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