developed from genomic DNA sequences of an F_2 population from the cross Sudan green (SG) × bast fiber shy (*bfs*) mutant of *C. olitorius* [20]. With the recent publication of draft genomes of the two jute species *C. capsularis* and *C. olitorius* [25], we found that >82.5% and 24.6% eSSR markers have matching primer-binding sequences and

thus are potential to map on the respective genomes for breeding applications.

4.3. *Jute eSSRs linked to diverse functions*

Compared to SSRs from noncoding portions of the genome, SSRs identified from expressed genes can be potentially linked to a phenotypic trait or to functional diversity, owing to their possible role in altering transcription or translation processes [31]. In the present study, functional annotation of 67.2% of jute eSSRs containing unigenes showed significant association with various biological functions and molecular processes. The majority of these unigenes were linked to metabolism and regulation of transcription and translation. In the biological process category, top hits were associated with plant development, reproduction, and stress signaling process. A high proportion (84.6%) of jute unigenes containing eSSR loci were predicted to encode functional proteins, in which 47.5% of eSSRs lay within ORF coding regions, suggesting their potential use in mapping functional diversity. The results justify optimism for the development of candidate gene-based functional markers, as demonstrated in barley [39] and sugarcane [40]. These markers will serve as a key resource for mapping economically important traits linked to genes using existing mapping populations in jute [9,41].

4.4. *Jute eSSRs are potential cross-species marker for genetic polymorphism analysis*

Owing to the advantages of functional markers being directly or indirectly associated with traits, SSRs derived from the expressed regions of a genome can significantly contribute to direct allele selection. These SSR-flanking markers, when polymorphic, are efficient in identifying meaningful allelic variations across genotypes and species [31]. Thus, from a marker-utility perspective, it makes sense to validate the jute eSSR markers in different jute accessions and related wild jute species. In our study, an assessment of 74 random jute eSSR primer pairs (21 dinucleotide, 50 trinucleotide, and three tetranucleotide repeats) yielded 71.6% success in cross-species amplifications. The primers for validation were chosen to include more di- and trinucleotides with higher repeat numbers, as they are potentially variable in nature [42]. From our results, it was evident that the di- and trinucleotide repeats with higher numbers of iterations were highly polymorphic and that the majority of these SSR loci were located in the untranslated regions (UTRs) of predicted genes. Among the unsuitable primers, 5.4% did not yield any amplifications and the rest produced multiple bands, most likely owing to the presence of introns [43] or redundant primer-binding complementary sites [44], respectively. The proportion of polymorphic primers (58.1%) among the validated primers, is low in comparison with that in the Chinese jute variety Huangma 179 (83.6%) [14] and with ramie (81.0%) [35], but was similar to that in kenaf (53.9%) [34]. However, the reduced proportion of primer polymorphism is highly influenced by the random selection of primers and scoring method adopted. The average number of alleles and PIC values, which indicate the efficiency of the primers for distinguishing genotypes, were high (2.74 alleles and a PIC of 0.34) compared to those values for the Chinese jute variety Huangma 179 (1.4 alleles and PIC of 0.09) [14] and were similar to those for flax EST-SSRs (2.3 alleles and PIC of 0.35) [36]. In our study, the eSSR markers showed distinct variation in allele size among different *Corchorus* species, such as between *C. capsularis* and *C. olitorius* and wild jute species. The allelic variations were sufficient to generate a clustering dendrogram and distinguish the jute accession by clade. Accordingly, these alleles can serve as a marker of choice for characterization of populations, varieties, and germplasm, as reported in wheat [45], jatropha [46], and oil palm [47]. The small number of distinct alleles observed in wild jute profiles are interesting, given that it can be exploited in developing markers for interspecific hybrid identification [48]. In a parental polymorphism demonstration between a cultivated jute female parent and a wild jute male parent, eSSR primers could clearly distinguish these parents, leading to the proposition that these markers could be a very good candidate for hybrid identification in a wide-crossing program.

4.5. *JuteMarkerdb as a resource for marker selection*

With the rising importance of developing functional markers in jute, researchers and academicians may benefit by selecting functional SSR markers from compiled information. For this reason, besides discovery of eSSR markers, it was considered highly desirable to develop a user-friendly database of jute eSSR markers. Similar databases have been published in several crops, such as chickpea [49], pigeon pea [50], tomato [51], and cotton [52]. To our knowledge, no molecular marker database has been developed for jute and related bast fiber crops.

Here we report the development of a user-friendly and a comprehensive web-accessible database for jute functional markers (JuteMarkerdb) with detailed information on SSRs located in the *C. capsularis* assembled unigene sequences. The nonredundant set of 2079 primer pairs will help users to select markers for a variety of applications, such as diversity analysis, genome evolution and population structure studies, hybrid identification, gene tagging, and breeding applications. Our categorization of eSSR markers is also expected to assist in the selection of trait-specific functional markers, such as for disease resistance or stress responses. The database will allow users to browse all of the jute eSSR markers along with related information or to batch-download them with a variety of filtering options, such as repeat type, repeat unit size, location of SSRs in ORFs, and unigene annotation. A query-based approach is also included to allow specific filtration of the jute eSSR markers.

5. Conclusions

We have developed a large set of 2079 nonredundant eSSR markers from publicly available jute unigene sequences from di-, tri-, and tetranucleotide repeats, which are efficient in polymorphism detection. Only eighteen eSSR markers were placed on the existing RAD map in five linkage groups, but majority of them hold promise of being placed on the physical map of jute as the genome sequence became available. The eSSR resource in jute was sampled with randomly selected primers to

demonstrate their potential for cross-species use. One limitation with the eSSR markers is that only a small fraction of them was experimentally validated in different jute accessions. More number of these markers need wet-lab validations and cross-transferability analysis in other related fiber crops, such as kenaf and roselle for their efficient utilization. Finally, the analysis and findings were compiled comprehensively in a user-friendly web database, probably the first in jute and allied bast fiber crops. This jute eSSR resource is expected to provide a huge benefit to jute and allied fiber researchers and academicians to explore and use this primer set in a variety of applications, including analysis of genetic diversity and population structure, hybrid identification, gene tagging, trait association, and breeding applications.

Supplementary data for this article can be found online at http://dx.doi.org/10.1016/j.cj.2017.02.006.

Acknowledgments

The authors acknowledge Indian Council of Agricultural Research-Central Research Institute for Jute and Allied Fibres (ICAR-CRIJAF), Barrackpore, for providing necessary infrastructure and financial support for this work. The authors thank ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India for assistance in hosting the database.

REFERENCES

[1] V.H. Heywood, R.K. Brummitt, A. Culham, O. Seberg, Flowering Plant Families of the World, Kew Publishing, Royal Botanic Gardens, Kew, Richmond, Surrey, UK, 2007.
[2] R.M. Rowell, H.P. Stout, Jute and kenaf, in: M. Lewin (Ed.), Handbook of Fiber Chemistry, third ed.CRC Press, Boca Raton, Florida, USA 2007, pp. 405–452.
[3] C. Roul, The International Jute Commodity System, Northern Book Centre, New Delhi, India, 2009.
[4] M.S. Rahman, Jute: a versatile natural fibre. Cultivation, extraction and processing, in: J. Müssig (Ed.), Industrial Applications of Natural Fibres: Structure, Properties and Technical Application, John Wiley & Sons Ltd, Chichester, UK 2010, pp. 135–161.
[5] C.S. Kar, A. Kundu, D. Sarkar, M.K. Sinha, B.S. Mahapatra, Genetic diversity in jute (*Corchorus* spp) and its utilization: a review, India, J. Agric. Sci. 79 (2009) 575–586.
[6] S. Maity, S. Chowdhury, A.K. Datta, Jute biology, diversity, cultivation, pest control, fiber production and genetics, in: Eric Lichtfouse (Ed.), Organic Fertilization, Soil Quality and Human Health, Sustainable Agriculture Reviews 9, Springer, Netherlands 2012, pp. 227–262.
[7] M. Das, S. Banerjee, R. Dhariwal, S. Vyas, R.R. Mir, N. Topdar, A. Kundu, J.P. Khurana, A.K. Tyagi, D. Sarkar, M.K. Sinha, H.S. Balyan, P.K. Gupta, Development of SSR markers and construction of a linkage map in jute, J. Genet. 91 (2012) 21–31.
[8] Y.P. Chen, L.W. Zhang, J.M. Qi, H. Chen, A.F. Tao, J.T. Xu, L.H. Lin, P.P. Fan, Genetic linkage map construction for white jute (*Corchorus capsularis* L.) using SRAP, ISSR and RAPD markers, Plant Breed. 133 (2014) 777–781.
[9] A. Kundu, A. Chakraborty, N.A. Mandal, D. Das, P.G. Karmakar, N.K. Singh, D. Sarkar, A restriction-site-associated DNA (RAD) linkage map, comparative genomics and identification of QTL for histological fibre content coincident with those for retted bast fibre yield and its major components in jute (*Corchorus olitorius* L., *Malvaceae* s. l.), Mol. Breed. 35 (2015) 19.
[10] C. Biswas, P. Dey, P.G. Karmakar, S. Satpathy, Discovery of large-scale SNP markers and construction of linkage map in a RIL population of jute (*Corchorus capsularis*), Mol. Breed. 35 (2015) 119.
[11] L.W. Zhang, M.H. Yuan, A.F. Tao, J.T. Xu, L.H. Lin, P.P. Fang, J.M. Qi, Genetic structure and relationship analysis of an association population in jute (*Corchorus* spp.) evaluated by SSR markers, PLoS One 10 (2015) e0128195.
[12] A. Chakraborty, D. Sarkar, P. Satya, P.G. Karmakar, N.K. Singh, Pathways associated with lignin biosynthesis in lignomaniac jute fibres, Mol. Gen. Genomics. 290 (2015) 1523–1542.
[13] P. Samanta, S. Sadhukhan, A. Basu, Identification of differentially expressed transcripts associated with bastfibre development in *Corchorus capsularis* by suppression subtractive hybridization, Planta 241 (2015) 371–385.
[14] L.W. Zhang, Y.R. Li, A.F. Tao, P.P. Fang, J.M. Qi, Development and characterization of 1,906 EST-SSR markers from unigenes in jute (*Corchorus* spp.), PLoS One 10 (2015) e0140861.
[15] A. Kundu, D. Sarkar, A. Bhattacharjee, N. Topdar, M.K. Sinha, B.S. Mahapatra, A simple ethanol wash of the tissue homogenates recovers high-quality genomic DNA from *Corchorus* species characterized by highly acidic and proteinaceous mucilages, Electron. J. Biotechnol. 14 (2011) 1.
[16] T. Thiel, W. Michalek, R. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L), Theor. Appl. Genet. 106 (2003) 411–422.
[17] L.C. da Maia, D.A. Palmieri, V.Q. de Souza, M.M. Kopp, F.I. de Carvalho, A.C. de Oliveira, SSR Locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation, Int. J. Plant Genomics 412696 (2008).
[18] M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics 19 (2003) ii215–ii 225.
[19] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.
[20] R. Van Berloo, GGT 2.0: versatile software for visualization and analysis of genetic data, J. Hered 99 (2008) 232–236.
[21] X.J. Min, G. Butler, R. Storms, A. Tsang, OrfPredictor: predicting protein-coding regions in EST-derived sequences, Nucleic Acids Res. 33 (2005) W677–W680.
[22] S. Hunter, R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A.L. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A.F. Quinn, J.D. Selengut, C.J.A. Sigrist, M. Thimma, P.D. Thomas, F. Valentin, D. Wilson, C.H. Wu, Corin Yeats, InterPro: the integrative protein signature database, Nucleic Acids Res 37 (2009) D211–D215.
[23] K.J. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129.
[24] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[25] M.S. Islam, J.A. Saito, E.M. Emdad, B. Ahmed, M.M. Islam, A. Halim, Q.M.M. Hossen, M.Z. Hossain, R. Ahmed, M.S. Hossain, S.M.T. Kabir, M.S.A. Khan, M.M. Khan, R. Hasan, N. Aktar, U. Honi, R. Islam, M.M. Rashid, X.H. Wan, S.B. Hou, T. Haque, M.S. Azam, M.M. Moosa, S.M. Elias, A.M.M. Hasan, N. Mahmood, M. Shafiuddin, S. Shahid, N.S. Shommu, S. Jahan, S. Roy, A. Chowdhury, A.I. Akhand, G.M. Nisho, K.S. Uddin, T. Rabeya, S.M.E. Hoque, A.R. Snigdha, S. Mortoza, S.A. Matin, M.K. Islam, M.Z.H. Lashkar, M. Zaman, A. Yuryev, M.K. Uddin, M.S. Rahman, M.S. Haque, M.M. Alam, H. Khan, M. Alam,

Comparative genomics of two jute species and insight into fibre biogenesis, Nat. Plants 3 (2017) 16223.
[26] A. Roy, A. Bandyopadhyay, A.K. Mahapatra, S.K. Ghosh, N.K. Singh, K.C. Bansal, K.R. Koundal, T. Mohapatra, Evaluation of genetic diversity in jute (*Corchorus* species) using STMS, ISSR and RAPD markers, Plant Breed. 125 (2006) 292–297.
[27] J.I. Mir, P.G. Karmakar, S. Chattopadhyay, S.K. Chaudhury, S.K. Ghosh, A. Roy, SSR and RAPD profile based grouping of selected jute germplasm with respect to fibre fineness trait, J. Plant Biochem. Biotechnol. 17 (2008) 29–35.
[28] R.R. Mir, S. Banerjee, M. Das, V. Gupta, A.K. Tyagi, M.K. Sinha, H.S. Balyan, P.K. Gupta, Development and characterization of large-scale simple sequence repeats in jute, Crop Sci. 49 (2009) 1687–1694.
[29] Y.J. Park, J.K. Lee, N.S. Kim, Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops, Molecules 14 (2009) 4546–4569.
[30] B.D. Singh, A.K. Singh, Marker-Assisted Plant Breeding: Principles and Practices, Springer, New Delhi, India, 2015.
[31] R.K. Varshney, A. Graner, M.E. Sorrells, Genic microsatellite markers in plants: features and applications, Trends Biotechnol. 23 (2005) 48–55.
[32] J.R. Ellis, J.M. Burke, EST-SSRs as a resource for population genetic analyses, Heredity 99 (2007) 125–132.
[33] L.W. Zhang, R. Ming, J.S. Zhang, A.F. Tao, P.P. Fang, J.M. Qi, De novo transcriptome sequence and identification of major bast-related genes involved in cellulose biosynthesis in jute (*Corchorus capsularis* L.), BMC Genomics 16 (2015) 1062.
[34] L.W. Zhang, X.B. Wan, J.T. Xu, L.H. Lin, J.M. Qi, De novo assembly of kenaf (*Hibiscus cannabinus*) transcriptome using Illumina sequencing for gene discovery and marker identification, Mol. Breed. 35 (2015) 192.
[35] T.M. Liu, S.Y. Zhu, L.L. Fu, Q.M. Tang, Y.T. Yu, P. Chen, M.B. Luan, C.B. Wang, S.W. Tang, Development and characterization of 1,827 expressed sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud), PLoS One 8 (2013) e60346.
[36] S. Cloutier, Z.X. Niu, R. Datla, S. Duguid, Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.), Theor. Appl. Genet. 119 (2009) 53–63.
[37] H. Ellegren, Microsatellites: simple sequences with complex evolution, Nat. Rev. Genet. 5 (2004) 435–445.
[38] H. Sonah, R.K. Deshmukh, A. Sharma, V.P. Singh, D.K. Gupta, R.N. Gacche, J.C. Rana, N.K. Singh, T.R. Sharma, Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*, PLoS One 6 (2011), e21298. .
[39] L.C. Emebiri, An EST-SSR marker tightly linked to the barley male sterility gene (*msg6*) located on chromosome 6H, J. Hered. 101 (2010) 769–774.
[40] V. Diola, M.H.P. Barbosa, C.F.M. Veiga, E.C. Fernandes, Molecular markers EST-SSRs for genotype-phenotype association in sugarcane, Sugar Tech 16 (2014) 241–249.
[41] J.I. Mir, A. Roy, S.K. Ghosh, P.G. Karmakar, Development of linkage map in F_2 population of selected parents with respect to *Macrophomina phaseolina* resistance trait using screened polymorphic RAPD and developed SCAR markers of jute, Arch. Phytopathol. Plant Prot. 44 (2011) 671–683.
[42] X.B. Li, F. Jin, L. Jin, A. Jackson, C. Huang, K.H. Li, X.L. Shu, Development of *Cymbidium ensifolium* genic-SSR markers and their utility in genetic diversity and population structure analysis in cymbidiums, BMC Genet 15 (2014) 124.
[43] Z. Wang, G.H. Yu, B.B. Shi, X.M. Wang, H.P. Qiang, H.W. Gao, Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and *in silico* mapping onto the *M. truncatula* genome, PLoS One 9 (2014) e92029.
[44] A.L. Westman, S. Kresovich, The potential of cross taxa simple sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop brassicas, Theor. Appl. Genet. 96 (1998) 272–278.
[45] F. Leigh, V. Lea, J. Law, P. Wolters, W. Powell, P. Donini, Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat, Euphytica 133 (2003) 359–366.
[46] M.F. Wen, H.Y. Wang, Z.Q. Xia, M.L. Zou, C. Lu, W.Q. Wang, Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L, BMC Res. Notes 3 (2010) 42.
[47] N.C. Ting, N.M. Zaki, R. Rosli, E.T. Low, M. Ithnin, S.C. Cheah, S.G. Tan, R. Singh, SSR mining in oil palm EST database: application in oil palm germplasm diversity studies, J. Genet. 89 (2010) 135–145.
[48] C. Turchetto, A.L. Segatto, J. Beduschi, S.L. Bonatto, L.B. Freitas, Genetic differentiation and hybrid identification using microsatellite markers in closely related wild species, AoB Plants 7 (2015) plv084.
[49] D. Doddamani, M.A. Katta, A.W. Khan, G. Agarwal, T.M. Shah, R.K. Varshney, CicArMiSatDB: the chickpea microsatellite database, BMC Bioinformatics 15 (2014) 212.
[50] V. Sarika, M.A. Arora, A. Iquebal, D. Rai, Kumar, PIPEMicroDB: Microsatellite Database and Primer Generation Tool for Pigeonpea Genome, Database 2013, 2013 bas054.
[51] M.A. Iquebal, Sarika, V. Arora, N. Verma, A. Rai, D. Kumar, First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification, BMC Plant Biol 13 (2013) 197.
[52] A. Blenda, J. Scheffler, B. Scheffler, M. Palmer, J.M. Lacape, J.Z. Yu, C. Jesudurail, S. Jung, S. Muthukumar, P. Yellambalase, S. Ficklin, M. Staton, R. Eshelman, M. Ulloa, S. Saha, B. Burr, S.L. Liu, T.Z. Zhang, D.Q. Fang, A. Pepper, S. Kumpatla, J. Jacobs, J. Tomkins, R. Cantrell, D. Main, CMD: a cotton microsatellite database resource for *Gossypium* genomics, BMC Genomics 7 (2006) 132.