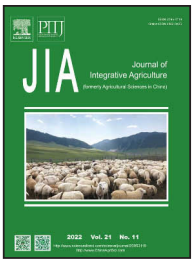
Journal of Integrative Agriculture 2022, 21(11): 3230–3244

Available online at www.sciencedirect.com

ScienceDirect



RESEARCH ARTICLE

Genome-wide development of interspecific microsatellite markers for *Saccharum officinarum* and *Saccharum spontaneum*

LIU Lei[1], WANG Heng-bo[1], LI Yi-han[1, 2], CHEN Shu-qi[1], WU Ming-xing[1], DOU Mei-jie[1], QI Yi-yin[1], FANG Jing-ping[3], ZHANG Ji-sen[1, 2]

[1] *Center for Genomics and Biotechnology/Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology/Key Laboratory of Genetics, Fujian Agriculture and Forestry University, Fuzhou 350002, P.R.China*

[2] *State Key Laboratory for Conservation and Utilization of Subtropical Agro-Biological Resources & Guangxi Key Laboratory for Sugarcane Biology, Guangxi University, Nanning 530004, P.R.China*

[3] *College of Life Sciences, Fujian Normal University, Fuzhou 350007, P.R.China*

Abstract

Sugarcane has a large, complex, polyploid genome that has hindered the progress of genomic research and molecular marker-assisted selection. The user-friendly SSR markers have attracted considerable attention owing to their ideal genetic attributes. However, these markers were not characterized and developed at the genome-wide scale due to the previously lacking high-quality chromosome-level assembled sugarcane genomes. In this present study, 744 305 and 361 638 candidate SSRs were identified from the genomes of *S. officinarum* and *S. spontaneum*, respectively. We verified the reliability of the predicted SSRs by using 1 200 interspecific SSR primer pairs to detect polymorphisms among 11 representative accessions of *Saccharum*, including *S. spontaneum*, *S. officinarum*, *S. robustum*, and modern sugarcane hybrid. The results showed that 660 SSR markers displayed interspecific polymorphisms among these accessions. Furthermore, 100 SSRs were randomly selected to detect the genetic diversity for 39 representative *Saccharum* accessions. A total of 320 alleles were generated using 100 polymorphic primers, with each marker ranging from two to seven alleles. The genetic diversity analysis revealed that these accessions were distributed in four main groups, including group I (14 *S. spontaneum* accessions), group II (two *S. officinarum* accessions), group III (18 modern sugarcane hybrid accessions), and group IV (five *S. robustum* accessions). Experimental verification supported the reliability of the SSR markers based on genome-wide predictions. The development of a large number of SSR markers based on wet experiments is valuable for genetic studies, including genetic linkage maps, comparative genome analysis, genome-wide association studies, and marker-assisted selection in *Saccharum*.

Keywords: *Saccharum*, genome-wide, simple repeat sequences (SSR), microsatellite, molecular markers

Received 28 January, 2022 Accepted 14 July, 2022
Correspondence FANG Jing-ping, E-mail: jinphia@163.com; ZHANG Ji-sen, E-mail: zjisen@126.com

doi: 10.1016/j.jia.2022.08.129

1. Introduction

Sugarcane (*Saccharum* spp. hybrid) taxonomically belongs to the genus *Saccharum*, which is placed under the Poaceae family of Andropogoneae tribe and is an important commercial worldwide, providing sustainable

economic growth and food security for tropical and subtropical regions of the world (Brand-Miller and Barclay 2017). Around 80% of sugar and 50% of alcohol are produced from sugarcane in the world (Waclawovsky *et al.* 2010). The *Saccharum* genus contains six main inter-breeding species: the two wild species are *S. spontaneum and S. robustum*, and the four cultivated species are *S. officinarum*, *S. sinense*, *S. barberi*, and *S. edule* (Hunsigi 1994). Dutch breeders in Java made the initial inter-specific crosses between noble cane *S. officinarum* (female) and wild cane *S. spontaneum* (male), followed by "Nobilization" of the hybrids by repeated backcrossing to *S. officinarum*. They thus obtained the disease and stress resistance ability of *S. spontaneum* while recovering the high biomass and high sugar content of *S. officinarum* (Brandes and Sartoris 1936). Modern sugarcane cultivars (2n=100–130) are genetically unstable, highly polyploid, heterozygous, and aneuploid with multi-specific origins and giant genome sizes (~10 Gb). They are inter-specific hybrids, with 70–80% of the chromosomes derived from *S. officinarum* (Souza *et al.* 2013), 10–15% from *S. spontaneum*, and the remaining 5–10% from interspecific recombination (D'Hont *et al.* 1996). All commercial sugarcane breeding populations worldwide share a narrow genetic background and high linkage disequilibrium (LD) due to the limited number of original parental lines involved in sugarcane crosses (Arruda 2012). The genetic complexity, narrow genetic base, and huge genome size of sugarcane create many challenges for breeding. Information about the genetic background of wild progenitors of modern sugarcane cultivars is a vital resource for sugarcane improvement.

In contrast to the labor-intensive and time-consuming procedures of conventional breeding, molecular markers can greatly accelerate the breeding progress by efficiently selecting key agronomic and quality traits. However, sugarcane marker-assisted breeding aimed at further enhancing sugarcane sucrose and biomass production have largely been challenged by the lack of trait-associated makers. For decades molecular markers have had numerous applications allowing important advances in many fields, including cultivar identification (Cordeiro *et al.* 2003), marker-assisted breeding (Pan *et al.* 2006), molecular fingerprint (Pang *et al.* 2019), genetic population and diversity analysis (Bertani *et al.* 2021), genetic mapping (You *et al.* 2019), and the fine mapping and cloning of genes (Wang *et al.* 2017). In sugarcane, various molecular markers are being developed and employed to characterize sugarcane germplasm, including amplified fragment length polymorphisms (AFLPs) (Besse *et al.* 1998; Aitken *et al.* 2006), restriction fragment length polymorphisms (RFLPs) (Creste *et al.* 2014), random amplification polymorphism DNAs (RAPDs) (Da Silva *et al.* 2008; Nawaz *et al.* 2010; Ramakrishnan *et al.* 2016), simple sequence repeats (SSRs) (Xu *et al.* 1999; Aitken *et al.* 2005; Cai *et al.* 2005; Andru *et al.* 2011; Nayak *et al.* 2014; Zawedde *et al.* 2015), inter-simple sequence repeats (ISSRs) (Almeida *et al.* 2009; Devarumath *et al.* 2012), and single nucleotide polymorphisms (SNPs) (Yang *et al.* 2018). Among these molecular markers, the user-friendly SSR markers have attracted much attention owing to their ideal genetic attributes (Kantety *et al.* 2002), such as richness, high-polymorphism levels, multi-allelic nature, and co-dominant inheritance (Amar *et al.* 2011; Baert-Desurmont *et al.* 2016) and enjoyed great success in sugarcane genetic diversity research (Wang *et al.* 2020), genetic mapping (Pan and Liu 2010), and marker-assisted selection (MAS) (Olatoye *et al.* 2019). Previous studies focused more on the development of expressed sequence tag (EST)-based markers (Oliveira *et al.* 2007) than genomic ones due to the lack of genomic sequence data. Although EST-SSRs are useful for genetic analysis, they have disadvantages. Specifically, they have relatively low numbers of polymorphisms and high concentrations in gene-rich regions of the genome, which may limit their usage, especially for the construction of linkage maps (Temnykh *et al.* 2000). In addition, considering the high levels of polyploidy in the sugarcane genome, more genomic markers are needed for the advancement in genetics, genomics, and breeding of sugarcane.

Identifying SSRs from available genome sequences has proved to be a robust, rapid, and widely applicable strategy (Zhang *et al.* 2012; Liu C *et al.* 2016; Cui *et al.* 2017). The advances in sequencing technologies have taken a significant leap in the last decade. Next generation sequencing (NGS) and third-generation DNA sequencing technologies have enabled the completion of a rapidly growing number of whole-genome sequencing projects. More and more genome sequence information can be obtained from public databases, promoting the development and application of SSR markers (Baraket *et al.* 2011; Giovannini *et al.* 2012; Potts *et al.* 2012; Emanuelli *et al.* 2013; Hajmansoor *et al.* 2013). Recently, a haploid *S. spontaneum*, AP85-441, was assembled into 32 pseudo-chromosomes comprising eight homologous groups of four members each (Zhang *et al.* 2018). Moreover, the genome of *S. officinarum*, Louisiana Purple (LA-purple), was assembled into 80 pseudo-chromosomes in 10 homologous groups (unpublished data in our lab), which provided an opportunity to screen for SSRs across the sugarcane genomes of wild progenitors more thoroughly and efficiently.

The present study used the available dataset of sugarcane genome sequences to mine the SSRs

throughout sugarcane assembled genomes using bioinformatics methods (Zhang *et al.* 2018). We identified SSR motifs in a genome-wide manner and characterized these SSR motifs. We also attempted to evaluate genetic diversity and population structure among the *S. officinarum* and *S. spontaneum* using SSR markers. This is the first report on genome-wide interspecific SSR identification and interspecific SSR markers development in sugarcane. The inventory of SSR markers proposed in the present study could facilitate sugarcane genetic and breeding studies. Our study provides a large-scale SSR marker dataset for *Saccharum* that could be widely applied in cross-breeding identification, basic research for MAS, and interspecific cross-genetic map construction.

2. Materials and methods

2.1. Plant materials and DNA extraction

In the present study, the available sugarcane genomes of *S. officinarum* (LA-purple) and *S. spontaneum* (AP85-441, 1n=4x=32) (Zhang *et al.* 2018) were used for SSR identification and validation. Eleven sugarcane elite accessions (1, Badila; 2, Black Cheribon; 3, Crystalina; 4, Loethers; 5, LA-purple; 6, SES-208; 7, FU89-1-1; 8, Guangdong21; 9, SC79-2011; 10, POJ-2878; 11, ROC-22) were used for SSR validation. Besides, 39 sugarcane accessions, including *S. officinarum* (nine accessions), *S. spontaneum* (14 accessions), *S. robustum* (five accessions), and modern sugarcane cultivars (11 accessions), were selected for the diversity analysis (Table 1). Total genomic DNA was isolated from fresh leaves by an optimized CTAB method (Yao *et al.* 2005). The concentration and quality of all DNA samples were determined using the NanoDrop spectrophotometer (ND-1000 version 3.1.1, USA), and all DNA samples were diluted to 25 ng L^{-1} to enable subsequent PCR-based *in vitro* amplification.

2.2. SSR marker identification and primer development

The LA-purple genome consists of 80 pseudo-chromosomes with 10 homologous groups of eight members each (http://sugarcane.zhangjisenlab.cn/sgd/html/download.html). The haploid *S. spontaneum*, AP85-441, is an assembly of 32 pseudo-chromosomes comprised of eight homologous groups of four members each.

The genome sequences were used to identify SSR loci using a Perl program MISA, MIcroSAtellite identification tool (Thiel *et al.* 2003) available at the

Table 1 Sugarcane (*Saccharum* spp.) accessions used in the study[1)]

No.	Accession	Species	No.	Accession	Species
1	●2013-8	*S. spontaneum*	23	●KARA KARA WA	*S. officinarum*
2	●2015-68	*S. spontaneum*	24	●Manjar	*S. officinarum*
3	●FJ79-II-9	*S. spontaneum*	25	●Keong Java	*S. officinarum*
4	●Guangdong30	*S. spontaneum*	26	●NC20	*S. officinarum*
5	●GZ78-I-11	*S. spontaneum*	27	▲●LA-purple	*S. officinarum*
6	●HNLD-1	*S. spontaneum*	28	●CZ-2	*S. hybrid*
7	●Laos2	*S. spontaneum*	29	●ROC1	*S. hybrid*
8	●HN2012-124	*S. spontaneum*	30	▲●P0J2878	*S. hybrid*
9	●SC79-I-1	*S. spontaneum*	31	▲●ROC22	*S. hybrid*
10	●Vietnam3	*S. spontaneum*	32	●Co281	*S. hybrid*
11	●YN82-25	*S. spontaneum*	33	●F134	*S. hybrid*
12	●YN82-91	*S. spontaneum*	34	●Co419	*S. hybrid*
13	●YN-4	*S. spontaneum*	35	●Co290	*S. hybrid*
14	▲FU89-1-1	*S. spontaneum*	36	●Co213	*S. hybrid*
15	▲Guangdong21	*S. spontaneum*	37	●R570	*S. hybrid*
16	▲SC79-2011	*S. spontaneum*	38	●co1001	*S. hybrid*
17	▲●SES-208	*S. spontaneum*	39	●YN-*Robustum*	*S. robustum*
18	▲Loethers	*S. officinarum*	40	●51NG3	*S. robustum*
19	▲●Crystalina	*S. officinarum*	41	●51NG63	*S. robustum*
20	▲●Black Cheribon	*S. officinarum*	42	●FJ-*Robustum*	*S. robustum*
21	▲●Badila	*S. officinarum*	43	●57NG208	*S. robustum*
22	●Cana Blanca	*S. officinarum*			

[1)] ▲ indicated the 11 accessions of the genus *Saccharum* for validation of SSRs, ● indicated the 39 accessions for the genetic diversity analysis.

website (http://pgrc.ipkgatersleben.de/misa/misa). The SSR loci containing perfect repeat units of 1–6 nucleotides only were considered. The criteria for identifying the minimum repetition time of mono-, di-, tri-, tetra-, penta-and hexa-nucleotide repeats were 10, 7, 6, 5, 4, and 4, respectively. Compound microsatellites were defined as those with the interval between two repeats motifs shorter than 100 nt (Wang *et al.* 2015; Deng *et al.* 2016; Liu F *et al.* 2016).

To find the SSR loci with unique flanking sequences, the BLASTN (Jones *et al.* 2002) search of the genomic sequence was carried out using the repetitive motif and the flanking sequences on both sides of the 150 bp (*E*-value cut-off of 1e–5) and filtered with 100% identity and 100% alignment length of the flanking sequences. We identified SSR loci with a single hit for marker development. Primer 3 (https://sourceforge.net/projects/primer3/) was used to design PCR primers for all candidate loci to amplify 100–300 bp products with 40–60% GC content. Primer sizes ranged from 18 to 22 bp with an optimum size of 20 bp, and the annealing temperature was set at 60°C.

The precision of the designed primers was estimated through the electronic PCR (e-PCR) program. The genome sequences of *S. officinarum* and *S. spontaneum* were used as templates for primers to blast with the following criteria: 100 bp margin, 100–400 bp product sizes, and 1 bp mismatch. Primers obtained from *S. officinarum* species were compared to those obtained from *S. spontaneum* species and *vice versa*. Through e-PCR, the size of the PCR product for each pair of primers in each template was scored. If the size and position of PCR products were unique in two *Saccharum* species, this combination of primers was recognized as the interspecific SSR primers for *S. officinarum* and *S. spontaneum*.

2.3. Experimental validations of SSR markers

SSR length is an important factor for SSR marker development. According to the length of SSRs, SSRs can be divided into two groups. Class I SSRs (hypervariable markers) are composed of those with ≥20 bp repeats in length, while Class II SSRs (potentially variable markers) contain repeats <20 bp in length (Temnykh *et al.* 2001). Only the Class I SSRs and non-mononucleotides were selected for final genome-wide SSR marker development in *Saccharum* to improve their utilization efficiency.

Mononucleotides were filtered due to the difficulty in distinguishing bona fide microsatellites from sequencing or assembly error and because (A/T)n repeats in the coding region may be confused with polyadenylation tracks. SSR markers derived from longer perfect repeats (SSRs≥20 nucleotides in length) have been demonstrated to show high levels of polymorphisms from the experimental data in many organisms such as human (Iwasaki *et al.* 1992) and rice (Cho *et al.* 2000).

To validate the e-PCR results, PCR amplification was performed in a 20-μL PCR volume which was composed of 0.5 μL template DNA (25 ng μL^{-1}), 0.5 μL of each primer, 10 μL 2× *Taq* Master Mix (Boshang Biological, Shanghai, China), and 8.5 μL ddH_2O. PCR amplification was completed using the following touchdown program: 94°C for 90 s for initial denaturation, then 34 cycles at 94°C (20 s)/59.5°C (20 s)/72°C (30 s), and the last extension step was 72°C for 5 min. The PCR products were identified by polyacrylamide gel electrophoresis. The fragment sizes of the PCR products were examined by identification with the DL50 DNA ladder (TaKaRa, Dalian, China).

2.4. Cloning and sequencing of PCR products

PCR products of SSRs that showed amplification in all sugarcane genotypes were excised from the 5% agarose gel through Gel Extraction Kit (OMEGA Bio-Tek, USA), cloned into the TA cloning vector (pEASYTM-T5 Zero, Transgen Biotech, Beijing, China), and sequenced to examine their conservation and cross-species transferability. The final edited sequences belonging to each locus were compared with the respective original SSR sequence using DNAMAN (Lynnon BioSoft, Vaudreuil, Quebec, Canada) for ascertaining the target domain/SSR conservation (Battilana *et al.* 2009).

3. Results

3.1. SSR motif content in the sugarcane genome

The genome sequences of *S. officinarum* and *S. spontaneum* were used for searching the SSRs with various types of required repeating motifs, from single nucleotides to hexanucleotides. The results of MISA showed that the assembled genome size of *S. officinarum* was 6 804 994 kb, in which a total of 744 305 microsatellites were identified with an average interval length of 9.14 kb between SSR loci (Table 2); the assembled genome size of *S. spontaneum* was 3 140 615 kb, in which a total of 361 638 microsatellites were identified with an average interval length 8.68 kb for each SSR loci (Table 2). Based on the assembled genomes of *S. officinarum* and *S. spontaneum*, the distribution of the chromosome of microsatellites was investigated. The distribution of each chromosome showed that

Table 2 Distribution of SSR loci in the genomes of *Saccharum officinarum* LA-purple and *Saccharum spontaneum* AP85-441

Repeat unit type	*S. officinarum* LA-purple			*S. spontaneum* AP85-441		
	SSR loci	Ratio (%)	Motif	SSR loci	Ratio (%)	Motif
Mono-	414 310	55.66	4	210 400	58.18	4
Di-	146 393	19.67	12	64 766	17.91	12
Tri-	73 621	9.89	60	33 323	9.21	54
Tetra-	16 992	2.28	236	8 746	2.42	219
Penta-	24 102	3.24	958	11 784	3.26	873
Hexa-	19 637	2.64	2 353	10 678	2.95	1 983
c[1)]	49 250	6.62	39 738	21 941	6.07	19 006
Total	744 305	100	43 361	361 638	100	22 151

[1)] c, compound.

microsatellite markers are uniformly distributed throughout the genome (Fig. 1; Appendices A and B).

Of these SSRs, mono-nucleotide repeats accounted for 55.66 and 58.18%, which were the largest proportion, followed by di-nucleotide (19.67 and 17.91%) and tetra-nucleotide repeats (9.89 and 9.21%) in the two *Saccharum* species, respectively (Table 2). The distribution of SSR motifs in *S. officinarum* was consistent with *S. spontaneum*. The dominance of di-nucleotides was attributed to an overrepresentation of AT/TA motifs. TGT/ACA motifs were dominant in tri-nucleotide SSRs, and AAAT/TTTA and AAAT/TTAA motifs were dominant in tetra-nucleotide SSRs (Fig. 2). The total repeat number of T/A was significantly greater than that of C/G in all of the genome-wide SSRs. There were many types of penta- and hexa-nucleotide SSRs, each with low frequencies, ranging from 4.84 to 13.25% across two *Saccharum* species genomes (Xu *et al.* 2013). The numbers of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide, and compound-type motifs are listed in Table 2. The frequencies of SSR motifs with regard to the repeat number were examined (Fig. 3). We found that for each motif type, the SSR abundance decreased dramatically with an increased number of repeat units. The most significant rate of decrease was for the hexanucleotide motifs.

3.2. Polymorphisms of the interspecific SSR markers between *S. officinarum* and *S. spontaneum*

Primers were designed for the candidate SSR loci using Primer3 Software. Amplification primer pairs were designed for 63 414 and 68 214 SSRs in *S. officinarum* and *S. spontaneum*, respectively. To test the amplification efficiency, specificity, and polymorphisms of the newly developed SSR primers, the genome sequences of *S. officinarum* and *S. spontaneum* were used as templates for e-PCR analysis. After SSR detection, a total of 41 725 (65.79%), 5 292 (8.35%), 3 260 (5.14%), and 13 137 (20.72%) of these primers could generate 1, 2, 3, >3 loci in *S. officinarum*, respectively. A total number of 42 888 (62.87%), 8 549 (12.53%), 6 140 (9.10%), and

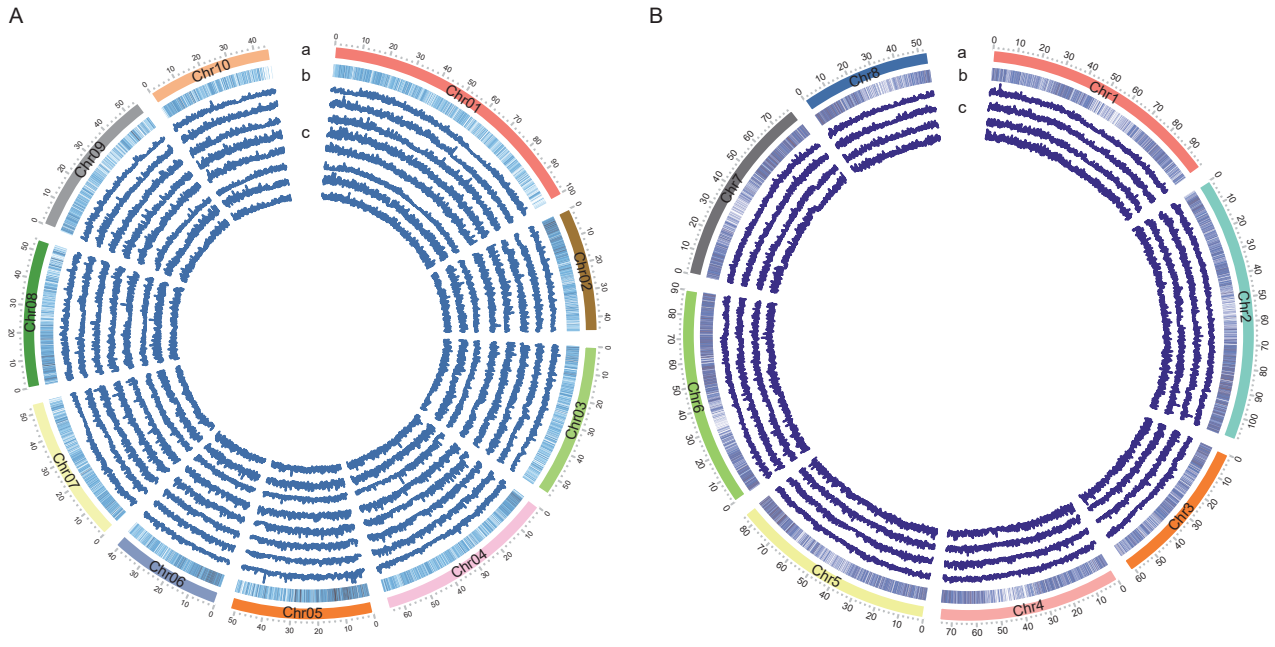


Fig. 1 Genome-wide distribution of SSR markers on the chromosomes of the *Saccharum officinarum* (A) and *Saccharum spontaneum* (B) genome. a, haploid chromosome; b, SSR distribution; c, SSR density on each homologous chromosome.

Fig. 2 Percentage of repeats in different types of genomic SSR in the genome of *Saccharum officinarum* LA-purple and *Saccharum spontaneum* AP85-441. A1–F1 indicated the different motif types in the *S. officinarum* LA-purple, and A2–F2 indicated the different motif types in the *S. spontaneum* AP85-441.

10 636 (15.59%) of these primers could generate 1, 2, 3, >3 loci in *S. spontaneum*, respectively. The amplification efficiency of these primers was 100%.

By comparing SSR sequences between the two genomes, a new dataset containing 1 807 SSR markers was established. These SSRs had unique sizes and specific chromosomal positions for PCR products in both *Saccharum* genomes, indicating that they are valuable as genetic markers (Fig. 1). The markers are evenly distributed on each chromosome of sugarcane and can be used as representative candidate markers for subsequent experimental analysis. Among these markers, SSRs with di-, tri-, tetra-, penta-, and hexa-nucleotide motifs accounted for 46.21, 15.71, 5.53, 9.07, 4.76, and 18.71%, respectively (Fig. 4).

3.3. Development of SSR markers for interspecific identification of *Saccharum*

Finally, 1 200 SSR markers were selected to check their amplification efficiency in the *Saccharum* genus (*S. officinarum*, *S. spontaneum*, and sugarcane modern

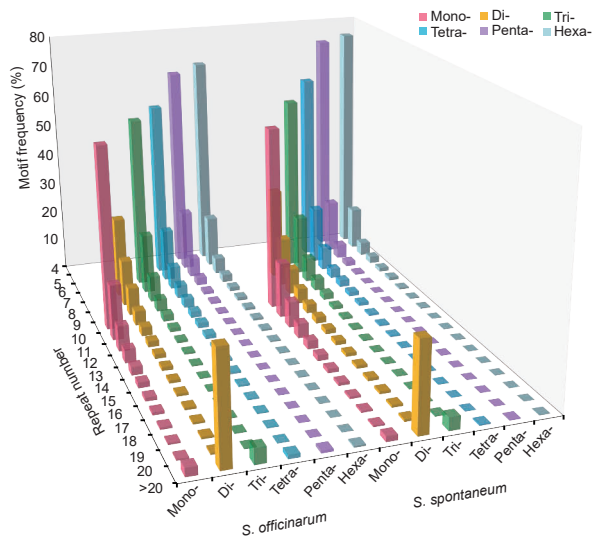


Fig. 3 The frequency of motif distribution of mono- to hexa-nucleotide motif types with varied repeat numbers (from 4 to 20) in the genomic sequences of *Saccharum officinarum* and *Saccharum spontaneum*.

cultivars). We randomly selected 15 pairs of primers from each of the 10 homologous groups of *S. officinarum* genome for PCR amplification. Initially, the effectiveness of these primer pairs was estimated for the *Saccharum* species, including *S. officinarum* (five accessions), *S. spontaneum* (four accessions), and modern sugarcane cultivars (two accessions) (Table 1) using polyacrylamide gel electrophoresis (PAGE). A total of 1 152 (96%) SSR markers could amplify at least one band, of which 660 showed interspecific polymorphisms among the *Saccharum* plant materials (Appendix C). These SSRs were co-dominant markers, and their main bands were clear, stable, and easily distinguished in each accession, while the remaining markers were dominant and displayed a weak band at the expected position (Fig. 5-A). One of the reasons might be that the e-PCR would underestimate the complexity of sugarcane and there were more multiple loci amplifications in experimental PCR than e-PCR. The incomplete *Saccharum* genome sequences used in this study might be responsible for the underestimation of the genome complexity.

To evaluate polymorphisms in newly developed SSR markers, 100 validated microsatellite markers were amplified in the *Saccharum* genus, including 39 accessions: nine *S. officinarum*, 17 *S. spontaneum*, five *S. robustum*, and 11 modern sugarcane cultivars (Table 1; Appendix D). All the markers displayed interspecific polymorphisms in these four sugarcane species (Fig. 5-B). The highly successful amplification rate in *Saccharum* demonstrated that the newly developed 1 807 microsatellite markers for the *Saccharum* genome can be used for species identification and extend the sugarcane marker pool for functional genomics and molecular breeding.

To determine the effectiveness of the SSR markers, sequence similarities of the PCR products were compared among four *Saccharum* species. The bands presented in all species were sequenced. All identical sequences of PCR bands had high similarities (85%) across *S. officinarum*, *S. spontaneum*, *S. robustum*, and modern sugarcane cultivars. Marker polymorphisms among the four species were mainly caused by the variation in the number of repeats of SSR motifs. Sequence alignment showed that all bands contained the (CCT)n and (TTTA)n repeat motifs. We found that the same primers were consistent in the amplification of different materials, the accession of *S. spontaneum* YN-4 were observed to have six CCT, and four CCT were found in *S. officinarum* NC20. The sizes of the fragments were also different when amplified by the same primer in the same material. In *S. hybrid* cultivar Co419, the two bands were found to have four and six tri-nucleotide repeat motif CAG, respectively (Fig. 5-C). These sequence files were

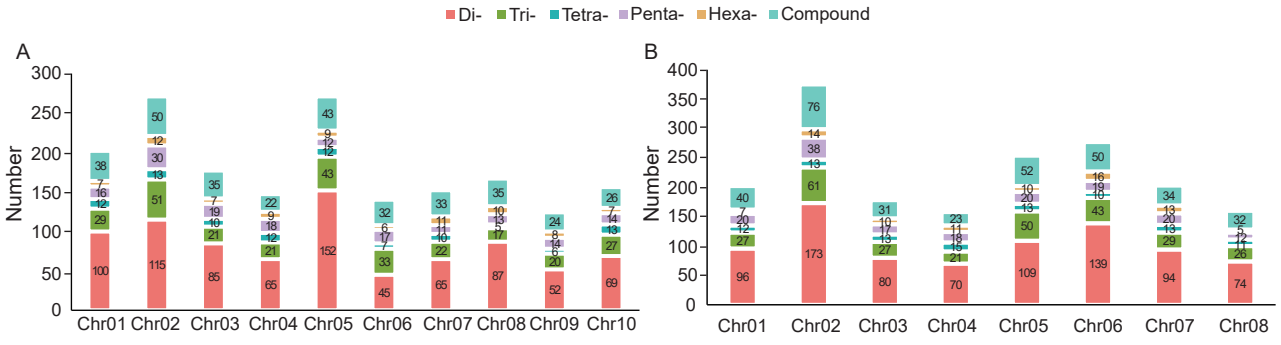


Fig. 4 The distribution types of candidate interspecific markers on the chromosomes of *Saccharum officinarum* LA-purple (A) and *Saccharum spontaneum* AP85-441(B).

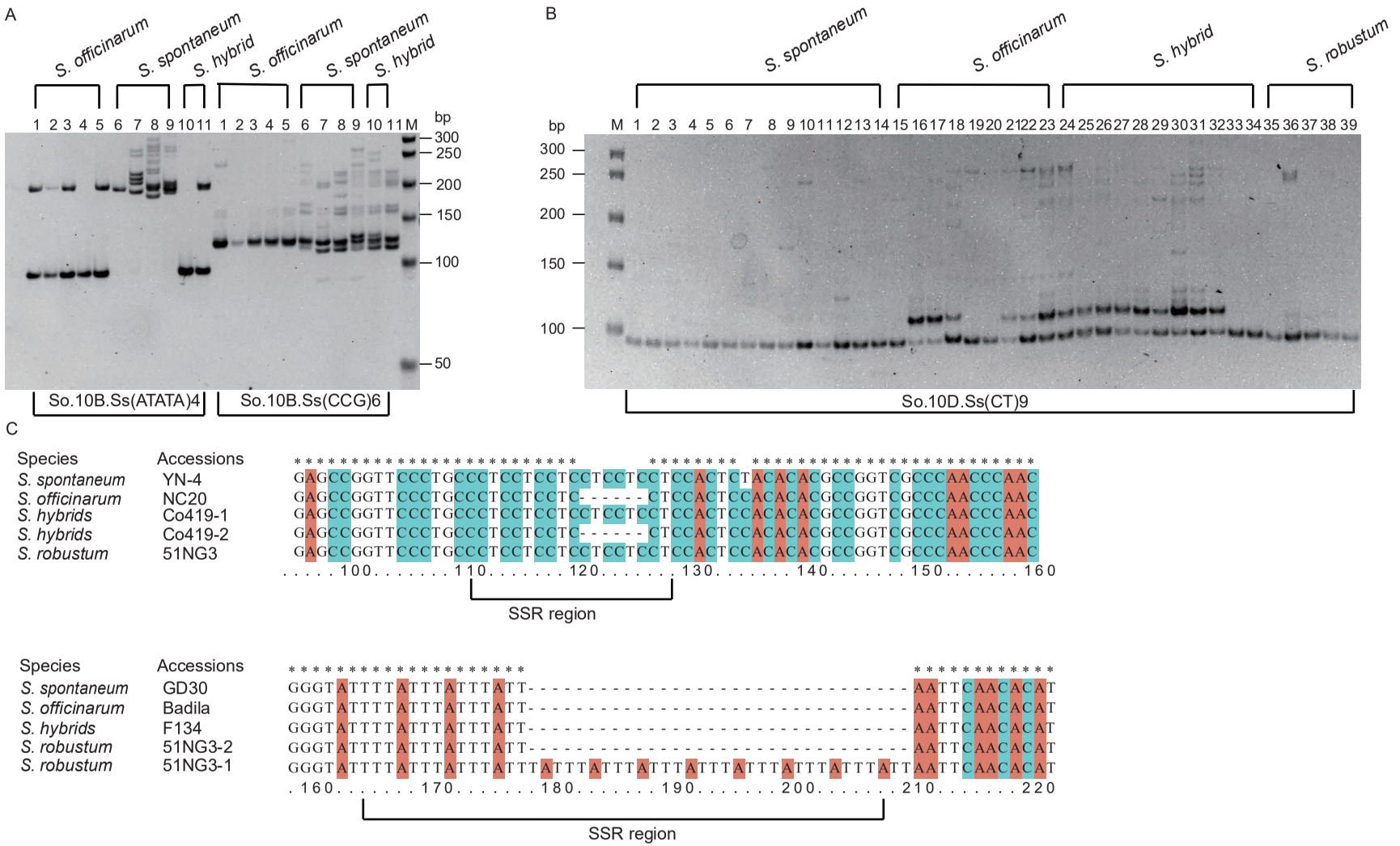


Fig. 5 Representative SSR gel electrophoresis and the sequence alignment of SSRs. A, SSRs verification profile of 11 accessions from *Saccharum* detected using polyacrylamide gel electrophoresis. 1, Badila; 2, Black Cheribon; 3, Crystalina; 4, Loethers; 5, LA-purple; 6, SES-208; 7, FU89-1-1; 8, Guangdong21; 9, SC79-2011; 10, POJ-2878; 11, ROC-22. B, electrophoretic patterns of SSR primers So.10D.Ss(CT)9 in 39 tested *Saccharum* accessions. 1, 2013-8; 2, 2015-68; 3, FJ79-II-9; 4, Guangdong30; 5, GZ78-I-11; 6, HNLD-1; 7, Laos2; 8, HN2012-124; 9, SC79-I-1; 10, Vietnam3; 11, YN82-25; 12, YN82-91; 13, YN-4; 14, SES-208; 15, Crystalina; 16, Black Cheribon; 17, Badila; 18, Cana Blanca; 19, KARA KARA WA; 20, Manjar; 21, Keong Java; 22, NC20; 23, LA-purple; 24, CZ-2; 25, ROC1; 26, P0J2878; 27, ROC22; 28, Co281; 29, F134; 30, Co419; 31, Co290; 32, Co213; 33, R570; 34, co1001; 35, YN-*Robustum*; 36, 51NG3; 37, 51NG63; 38, FJ-*Robustum*; 39, 57NG208. C, multiple sequence alignments of the size variant fragments amplified from four species of *S. spontaneum* (YN-4 and GD30), *S. officinarum* (NC20 and Badila), *S.* hybrids (Co419 and F134) and *S. robustum* (51NG3) for the SSR marker, the marked parts indicate the repeat motif sequences.

analyzed and unequivocally confirmed the cross-species conservation and transferability.

3.4. Genetic distance-based clustering analysis

Cluster analysis based on Nei's genetic similarity coefficient using the UPGMA algorithm is shown in Fig. 6. The 39 sugarcane accessions were divided into four groups (I, II, III, and IV) according to the clustering results. The similarity among all the accessions varied from 0.51 to 0.85, with an average value of 0.68. When the genetic similarity coefficient was 0.64, 39 sugarcane accessions could be grouped into four clusters. Cluster I comprised 14 accessions, which belonged to *S. spontaneum*. Cluster II and Cluster III grouped two and 18 accessions, respectively, together belonging to *S. officinarum* and cultivated sugarcane species. Cluster IV had five genotypes within *S. robustum*. The UPGMA dendrogram produced clusters separating *S. spontaneum* (Cluster I) and *S. officinarum*, *S. hybrid*, and *S. robustum* (Cluster II, III, and IV) into two distinct groups when the genetic similarity was 0.51. The results were highly consistent with the material source information and further indicated that the newly developed sugarcane SSRs can be used to analyze the genetic diversity of the *Saccharum* genus.

4. Discussion

Genetic analysis of sugarcane has been hampered because of its complex polyploid genome and lack of sufficiently informative markers. Thus, large numbers of useful polymorphic markers are required for germplasm diversity and analysis of the genetic relationship of sugarcane because of its large genome and complex genetic background. SSRs have increasingly become the marker of choice for these sorts of analyses as they have the advantage of simplicity, high abundance, ubiquity, high level of polymorphisms, stable co-dominance, and multiple-*allelism* among genomes (Fraser *et al.* 2004; Chen *et al.* 2006; Hirano *et al.* 2011). In previous studies, the *de novo* development of SSRs was a costly, labor-intensive, and time-consuming endeavor due to a lack of genome sequence information (Zane *et al.* 2010). The increasing availability of genomic sequence data has made it faster and less expensive to characterize and utilize genome-wide SSR markers from many species (Zhang *et al.* 2018). In recent years, considerable progress has been made in sugarcane genomics, including the sequencing of *S. spontaneum* AP85-441 (Zhang *et al.* 2018) and *S. officinarum* LA-purple (unpublished data in our lab), which facilitated successful efforts in gene family identification, genetic mapping, population genetic, and genome evolution analysis in sugarcane (Ma *et al.* 2020; Li *et al.* 2021; Zhang *et al.* 2021). In this paper, the newly assembled genomes of *S. officinarum* and *S. spontaneum* were used for large-scale genome-wide SSR mining and generated 744 305 and 361 638 candidate SSRs, respectively. The motifs

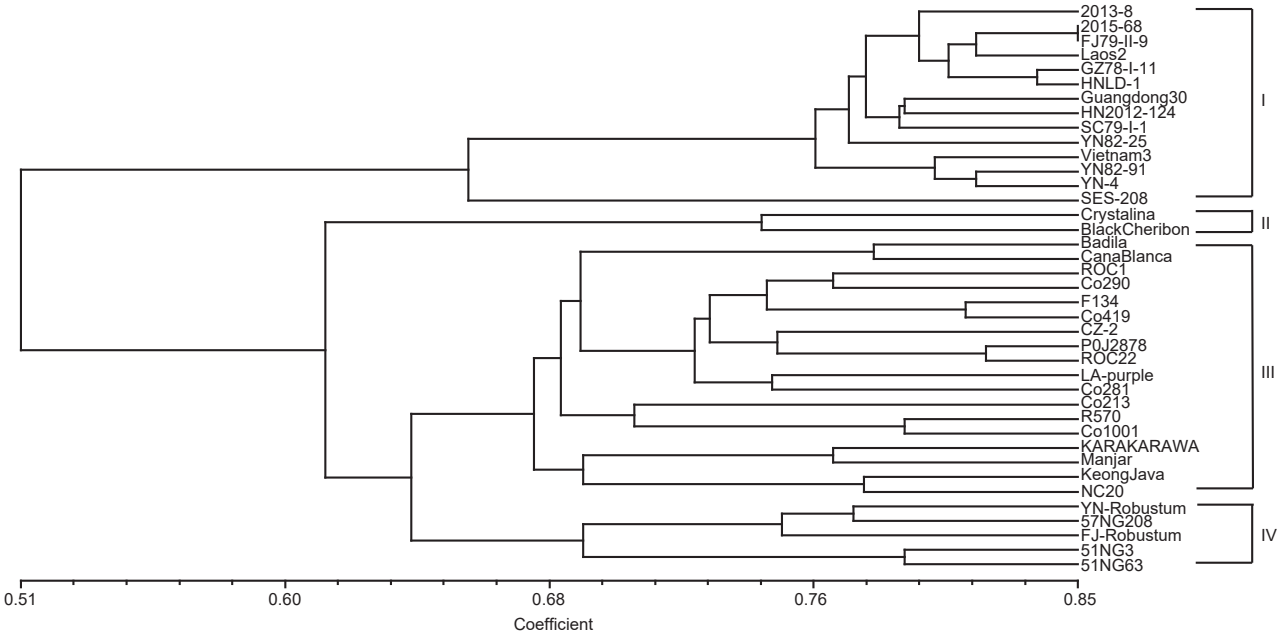


Fig. 6 Dendrogram for 39 sugarcane accessions derived from UPGMA cluster analysis based on 100 SSR markers. The 39 sugarcane accessions were grouped into four clades (I, II, III, and IV).

which have more than 7 nt were not considered in this paper because of their small proportion (Xu *et al.* 2017).

4.1. Frequency and characteristics of SSRs in *Saccharum* genomes

In the current study, the average SSR density was one SSR every 9.14 and 8.68 kb in *S. officinarum* and *S. spontaneum*. Based on the survey across other plant genomes, on average, one SSR was found every 1.14 kb in *Arabidopsis* (Lawson and Zhang 2006), 3.6 kb in rice (Zhang *et al.* 2007), 220 kb in sorghum (Iniguez-Luy *et al.* 2008), 15.48 kb in maize (Xu *et al.* 2013), 220 kb in sorghum (Iniguez-Luy *et al.* 2008), and 578 kb in wheat (Morgante *et al.* 2002). Compared with other plants, we found that the *Saccharum* species with large genome sizes have a moderate SSR frequency, which agreed well with the generally negative correlation between genome sizes and SSR frequencies in plants (Deguchi *et al.* 1996).

In sugarcane, the most common SSR types were mono-, di-, and tri-nucleotide repeats, which together account for 90% of all SSRs. Mono-, di-, and tri-nucleotides occupy 63, 16, and 11% of the total SSRs in *S. officinarum* and 62, 20, and 8% in *S. spontaneum* (Table 2), respectively. The frequency of SSRs within certain nucleotide compositions shows species and genome fraction specificity among eukaryotic genomes. The motifs containing A/T were more common than the motifs containing G/C (Table 2). A similar trend has also been observed in other plant genomes such as rice, sorghum, apple, and tobacco (Tong *et al.* 2015; Fu *et al.* 2016) that also exhibit a strong bias to AT-richness in SSR motifs, but this was not the trend in *Brachypodium distachyon* and maize (Liu *et al.* 2012). AT/TA was the most common repeat unit across sugarcane genomes, representing 56.71% in *S. officinarum* and 57.92% in *S. spontaneum*. These results were consistent with the analysis of sorghum, tobacco, and apple, but not other Poaceae plants including wheat (Yu *et al.* 2004), rice (Liu *et al.* 2012), and maize (Duan *et al.* 2006), in which AG motif were the most abundant. The TGT/ACA motif (16.04%) appeared the most frequently in sugarcane genomes. This AT-rich trend was similar to *Arabidopsis* (AAG/TTC) (Lawson and Zhang 2006), tobacco (AAC/TTG) (Lim *et al.* 2004), and apple (AAC/GTT) (He *et al.* 2011) but different from barley, wheat, maize, rice, sorghum, and rye (Cho *et al.* 2000; Thiel *et al.* 2003; Lv *et al.* 2019; Wang *et al.* 2019), in which CCG/GCC is the most prevalent triplet while ATT/TAA is the least. The strong bias towards AT-rich in motifs especially in 1–3 nucleotide motifs could be attributed to two possible reasons: first, methylated C residue is more easily transformed into T (Levinson and Gutman 1987); second, SSR sites are generated by the high rates of strand slippage during DNA replication (Halsted and Devlin 2001; Andrés and Bogdanowicz 2012), during which time, the SSR motif based on the A-T structure type requires less energy than the G-C type, making the A-T base-rich structure type easier to produce than the G-C type. However, with the increased nucleotide repetition motif from 3 to 6, the energy advantage of this hydrogen bond is relatively less.

4.2. Evaluation of genomic SSR markers in *Saccharum* species

SSR markers have been widely used for genetic analysis, high-throughput genotypic and map construction, and breeding in many crop species. Previous studies have used EST-derived markers in *Saccharum*, but only a limited number of genomic markers have been developed and evaluated (Da Silva 2001; Pinto *et al.* 2004, 2006; Aitken *et al.* 2005; Parida *et al.* 2009). The limitation of previously developed genomic SSR markers in trait dissection and discovery of candidate genes restricts their use in sugarcane genetic studies and molecular breeding. In this study, a total of 1 200 interspecific genomic SSR primer pairs were designed, and 1 152 (96%) primer pairs proved effective. Due to the self-complementary requirement for dimer formation, AT/TA was not used to develop markers in our analysis (Wang *et al.* 2011). Among these primers, 100 pairs of SSR primers with polymorphisms were used to analyze the SSR polymorphisms among 39 sugarcane materials (Appendix D), and the total number of amplified bands was 313. The genetic similarity indices among the 39 sugarcane germplasms varied between 0.51 and 0.85 with an average of 0.68, which reflects a substantial level of genetic diversity within the different *Saccharum* species and cultivated varieties that come from diverse geographic locations across the world. The results of clustering analysis showed that the sugarcane accessions of each group were derived from their respective progenitors and shared genetic intimacy with each other. SSR genetic similarity and cluster analysis of this study revealed that the offspring from the same cross combination had higher genetic similarity with each other, and the same series of sugarcane accessions mostly grouped in one cluster. The clustering results based on genetic similarity of SSR among sugarcane germplasms are largely consistent with the material source. For example, F134 (Co290×POJ2878) and Co419 (POJ2878×Co290) from the same parents had very high genetic similarity

coefficients, therefore being located in the same cluster.

However, some results of inheritance similarity and clustering analysis are irreconcilable with the sources of sugarcane materials, which may be closely related to the cross-breeding process of sugarcane or the genetic composition of sugarcane. Sugarcane is an allopolyploid plant with high heterozygosity, and cross-breeding, backcrossing, and self-crossing will lead to high-level separations in offspring, so the offspring from the same cross combination may have a large genetic difference, while the offspring from different combinations may also have a small genetic difference. In addition, genealogical records of some cultivars failed to reflect their genetic relationships due to the phenomenon of self-crossing, powdering, and seed mixing.

The cluster analysis revealed a distant genetic relationship among *S. spontaneum*, *S. officinarum*, cultivated sugarcane species, and *S. robustum*. One of the main reasons is the northern India origin of the diversity of *S. spontaneum*, where the lowest chromosome number of accessions of this species (2n=40) were found (Zhang *et al.* 2013). *Saccharum officinarum* likely evolved from *S. robustum* as they share the same center of origin and diversity in Papua New Guinea (Skinner *et al.* 1987). High genetic similarity between the *S. officinarum* clones and commercial cultivars indicates a large amount of introgression from *S. officinarum* into modern cultivars (Selvi *et al.* 2003; Singh *et al.* 2011; Chandra *et al.* 2013), confirming the consensus that modern sugarcane cultivars are aneuploidy, interspecific hybrids with chromosomes ranging from 2n=100 to 130, comprising 70–80% *S. officinarum*, 10–20% *S. spontaneum*, and 10% recombinant chromosomes between the two species (Daugrois *et al.* 1996).

The PCR-based DNA markers facilitate the determination of the genetic variability that occurs among conventional progenitor species and commercial cultivars (Singh *et al.* 2019). The genomic constitution of different plant genetic resources can be accurately and finely evaluated by determining the population structure with different DNA markers (Nachimuthu *et al.* 2015; Singh *et al.* 2017). Based on the advances in genomic resources, SSR markers are preferred to be applied in rapid analysis of the genetic structures of breeding germplasms (Liu C *et al.* 2016; Singh *et al.* 2019).

5. Conclusion

Our study is the first systematic and comprehensive analysis of the SSR markers from the sugarcane genome. We predicted 744 305 and 361 638 candidate SSRs in the genomes of *S. officinarum* and *S. spontaneum*, respectively. Out of 1 200 SSR markers, 660 displayed interspecific polymorphisms among the 11 representative accessions, proving the reliability of the examined SSRs. We also used 100 SSRs to detect the genetic diversity for 39 representative *Saccharum* accessions. Our study provides a valuable resource for the genetic analysis of sugarcane.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2021YFF1000101-5), the Science and Technology Planting Project of Guangdong Province, China (2019B020238001), the Natural Science Foundation of Fujian Province, China (2019J05066), the National Natural Science Foundation of China (41906096), and the Guangdong Laboratory for Lingnan Modern Agriculture and the State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, China.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendices associated with this paper are available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

References

Aitken K S, Jackson P A, McIntyre C L. 2005. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics*, **110**, 789–801.

Aitken K S, Li J C, Jackson P, Piperidis G, McIntyre C L. 2006. AFLP analysis of genetic diversity within *Saccharum officinarum* and comparison with sugarcane cultivars. *Australian Journal of Agricultural Research*, **57**, 1167–1184.

Almeida C M A D, Lima S E N D, Lima G S D A, Brito J Z D, Silva M V D. 2009. Caracterização molecular de cultivares de cana-de-açúcar utilizando marcadores ISSR. *Ciência E Agrotecnologia*, **33**, 1771–1776. (in Portuguese)

Amar M H, Biswas M K, Zhang Z, Guo W W. 2011. Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of *Citrus* germplasm collection. *Scientia Horticulturae*, **128**, 220–227.

Andrés J A, Bogdanowicz S M. 2012. Isolating microsatellite loci: Looking back, looking ahead. In: Orgogozo V, Rockman M, eds., *Molecular Methods for Evolutionary Genetics.* vol. 772. Springer, New York, USA. pp. 211–232.

Andru S, Pan Y B, Thongthawee S, Burner D M, Kimbeng C A J T. 2011. Genetic analysis of the sugarcane (*Saccharum*

spp.) cultivar 'LCP 85-384'. I. Linkage mapping using AFLP, SSR, and TRAP markers. *Theoretical and Applied Genetics*, **123**, 77–93.

Arruda P. 2012. Genetically modified sugarcane for bioenergy generation. *Current Opinion in Biotechnology*, **23**, 315–322.

Baert-Desurmont S, Charbonnier F, Houivet E, Ippolito L, Mauillon J, Bougeard M, Abadie C, Malka D, Duffour J, Desseigne F. 2016. Clinical relevance of 8q23, 15q13 and 18q21 SNP genotyping to evaluate colorectal cancer risk. *European Journal of Human Genetics*, **24**, 99–105.

Baraket G, Chatti K, Saddoud O, Abdelkarim A B, Mars M, Trifi M, Hannachi A S. 2011. Comparative assessment of SSR and AFLP markers for evaluation of genetic diversity and conservation of fig, *Ficus carica* L., genetic resources in Tunisia. *Plant Molecular Biology Reporter*, **29**, 171–184.

Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando M S. 2009. The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theoretical and Applied Genetics*, **118**, 653–669.

Bertani R P, Perera M F, Joya C M, Henriquez D D, Funes C, Chaves S, González V, Welin B, Cuenya M I, Castagnaro A P. 2021. Genetic diversity and population structure of *Acidovorax avenae* subsp. *avenae* isolated from sugarcane in Argentina. *Plant Pathology*, **70**, 1719–1732.

Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre C L. 1998. Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. *Genetica*, **104**, 143–153.

Brand-Miller J C, Barclay A W. 2017. Declining consumption of added sugars and sugar-sweetened beverages in Australia: a challenge for obesity prevention. *The American Journal of Clinical Nutrition*, **105**, 854–863.

Brandes E W, Sartoris G B. 1936. Sugarcane: Its origin and improvement. In: *Yearbook of the United States Department of Agriculture*. United States Government Printing Office, Washington, USA. pp. 561–624.

Cai Q, Aitken K S, Fan Y H, Piperidis G, Jackson P, McIntyre C L. 2005. A preliminary assessment of the genetic relationship between *Erianthus rockii* and the "Saccharum complex" using microsatellite (SSR) and AFLP markers. *Plant Science*, **169**, 976–984.

Chandra A, Jain R, Solomon S, Shrivastava S, Roy A K. 2013. Exploiting EST databases for the development and characterisation of 3 425 gene-tagged CISP markers in biofuel crop sugarcane and their transferability in cereals and orphan tropical grasses. *BMC Research Notes*, **6**, 47.

Chen C, Zhou P, Choi Y A, Huang S, Gmitter F G. 2006. Mining and characterizing microsatellites from citrus ESTs. *Theoretical and Applied Genetics*, **112**, 1248–1257.

Cho Y G, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S R, Park W D, Ayres N, Cartinhour S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, **100**, 713–722.

Cordeiro G M, Pan Y B, Henry R J. 2003. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. *Plant Science*, **165**, 181–189.

Creste S, Sansoli D M, Tardiani A C S, Silva D N, Pinto L R. 2014. Comparison of AFLP, TRAP and SSRs in the estimation of genetic relationships in sugarcane. *Toxicology*, **12**, 150–154.

Cui J, Cheng J, Nong D, Peng J, Hu Y, He W, Zhou Q, Dhillon N P S, Hu K. 2017. Genome-wide analysis of simple sequence repeats in bitter gourd (*Momordica charantia*). *Frontiers in Plant Science*, **8**, 1103.

D'Hont A, Grivet L, Feldmann P, Glaszmann J C, Rao S, Berding N. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics*, **250**, 405–413.

Daugrois J H, Grivet L, Roques D, Hoarau J Y, Lombard H, Glaszmann J C, D'Hont A. 1996. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. *Theoretical and Applied Genetics*, **92**, 1059–1064.

Deguchi K, Takeuchi H, Touge T, Kamoda M, Tsukaguchi M, Sasaki I, Nishioka M. 1996. Significance of cognitive function in the elicitation of sympathetic skin response in healthy humans. *Journal of the Autonomic Nervous System*, **61**, 123–127.

Deng P, Wang M, Feng K, Cui L, Tong W, Song W, Nie X. 2016. Genome-wide characterization of microsatellites in *Triticeae* species: abundance, distribution and evolution. *Scientific Reports*, **6**, 32224.

Devarumath R M, Kalwade S B, Kawar P G, Sushir K V. 2012. Assessment of genetic diversity in sugarcane germplasm using ISSR and SSR markers. *Sugar Tech*, **14**, 334–344.

Duan Y P, Chen W G, Li M S, Li X H, Zhang S H. 2006. The genetic diversity among 27 maize populations based on SSR data. *Scientia Agricultura Sinica*, **39**, 1102–1113. (in Chinese)

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater J M, Zyprian E, Moreira F M, Grando M S. 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology*, **13**, 39.

Fraser L G, Harvey C F, Crowhurst R N, De Silva H N. 2004. EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theoretical and Applied Genetics*, **108**, 1010–1016.

Fu P C, Zhang Y Z, Ya H Y, Gao Q B. 2016. Characterization of SSR genomic abundance and identification of SSR markers for population genetics in Chinese jujube (*Ziziphus jujuba* Mill.). *PeerJ*, **4**, e1735.

Giovannini D, Dettori M T, Leone A, Liverani A, Sirri S, Tellarini S, Verde I. 2012. Assessment of genetic variability in Italian heritage peach resources from Emilia-Romagna using microsatellite markers. *The Journal of Horticultural Science and Biotechnology*, **87**, 435–440.

Hajmansoor S, Bihamta M R, Alisoltani A. 2013. Genetic

diversity among and within Iranian and non-Iranian barely (*Hordeum vulgare* L.) genotypes using SSR and storage proteins markers. *Biochemical Systematics and Ecology*, **46**, 7–17.

Halsted C H, Devlin A M. 2001. Mutations in human glutamate carboxypeptidase II gene impacting folate metabolism, and detection of affected individuals. U.S. Patent Application 09/805,293[P]. 2002-1-24.

He Q, Li X W, Liang G L, Ji K, Guo Q G, Yuan W M, Zhou G Z, Chen K S, van de Weg W E, Gao Z S. 2011. Genetic diversity and identity of chinese loquat cultivars/Accessions (*Eriobotrya japonica*) using apple SSR markers. *Plant Molecular Biology Reporter*, **29**, 197–208.

Hirano R, Ishii H, Htun Oo T, Gilani S A, Kikuchi A, Watanabe K N. 2011. Propagation management methods have altered the genetic variability of two traditional mango varieties in Myanmar, as revealed by SSR. *Plant Genetic Resources*, **9**, 404–410.

Hunsigi G. 1993. *Production of Sugarcane: Theory and Practice.* Springer-Verlag, Berlin Heidelberg, Germany. ISBN: 978-3-642-78135-3.

Iniguez-Luy F L, Voort A V, Osborn T C. 2008. Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. *Theoretical and Applied Genetics*, **117**, 977–985.

Iwasaki H, Stewart P W, Dilley W G, Holt M S, Steinbrueck T D, Wells S A, Donis-Keller H. 1992. A minisatellite and a microsatellite polymorphism within 1.5 kb at the human muscle glycogen phosphorylase (PYGM) locus can be amplified by PCR and have combined informativeness of PIC 0.95. *Genomics*, **13**, 7–15.

Jones E, Dupal M, Dumsday J, Hughes L, Forster J. 2002. An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, **105**, 577–584.

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

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

Levinson G, Gutman G A. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, **4**, 203–221.

Li Z, Wang G, Liu X, Wang Z, Zhang M, Zhang J. 2021. Genome-wide identification and expression profiling of *DREB* genes in *Saccharum spontaneum*. *BMC Genomics*, **22**, 456.

Lim S, Notley-McRobb L, Lim M, Carter D A. 2004. A comparison of the nature and abundance of microsatellites in 14 fungal genomes. *Fungal Genetics and Biology*, **41**, 1025–1036.

Liu C, Zhang Q, Yao X, Zhong C, Yan C, Huang H. 2016. Characterization of genome-wide simple sequence repeats and application in interspecific genetic map integration in kiwifruit. *Tree Genetics & Genomes*, **12**, 21.

Liu F, Hu Z, Liu W, Li J, Wang W, Liang Z, Wang F, Sun X. 2016. Distribution, function and evolution characterization of microsatellite in *Sargassum thunbergii* (Fucales, Phaeophyta) transcriptome and their application in marker development. *Scientific Reports*, **6**, 18947.

Liu R, Koyanagi K O, Chen S, Kishima Y. 2012. Evolutionary force of AT-rich repeats to trap genomic and episomal DNAs into the rice genome: lessons from endogenous pararetrovirus. *The Plant Journal*, **72**, 817–828.

Lv B C, Su Q H, Xin X X, Sun J, Liu X X, Liu Y, Yang D L. 2019. Construction of genetic linkage map by SSR markers for wheat ILs population. *Journal of Gansu Agricultural University*, **54**, 35–42. (in Chinese)

Ma P, Zhang X, Chen L, Zhao Q, Zhang Q, Hua X, Wang Z, Tang H, Yu Q, Zhang M, Ming R, Zhang J. 2020. Comparative analysis of sucrose phosphate synthase (SPS) gene family between *Saccharum officinarum* and *Saccharum spontaneum*. *BMC Plant Biology*, **20**, 422.

Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, **30**, 194–200.

Nachimuthu V V, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian B A, Ponniah G, Gunasekaran K, Swaminathan M, Suji K K, Sabariappan R. 2015. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in *Oryza sativa*. *Rice*, **8**, 30.

Nawaz S, Khan F A, Tabasum S, Zakria M, Saeed A, Iqbal M Z. 2010. Phylogenetic relationships among *Saccharum* clones in Pakistan revealed by RAPD markers. *Genetics and Molecular Research*, **9**, 1673–1682.

Nayak S N, Song J, Villa A, Pathak B, Ayala-Silva T, Yang X, Todd J, Glynn N C, Kuhn D N, Glaz B. 2014. Promoting utilization of *Saccharum* spp. genetic resources through genetic diversity analysis and core collection construction. *PLoS ONE*, **9**, e110856.

Olatoye M O, Clark L V, Wang J, Yang X, Yamada T, Sacks E J, Lipka A E. 2019. Evaluation of genomic selection and marker-assisted selection in *Miscanthus* and energycane. *Molecular Breeding*, **39**, 171.

Oliveira K M, Pinto L R, Marconi T G, Margarido G R A, Pastina M M, Teixeira L H M, Figueira A V, Ulian E C, Garcia A A F, Souza A P. 2007. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Molecular Breeding*, **20**, 189–208.

Pan Y B, Liu P. 2010. Development of a genetic linkage map for Louisiana sugarcane: New microsatellite (SSR) DNA markers identified for LCP 85-384. *Journal of the American Society of Sugar Cane Technologists*, **30**, 137.

Pan Y B, Tew T L, Schnell R J, Viator R P, Richard E P, Grisham M P, White W H. 2006. Microsatellite DNA marker-assisted selection of *Saccharum spontaneum* cytoplasm-derived germplasm. *Sugar Tech*, **8**, 23–29.

Pang X, Tan X, Wei L, Liang F, Zhang J, Lyu P, Cheng Q, Huang Q, Zhou Q. 2019. Genetic diversity analysis for 35

sugarcane germplasm resources and construction of DNA fingerprint. *Journal of Southern Agriculture*, **50**, 1157–1164.

Parida S K, Kalia S K, Kaul S, Dalal V, Hemaprabha G, Selvi A, Pandit A, Singh A, Gaikwad K, Sharma T R, Srivastava P S, Singh N K, Mohapatra T. 2009. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. *Theoretical and Applied Genetics*, **118**, 327–338.

Pinto L R, Oliveira K M, Marconi T, Garcia A A F, Ulian E C, De Souza A P. 2006. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. *Plant Breeding*, **125**, 378–384.

Pinto L R, Oliveira K M, Ulian E C, Garcia A A F, De Souza A P. 2004. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome*, **47**, 795–804.

Potts S M, Han Y, Khan M A, Kushad M M, Rayburn A L, Korban S S. 2012. Genetic diversity and characterization of a core collection of malus germplasm using simple sequence repeats (SSRs). *Plant Molecular Biology Reporter*, **30**, 827–837.

Ramakrishnan M, Ceasar S A, Duraipandiyan V, Al-Dhabi N A, Ignacimuthu S. 2016. Using molecular markers to assess the genetic diversity and population structure of finger millet (*Eleusine coracana* (L.) Gaertn.) from various geographical regions. *Genetic Resources and Crop Evolution*, **63**, 361–376.

Selvi A, Nair N V, Balasundaram N, Mohapatra T. 2003. Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane. *Genome*, **46**, 394–403.

Da Silva C M, Mangolin C A, Mott A S, Machado M F P S. 2008. Genetic diversity associated with *in vitro* and conventional bud propagation of *Saccharum* varieties using RAPD analysis. *Plant Breeding*, **127**, 160–165.

Da Silva J A G. 2001. Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). *Genetics and Molecular Biology*, **24**, 155–159.

Singh R B, Singh B, Singh R K. 2017. Study of genetic diversity of sugarcane (*Saccharum*) species and commercial varieties through TRAP molecular markers. *Indian Journal of Plant Physiology*, **22**, 332–338.

Singh R B, Singh B, Singh R K. 2019. Development of potential dbEST-derived microsatellite markers for genetic evaluation of sugarcane and related cereal grasses. *Industrial Crops and Products*, **128**, 38–47.

Singh R K, Singh R B, Singh S P, Sharma M L. 2011. Identification of sugarcane microsatellites associated to sugar content in sugarcane and transferability to other cereal genomes. *Euphytica*, **182**, 335.

Skinner J C, Hogarth D M, Wu K K, Heinz D J. 1987. Sugarcane improvement through breeding. In: Heinz D J, ed., *Developments in Crop Science*. vol. 11. pp. 409–453. Elsevier, Amsterdam.

Souza M C P D, Silva J N, Almeida C. 2013. Differential detection of transposable elements between *Saccharum* species. *Genetics and Molecular Biology*, **36**, 408–412.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research*, **11**, 1441–1452.

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.). *Theoretical and Applied Genetics*, **100**, 697–712.

Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, **106**, 411–422.

Tong Z J, Jiao F C, Xiao B G. 2015. Analysis of SSR loci in *Nicotina tabacum* genome and its two ancestral species genome. *Scientia Agricultura Sinica*, **48**, 2108–2117. (in Chinese)

Waclawovsky A J, Sato P M, Lembke C G, Moore P H, Souza G M. 2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal*, **8**, 263–276.

Wang L, Zheng Y, Ding S, Zhang Q, Chen Y, Zhang J. 2017. Molecular cloning, structure, phylogeny and expression analysis of the invertase gene family in sugarcane. *BMC Plant Biology*, **17**, 109.

Wang P, Wang C Y, Zhang L X, Cong L, Zhu Z X, Lu X C. 2019. Development of SSR molecular markers with sorghum polymorphism using re-sequencing. *Biotechnology Bulletin*, **35**, 217–223.

Wang Q, Fang L, Chen J, Hu Y, Si Z, Wang S, Chang L, Guo W, Zhang T. 2015. Genome-wide mining, characterization and development of microsatellite markers in *Gossypium* species. *Scientific Reports*, **5**, 10638.

Wang Y W, Samuels T D, Wu Y Q. 2011. Development of 1,030 genomic SSR markers in switchgrass. *Theoretical and Applied Genetics*, **122**, 677–686.

Wang Z, Pan Y, Luo J, You Q, Xu L, Zhang H, Que Y. 2020. SSR-based genetic identity of sugarcane clones and its potential application in breeding and variety extension. *Sugar Tech*, **22**, 367–378.

Xu J, Liu L, Xu Y, Chen C, Rong T, Ali F, Zhou S, Wu F, Liu Y, Wang J, Cao M, Lu Y. 2013. Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Research*, **20**, 497–509.

Xu M, Liu X, Wang J W, Teng S Y, Shi J Q, Li Y Y, Huang M R. 2017. Transcriptome sequencing and development of novel genic SSR markers for *Dendrobium officinale*. *Molecular Breeding*, **37**, 18.

Xu M L, Melchinger A E, Xia X C, Lübberstedt T. 1999. High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. *Molecular and General Genetics*, **261**,

574–581.

Yao W, Yu A L, Xu J S, Zhou H, Zhang M Q, Chen R K. 2005. A simple and quick method for extracting sugarcane genomic DNA. *Journal of Agricultural Biotechnology*, **13**, 121–122. (in Chinese)

Yang S, Li X, Huang F, Huang Y, Liu X, Wu J, Wang Q, Deng Z, Chen R, Zhang M. 2018. A new method based on SNP of nrDNA-ITS to identify *Saccharum spontaneum* and its progeny in the genus *Saccharum*. *PLoS ONE*, **13**, e0197458.

You Q, Yang X, Peng Z, Islam M S, Sood S, Luo Z, Comstock J, Xu L, Wang J. 2019. Development of an Axiom Sugarcane100K SNP array for genetic map construction and QTL identification. *Theoretical and Applied Genetics*, **132**, 2829–2845.

Yu J K, La Rota M, Kantety R V, Sorrells M E. 2004. EST derived SSR markers for comparative mapping in wheat and rice. *Molecular Genetics and Genomics*, **271**, 742–751.

Zane L, Bargelloni L, Patarnello T. 2010. Strategies for microsatellite isolation: A review. *Molecular Ecology*, **11**, 1–16.

Zawedde B M, Ghislain M, Magembe E, Amaro G B, Grumet R, Hancock J. 2015. Characterization of the genetic diversity of Uganda's sweet potato (*Ipomoea batatas*) germplasm using microsatellites markers. *Genetic Resources and Crop Evolution*, **62**, 501–513.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, Wai C M, Zheng C, Shi Y, Chen S, Xu X, Yue J, Nelson D R, Huang L, Li Z, Xu H, *et al.* 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics*, **50**, 1565–1573.

Zhang J, Zhou M, Walsh J, Zhu L, Ming R. 2013. *Sugarcane Genetics and Genomics*. In: Moore P H, Botha F C, eds., *Sugarcane: Physiology, Biochemistry, and Functional Biology*. John Wiley & Sons, Inc., Hoboken, New Jersey, USA. pp. 623–643.

Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan M A, Zhou Y, Gu C, Zhang X, Han Z, Korban S S, Li S, Han Y. 2012. Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. *BMC Genomics*, **13**, 537.

Zhang Q, Qi Y, Pan H, Tang H, Wang G, Hua X, Wang Y, Lin L, Li Z, Li Y, Yu F, Yu Z, Huang Y, Wang T, Ma P, Dou M, Sun Z, Wang Y, Wang H, Zhang X, *et al.* 2022. Genomic insights into the recent chromosome reduction and polyploidization of complex autopolyploid sugarcane *S. spontaneum*. *Nature Genetics*, **54**, 885–896 .

Zhang Z, Deng Y, Tan J, Hu S, Yu J, Xue Q. 2007. A genome-wide microsatellite polymorphism database for the *indica* and *japonica* rice. *DNA Research*, **14**, 37–45.

Executive Editor-in-Chief HUANG San-wen
Managing Editor WENG Ling-yun