International Journal of *Molecular Sciences*

MDPI

Article

Comparative Analysis of Chloroplast Genome in *Saccharum* spp. and Related Members of '*Saccharum* Complex'

Sicheng Li [1,2#], Weixing Duan[3#], Jihan Zhao [1,2#], Yanfen Jing [4], Mengfan Feng [1,2], Bowen Kuang[1,2], Ni Wei[1,2], Baoshan Chen[1,2*] and Xiping Yang [1,2*]

[1] State Key Laboratory of Conservation and Utilization of Subtropical Agro-bioresources, Guangxi University, Nanning 530004; China. lamina0130@126.com (S.L.); fjnldxzxs@163.com (J.Z.); mengfanfeng@st.gxu.edu.cn (M.F.); kuangbowen97@163.com (B.K.); 1917301038@st.gxu.edu.cn (N.W.).
[2] Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, China
[3] Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/Sugarcane Research Center, Chinese Academy of Agricultural Sciences/Guangxi Key Laboratory of Sugarcane Genetic Improvement/Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture and Rural Affairs, Nanning, Guangxi 530007, China. duanweixing84@126.com (W.D.)
[4] Ruili Breeding Station of Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, Ruili, Yunnan 678600, China. rljyf@126.com (Y.J.)
#Co-first author; *corresponding author
* Correspondence: xipingyang@gxu.edu.cn (X.Y.); chenyaoj@gxu.edu.cn (B.C.).

Abstract: Sugarcane is widely cultivated in tropical and subtropical regions in more than 90 countries. The high ploidy of the sugarcane nuclear genome limits its genomic studies, while its chloroplast genome is small and conserved, which is suitable for phylogenetic relationship studies and molecular marker development, and has great potential for parent selection and cross verification in sugarcane breeding program. In this study, 19 chloroplast genomes had been sequenced and assembled including 3 species of the genera *Saccharum* (*S. robustum*, *S. sinense*, and *S. barberi* firstly reported in this study) and 2 sugarcane relative species, *Tripidium arundinaceum* and *Narenga porphyrocoma*. The gene content and arrangement were highly conserved among all analyzed species. Phylogram based on 74 protein-coding genes showed *T. arundinaceum* was distantly related to *Saccharum* genera, and *N. porphyrocoma* might be diverse from the same ancestor shared with *Saccharum*. The plastid phylogenetic signal demonstrated that *S. officinarum* and *S. robustum* shared a common ancestor, and cytoplasmic origin of *S. sinense* and *S. barberi* was much ancient than the *S. officinarum*/*S. robustum* linage. The nucleotide substitutions in these plastomes detected several highly variable regions, *rps16-trnQ*, *trnS-psbD*, *trnS-psbZ*, *trnR-rps14*, *ndhC-trnV*, *petL-petG*, *rpl33-rps18* IGS regions , and *rps12* intron. A total of 14 markers were developed, including 9 InDel markers at the inter- and intra-species level for distinguishing *Saccharum* and its relative species, 4 dCAPS markers for *S. officinarum* and *S. robustum*, and 1 dCAPS marker for *S. sinense* and *S. barberi*. These markers have been demonstrated their highly discriminatory power in the *Saccharum* genera. The results obtained from our studies will contribute to the understanding of Saccharinae classification, plastome evolution at inter- and intra-species level of *Saccharum* genera, and the discrimination of parents and reciprocal crosses within the genera *Saccharum* in sugarcane breeding program.

Keywords: Sugarcane; chloroplast genome; phylogenomics; InDel marker; dCAPS marker

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Int. J. Mol. Sci.* **2022**, *23*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Sugarcane (*Saccharum* spp.), belonging to Poaceae, Panicoideae, Andropogoneae, Saccharinae, is widely cultivated in tropical and subtropical regions in more than 90 countries (http://www.fao.org). The *Saccharum* genus consists of two wild species *S. spontaneum* and *S. robustum*, two ancient cultivars *S. barberi* for cultivars from India and *S. sinense*

from China, along with "noble" sugarcane *S. officinarum* that has been used for sugar production for hundred years. Modern sugarcane interspecific hybridization work undertaken in Java during the 1920s[1]. Since then, several founder sugarcane cultivars/hybrids were frequently used as parental material for breeding, resulting in a narrow genetic base for modern sugarcane cultivars, which had limited the continued improvement of commercial traits of varieties.

The *Saccharum* complex/Saccharinae subfamily has some genera closely related to sugarcane, which are divergent in a large number of biological and agronomic traits, but capable of interbreeding with sugarcane. Among them, *Tripidium arundinaceum* (2n=60)[2] and *Narenga porphyrocoma* (2n = 30)[3] has excellent field characteristics, such as high biomass, barren and drought tolerance, early maturity, strong stems, strong tillering ability, and mosaic disease resistance[4]. Particularly, considerable efforts had been made in introgressing robust traits from *Tripidium* species into sugarcane. Meanwhile, *N. porphyrocoma* also have been used for sugarcane breeding through distant hybridization[3]. The F_1 and BC_1 hybrid of sugarcane and *N. porphyrocoma* showed resistance to smut[5]. Since the *Saccharum* complex/Saccharinae comprises gene resources that are helpful for sugarcane variety improvement, it's necessary to obtain an accurate phylogenetic relationship between *Saccharum* and its purported relatives, particularly for those species from outside of the core *Saccharum*.

Chloroplasts (cp) are inherited maternally in the majority of plants, and each of these organelles contains a quadripartite circular molecule of double-stranded DNA that comprises two inverted repeats (IRs), and single-copy regions, the large and small single copy regions (LSC and SSC). Because of their relatively small size, simple structure, and conserved gene content, cpDNA sequences have been widely used for phylogenetic studies, and complete cp genome sequences could provide valuable datasets for resolving taxonomical complexes and/or DNA barcoding (e.g., *matK* and *rbcL*)[6]. In addition, the cp genomes have proven to be more informative than cp DNA fragments in revealing the phylogeny of land plants. With the advent of next-generation sequencing (NGS) technologies, plastome-scale data can be assembled for multipurpose tasks. *Saccharum* species have been restricted to investigate in nuclear genomics because of their high ploidy, and large genome size, while their chloroplast genomes are much smaller and relatively simple. Therefore, comparison of the complete cp genome sequences could reveal novel genome features such as single-nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and microsatellites. Moreover, SNPs-based dCAPS markers had been utilized for genetic analysis and selective breeding in allopolyploid genomes such as oats[7], wheat[8], and rape[9].

In genera *Saccharum*, the complete cp genome of *Saccharum hybrid* NCo310[10], SP80-3280[11], and RB867515[12]; *S. spontaneum* SES205A, SES234B[13], and Yunnan 83-184[14]; *S. officinarum* Badila and IJ76-514 were obtained in previous studies. More recently, the complete cp genome of *T. arundinaceum*[15,16] and *N. porphyrocoma*[17] had been sequenced and assembled. The analysis of chloroplasts in *Saccharum*-related focused on three aspects: (1) the structural features of cp genome were compared with *Saccharum* species and other agronomically and industrially important monocots such as *Zea mays*, and *Sorghum bicolor*; (2) molecular classification of all *Saccharum* species and its nearest evolutionary neighbors were elucidated by plastid-based studies; (3) the plastid view on sugarcane origins. These results have indicated the potential of this approach to elucidate the phylogenetic relationships within the '*Saccharum* complex'.

However, as the major member of *Saccharum* species, the cp genomes of *S. robustum*, *S. barberi*, and *S. sinense* have not been sequenced and assembled yet, which is urgently required for comprehensive understanding of the genus *Saccharum*. Meanwhile, the cp genomes of representative germplasm within the genus sugarcane sequenced and assembled were absent. Molecular sequence studies on the *Saccharum* genus to date have been small in scale, involving either limited marker studies or being based only on phylogenies derived from a few chloroplast regions or a few genomic regions. The situation is even

more complex when attempting to distinguish between cultivars that might only be a few decades divergent from one another. In this study, we aimed to obtain the complete cp genome sequence of *Saccharum* spp. and its relative genera to characterize those genome structures, gene content, and other characteristics. The whole-genome sequencing (WGS) paired-end reads were used to assemble the corresponding plastomes to (1) enrich, update and improve the database genomics of Saccharinae, (2) improve the phylogenetic resolution of the defined species complexes, (3) develop efficient markers for parent selection and cross verification in sugarcane breeding program.

2. Results

2.1. Plastid genome structure and organization

We sequenced and *de novo* assembled 19 cp genomes of Saccharinae members. Combined with 4 genomes assembled using NCBI SRA database, a total of 23 chloroplast genomes (Table S1) were included for plastid genome comparison, which included one species from genera *Tripidium*, one from genera *Narenga*, and twenty-one species from genera *Saccharum*, including four *S. spontaneum* species, four *S. robustum* species, two *S. sinense* species, three *S. barberi* species, four *S. officinarum* species, and four *S.* hybrids. The complete chloroplast genome of *N. porphyrocoma* (141,231 bp) (Figure 1a) contained a large single copy (LSC) region (83,107 bp), and a small single-copy (SSC) region (12,533 bp), which were separated by a pair of IRs (IRa and IRb; 22,794 bp each), 8 rRNA, and 38 tRNA, while that of *T. arundinaceum* (141,209 bp) (Figure 1b) consisted of an LSC (83,167 bp), an SSC (12,513 bp), and a pair of IRs (22,763 bp each), 8 rRNA and 38 tRNA.

In genera *Saccharum*, assembly sizes ranged from 141,161 bp (IND81-013, *S. spontaneum*) to 141,216 bp (SP79-9, *S. spontaneum*) (Figure 1c). All had a typical quadripartite structure, comprising a SSC (12,538–12,542 bp) and a LSC (82,036–83,093 bp) region separated by a pair of IRs (22,789–22,795 bp). The cp genome length varied from 141,161 to 141,216 bp with a medium-size of 141,177 bp. In the *Saccharum*, *S. spontaneum* ($\bar{x}$ = 141,183 ± 26 bp) has the most variable length, followed by *S. robustum* (141,171–141,187 bp, $\bar{x}$ = 141,179 ± 9 bp) and *S. officinarum* (141,182–141,183 bp, $\bar{x}$ = 141,183 ± 1 bp) (Fig1.c). Across the 21 chloroplast genome sequences in *Saccharum*, the cp genome length was most conserved in *S. sinense* (141,174 bp, $\bar{x}$ = 141,174 ± 0 bp) and *S. barberi* (141,174 bp, $\bar{x}$ = 141,174 ± 0 bp), as well as *S.* hybrid (141,182 bp, $\bar{x}$ = 141,182 ± 0 bp) (Table 1).

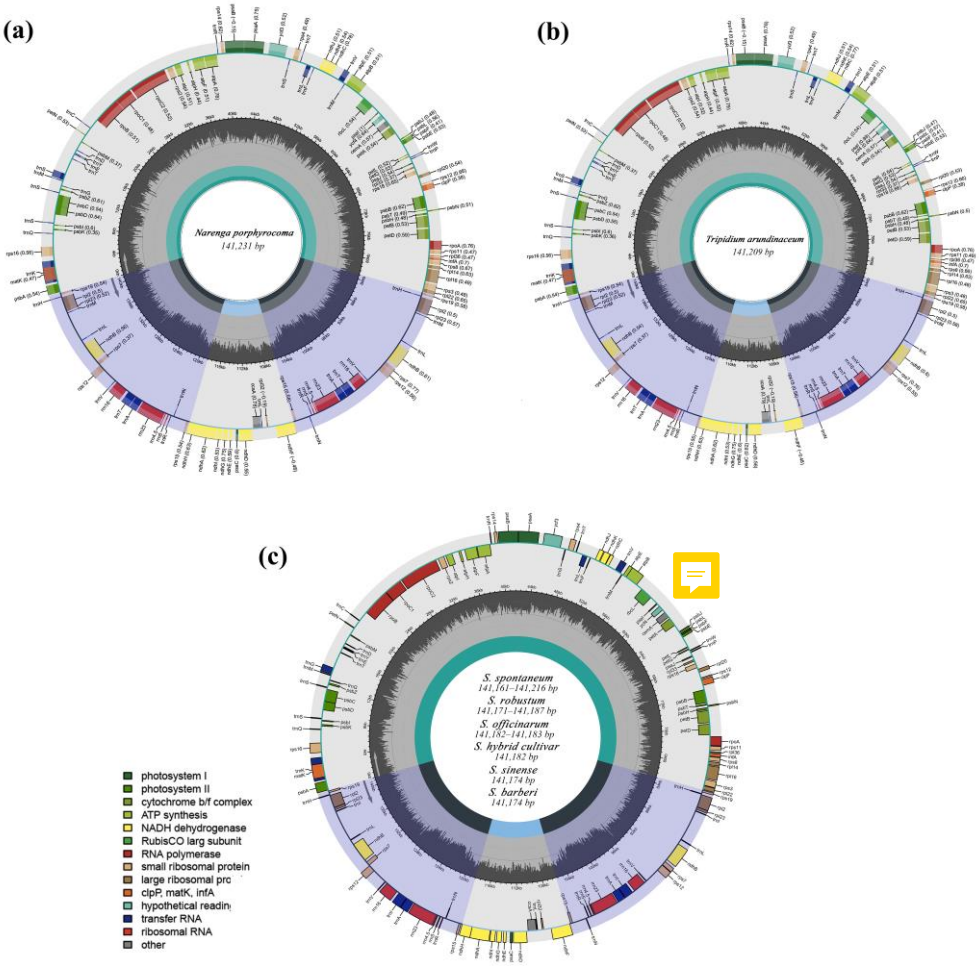


Figure 1. The complete structure of the assembled and annotated chloroplast genomes. The color of genes indicated their function. The genes presented outside the circle were transcribed counterclockwise, whereas the genes presented inside the circle were transcribed clockwise. The gene content and organization were similar for all *Saccharum* species, and therefore, one figure(c) was drawn as representative of all three species (a) *N. porphyrocoma*; (b) *T. arundinaceum*; (c) *Saccharum* spp..

In terms of synteny and gene number, both were highly conserved in the assembled accessions in genera *Saccharum*. The plastome consisted of 84 protein-coding genes (7 repeated/IR genes, *ndhB*, *rps12*, *rps19*, *rpl2*, *rpl23*, *rps7*, and *rps15*, and there were two copies of *rps15*), 29 tRNA genes (8 repeated/IR genes, *trnM-CAU*, *trnH-GUG*, *trnI-CAU*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG* and *trnN-GUU*, and there were two copies of *trnN-GUU*), and 4 rRNA genes (4 repeated/IR genes, *rrn23*, *rrn4.5*, *rrn5*, and *rrn16*) (Table S2). Among the annotated genes, there were 13 genes had introns in all species, of which there were two introns for *ycf3*, and one intron for the other 12 genes (*atpF*, *ndhA*, *ndhB*, *rps16*, *rpl16*, *rpl2*, t*rnK-UUU*, *trnS-CGA*, *trnL-UAA*, *trnV-UAC*, *trnT-CGU*, and *trnA-UGC*). Interestingly, the gene *rps12* had two introns but with a trans-splicing, with one of its exons located in the LSC region and the other duplicated by the IRs. The members of the *Saccharum* complex lost *accD* (encoded subunits of Acetyl-CoA carboxylase) and *ycf1*, which were also absent in the cp genomes of other Panicoideae grasses [16].

Table 1. The structure and length of the chloroplast genome sequences in *N. porphyrocoma*, *T. arundinaceum*, and *Saccharum* spp. assembled in this study.

Species	**Non-*Saccharum***		***S. spontaneum***				***S. robustum***				***S. sinense***	
Accession	BM87-36	HBW1	SES14	SP79-9	IN84-066	IND81-013	NG51-63	NG57-208	IS76-184	IN84-076	Guilinzhuzhe	Youba
Length/bp	141209	141231	141162	141216	141191	141161	141171	141171	141187	141187	141174	141174
LSC/bp	83167	83107	83039	83093	83068	83038	83036	83036	83056	83056	83043	83043
SSC/bp	12513	12533	12542	12542	12542	12542	12542	12542	12538	12538	12538	12538
IR/bp	22763	22794	22789	22789	22789	22789	22795	22795	22795	22795	22795	22795
Gene	130	130	130	130	130	130	130	130	130	130	130	130
PCGs	84	84	84	84	84	84	84	84	84	84	84	84
rRNA genes	8	8	8	8	8	8	8	8	8	8	8	8
tRNA genes	38	38	38	38	38	38	38	38	38	38	38	38

Species	***S. barberi***			***S. officinarum***				***S.* hybrid**			
Accession	Katha	Kewali	Chunne	Loether	Christalin	BlackChLBH	NG96-14	POJ2878	ROC25	GT42	ZhongzheNo1
Length/bp	141174	141174	141174	141183	141182	141182	141183	141182	141182	141182	141182
LSC/bp	83043	83043	83043	83048	83047	83047	83048	83047	83047	83047	83047
SSC/bp	12538	12538	12538	12542	12542	12542	12542	12542	12542	12542	12542
IR/bp	22795	22795	22795	22795	22795	22795	22795	22795	22795	22795	22795
Gene	130	130	130	130	130	130	130	130	130	130	130
PCGs	84	84	84	84	84	84	84	84	84	84	84
rRNA genes	8	8	8	8	8	8	8	8	8	8	8
tRNA genes	38	38	38	38	38	38	38	38	38	38	38

[1] PCGs :protein-coding genes.

2.2. Phylogenetic analysis of Saccharinae species

Maximum likelihood (ML) phylogenetic analysis was performed based on 74 protein-coding genes (PCGs), which were highly conserved single copy protein-coding genes shared between 36 species in Saccharinae (23 cp genome obtained in this study) (Table S3), rooted using *Sorghum bicolor* as outgroup. The phylogenetic tree showed that all *Saccharum* species were closely related (Figure 2). *S. spontaneum* diverged from other *Saccharum* species firstly, while *S. sinense* and *S. barberi* were clustered into the same clades, being sisters to *S. robustum* and *S. officinarum*. Our data was not possible to determine if *S. sinense* pre-dates *S. barberi* or vice versa. They probably had different *S. officinarum* accessions as an ancestor in the hybridization, contributing to their difference in terms of genome compositions. The phylogram showed that *S. officinarum* and sugarcane modern cultivar were more close to *S. robustum* than ancient hybrid species *S. sinense* and *S. barberi*.

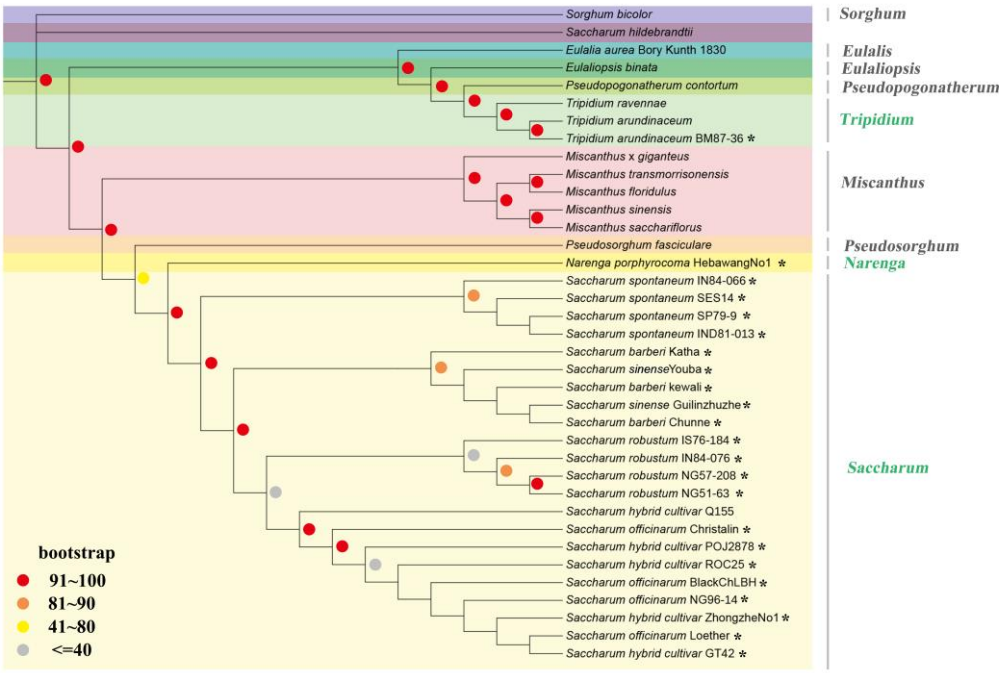


Figure 2. Phylogram based on 74 protein-coding genes shared between Saccharinae species (23 newly assembled in this study, which had been marked with *) and 13 published accessions of Saccharinae species in NCBI. *Sorghum bicolor* was included as outgroup.

2.3. Repeat Analysis and genetic diversity assessment

Chloroplast repeats with high diversity in copy numbers are usually candidate markers to investigate population genetics and biogeography of allied taxa. In the MISA analyses, a total of 36 microsatellites regions (> = 10 bp) were identified in the six *Saccharum* plastid genomes, 34 in *T. arundinaceum* and 33 in *N. porphyrocoma*. The overall distribution, types, and numbers of repeats are highly similar among the plastomes. The most abundant SSRs were mononucleotide, and the number of that was 26 in *T. arundinaceum* and *S. spontaneum*, and 27 in other species. The number of dinucleotide SSRs was the same in all species. Interestingly, trinucleotide SSRs were found in the plastomes of *T. arundinaceum*, and hexanucleotide SSRs were found in the plastomes of *N. porphyrocoma*, but were not detected in the other six plastomes. Detailed information about SSR type, SSR sequence, regions, and start/end positions of the aforementioned SSRs in the cp genome was presented in supplementary Table S4 and Table S5.

There were no large structural variations except for SNPs, and InDels in *Saccharum* cp genomes (Table S6 and Table S7). Among six *Saccharum* species, *S. spontaneum* had the largest number of SNPs (68) and InDels (43), while *S. officinarum* only had fewer number of SNPs (10) and InDels (5). *S. robustum* had 20 SNPs and12 InDels been detected, and *S. sinense* and *S. barberi* had the same variation and number of SNPs (22) and InDels (10). Most of the SNPs were found in intergenic sequences (52), and the maximum number of SNPs was found in *ndhF-rpl32*(8), followed by *ndhC-trnV-UAC* (7), *psbM-petN* (4), *trnG-GCC-trnM-CAU* (4), *rps16-trnQ-UUG* (3), and *petA-psbJ* (3). Seven SNPs were reported from introns of six genes, which was rps16 (1), *trnG-UCC* (1), *ycf3* (1), *trnV-UAC* (1), *rps12* (2), and *rps16* (1). Fifteen SNPs were located in the coding sequence (CDS) of genes, involved in a variety of biological processes in the chloroplast from transcription (*rpoC1* and *rpoC2*) to translation (*rps8* and *rps3*), the subunits of ATP synthase (*atpA*, *atpB*, and *atpF*), subunits of photosystem II (*psbE* and *ycf3*), c-type cytochrome synthesis (*ccsA*), and subunits of NADH-dehydrogenase (*ndhF* and *ndhD*). Most of InDels was found in intergenic sequences except for two InDels found in *ycf3* introns in *S. spontaneum*. Two long insertion fragments found in *S. spontaneum* were 27 bp and 50 bp, while one base insertion was found in other *Saccharum* species.

2.4. Comparison of chloroplast genome sequences

The LSC/IRs and IRs/SSC borders and their adjacent genes were compared in the six *Saccharum* cp genomes and other Saccharine species (Figure 3). The *rps19* gene was located in the IRb region entirely, 35 bp away from the LSC/IRb border. The *ndhF* gene was situated at the junction of IRb/SSC, extending 29 bp into the IRb region. Overall, the structure and gene content of the six chloroplast genomes were consistent, and no significant expansion or contraction of IR regions were found in the *Saccharum* spp. and two related species.

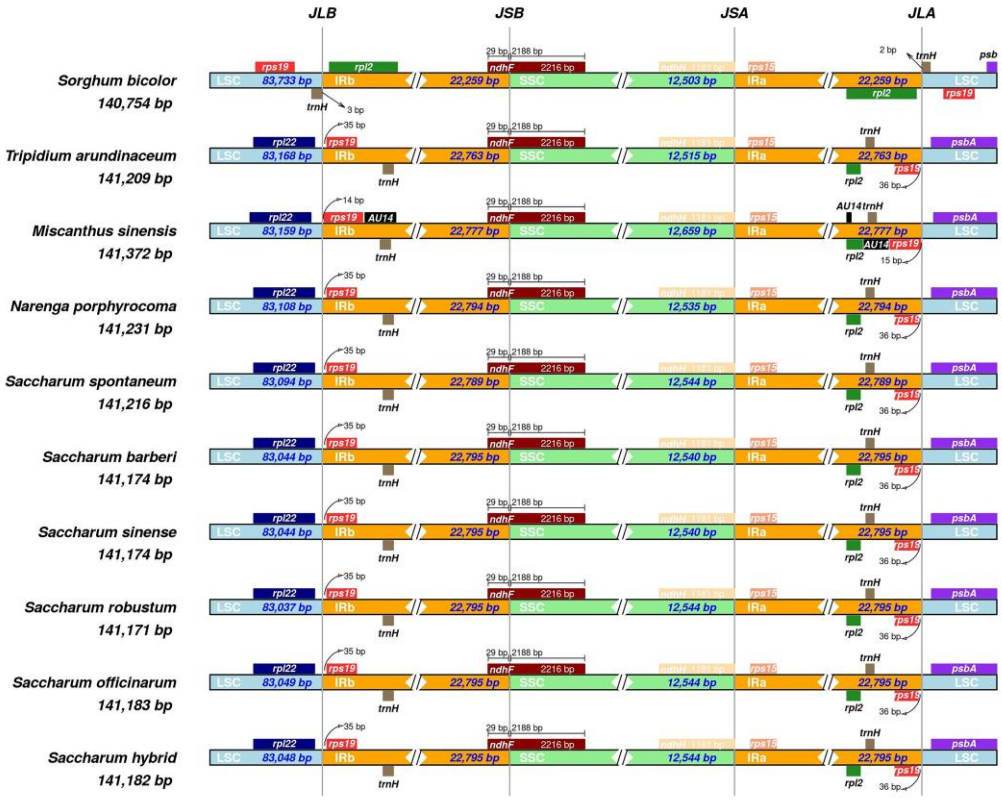


Figure 3. Comparison of border positions of LSC, SSC, and IR regions among nine Saccharinae species and *S. bicolor*. Genes are denoted by boxes, and the gap between the genes and the boundaries

was indicated by the number of bases unless the gene coincides with the boundary. Extensions of genes were also indicated above the boxes.

To investigate the similarities and differences of the cp genome sequences between genera *Saccharum* and other species of Saccharinae, a global alignment program was used to align these sequences. The result was plotted using the mVISTA tools with *sorghum bicolor* as a reference. Those closely related species had little difference in cp genome size, ranging from 141,174 bp to 141,231 bp. As shown in Figure 4, the structures of these cp genomes were conserved. The global patterns of sequence similarities among these sequences were very high.

Comparative sequence analysis of the eight Saccharinae species using mVISTA revealed a high sequence identity across the seven species, with identity ranging from 97.89% to 99.95%, suggesting that Saccharinae chloroplast genomes were fairly conserved. Overall, tRNA region was the most conservative region (identity = 99.95 ± 0.01 %), and exon regions (identity = 99.80 ± 0.00 %) was more conserved than intron region (identity = 99.04 ± 0.14 %) and intergenic regions (identity = 98.19 ± 0.19 %). Most high polymorphic regions were located in the intergenic regions (such as *rps16-trnQ*, *trnG-trnM*, *psbM-petN*, *ndhC-trnV*, *petA-psbJ*, and *psbJ-rpl33*). These regions may be undergoing more rapid nucleotide substitution at the species level, which indicated its potential application as molecular markers for phylogenetic analyses and plant identification in Saccharinae species. A few divergent regions were also observed in the genes coding regions of *matK*, *psbB*, *psaB*, and *rpoA*, and introns of *atpF*, *rps16*, *rpl16*, and *ycf3*.

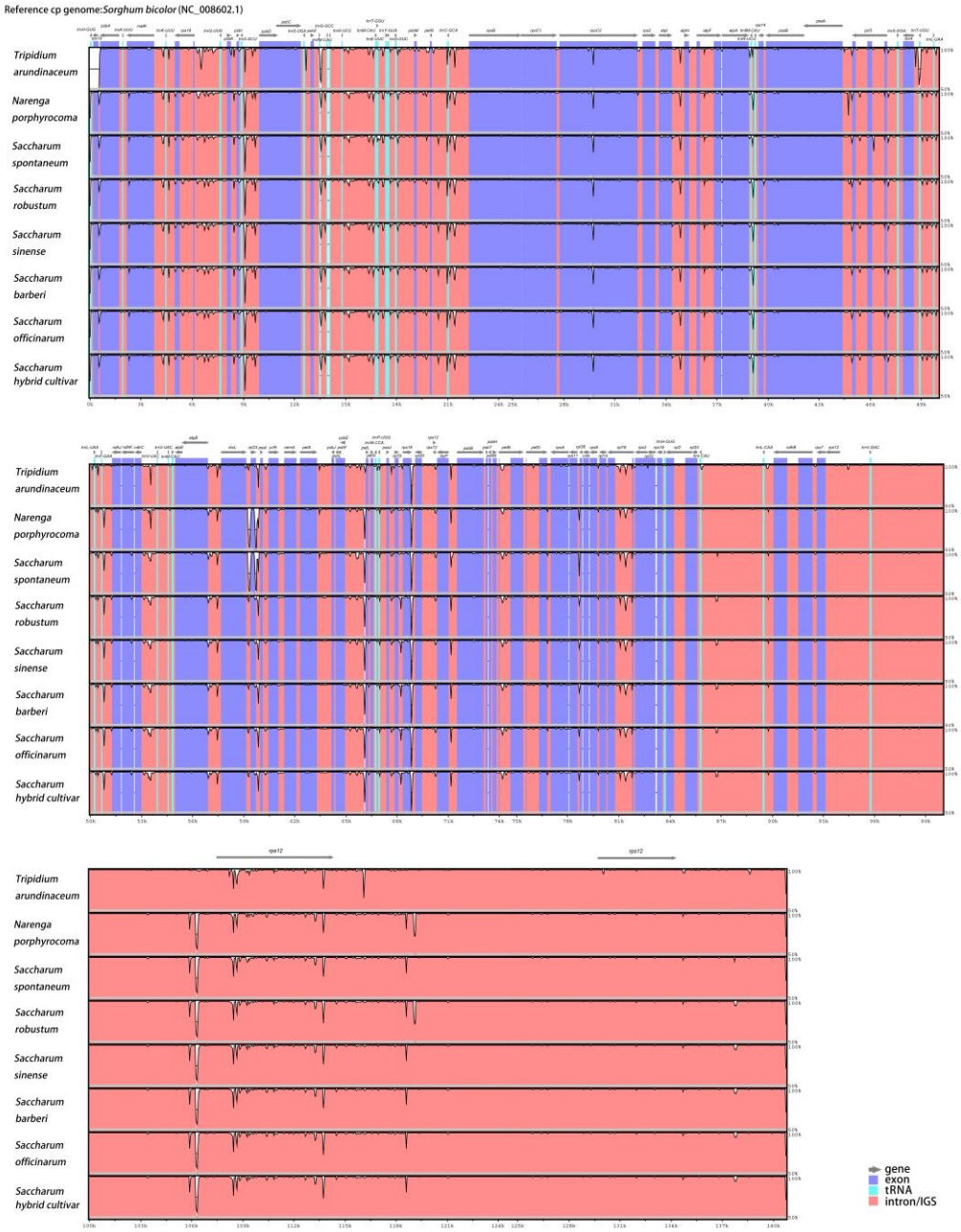


Figure 4. Chloroplast genomes were aligned using the mVISTA program with the annotation of *S. bicolor* as a reference. The X- and Y-scales indicated the coordinates within chloroplast genomes and the percentage of identity (50 %~100 %), respectively. Gray arrows indicated the direction of transcription of each gene. Annotated genes were displayed along the top. The red block was the intron region, the pearlecent Aqua block was the UTR region, and the purple block was the CDS region.

2.5. Development of markers to discriminate Saccharum and sugarcane-relative species

The nucleotide diversity (Pi) of the chloroplast genomes in the six *Saccharum* species was also calculated to assess the sequence divergence level of these species (Figure 5). The average Pi value was 0.000265 (range 0–0.00518) in the LSC region, and 0.000468 (range 0–0.00324) in the SSC region, while that of the IR region was 0. Most of the sequences with high Pi values were spacer regions between genes. Among these spacer regions, *rps16-trnQ*, *trnG-trnM*, *psbM-petN*, *petN-trnC*, *ndhC-trnV*, *petA-psbJ*, *psbJ-rpl33*, *ndhF-rpl32*, and

trnL-ccsA were the nine regions having Pi values bigger than 0.01000. Only one coding region (*ycf3*) had high Pi values over 0.01000. These hotspot regions of divergence could be applied in developing DNA markers for distinguishing *Saccharum* species.

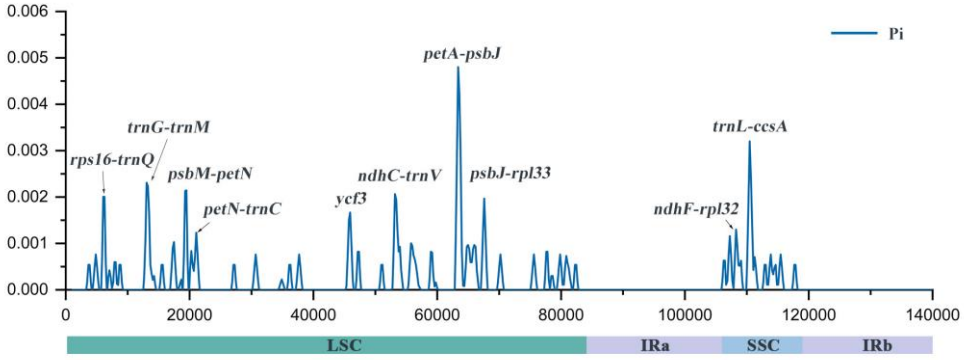


Figure 5. Nucleotide diversity (Pi) based on sliding window analysis in the aligned complete chloroplast genomes of *S. spontaneum*, *S. officinarum*, *S. robustum*, *S. sinense*, *S. barberi*, and *S.* hybrids.

Based on the alignment of complete cp genome sequences, nine most variable loci were selected as candidate DNA markers for validation (Figure 6a). Markers were derived from InDels in the intergenic regions including *rps16-trnQ*, *trnS-psbD*, *trnS-psbZ*, *trnR-rps14*, *ndhC-trnV*, *petL-petG*, *rpl33-rps18*, and *rps12* introns (Table 2). These nine markers were classified into three categories based on their differentiation ability. For the first category, SscpId4 in the *trnR-rps14* intergenic region was able to differentiate each of the three species, *T. arundinaceum*, *S. spontaneum*, and *S. officinarum*, having 188bp, 172bp, 198bp expected amplified fragment size respectively. For the second category, the marker SscpId7 (15bp InDel in the *rpl33-rps18* region) can differentiate *S. spontaneum* and *T. arundinaceum* from other *Saccharum* species. For the third category, all the rest markers could differentiate *T. arundinaceum* from *Saccharum* species, having 29bp, 31bp, 34bp, 21bp, 46bp, 18bp, and 33bp fragment length difference for SscpId1, SscpId2, SscpId3, SscpId5, SscpId6, SscpId8 and SscpId9, respectively. Validation experiments using six samples, two accessions from each species showed the power of these markers to separate the target samples (Figure 6b). These markers were valuable tools for accurate germplasm evaluation in sugarcane research and breeding programs.

Table 2. Information of nine InDel markers designed for *T. arundinaceum, S. spontaneum*, and *S. officinarum*.

Primer	Forward Reverse	Region	PCR product size (bp)		
			T. aru	*S. spon*	*S. off*
SscpId1	**GGATCGACTCTTTCCCAACA** **CCCAAAAACGACCAATCTTT**	*rps16-trnQ*	130	159	159
SscpId2	TGACAGAGGCAAGAAATAACGA GCCCGCTGAAGAGAAAATC	*trnS-psbD*	235	204	204
SscpId3	AAGACCGGAGCTATCAACCA TGAAAGTGGACTTGCATCTGA	*trnS-psbZ*	157	123	123
SscpId4	CCATTAGACAATGGACGCTTCT TGCTCCATATTCCTTCTTTATGA	*trnR-rps14*	188	172	198
SscpId5	AACCTAATGAAAATCGGATGATTA GCGAATTCCTATTTTGTTTGAA	*ndhC-trnV*	249	228	228
SscpId6	AGGGACTCATGTTCCGTCTG GCATGAAGGGGTTAAATTCC	*petL-petG*	191	145	145
SscpId7	TCGTGTGTTCGATTTTTCCA TTTTGATCCAGAACCAGAAGAA	*rpl33-rps18*	235	235	250
SscpId8	TTTGGCAAAGAAAAATAGAGTACG ATTCCTCAAAACGAGGCTCA	*rps12 intron*	155	173	173
SscpId9	TTTGGCATTAGTTCATGAGGA AAGGAAATGGCTATCAAGAACG	*rps12 intron*	143	176	176

[1] *T. aru : T. arundinaceum; S. spon: S. spontaneum; S. off: S. officinarum.*

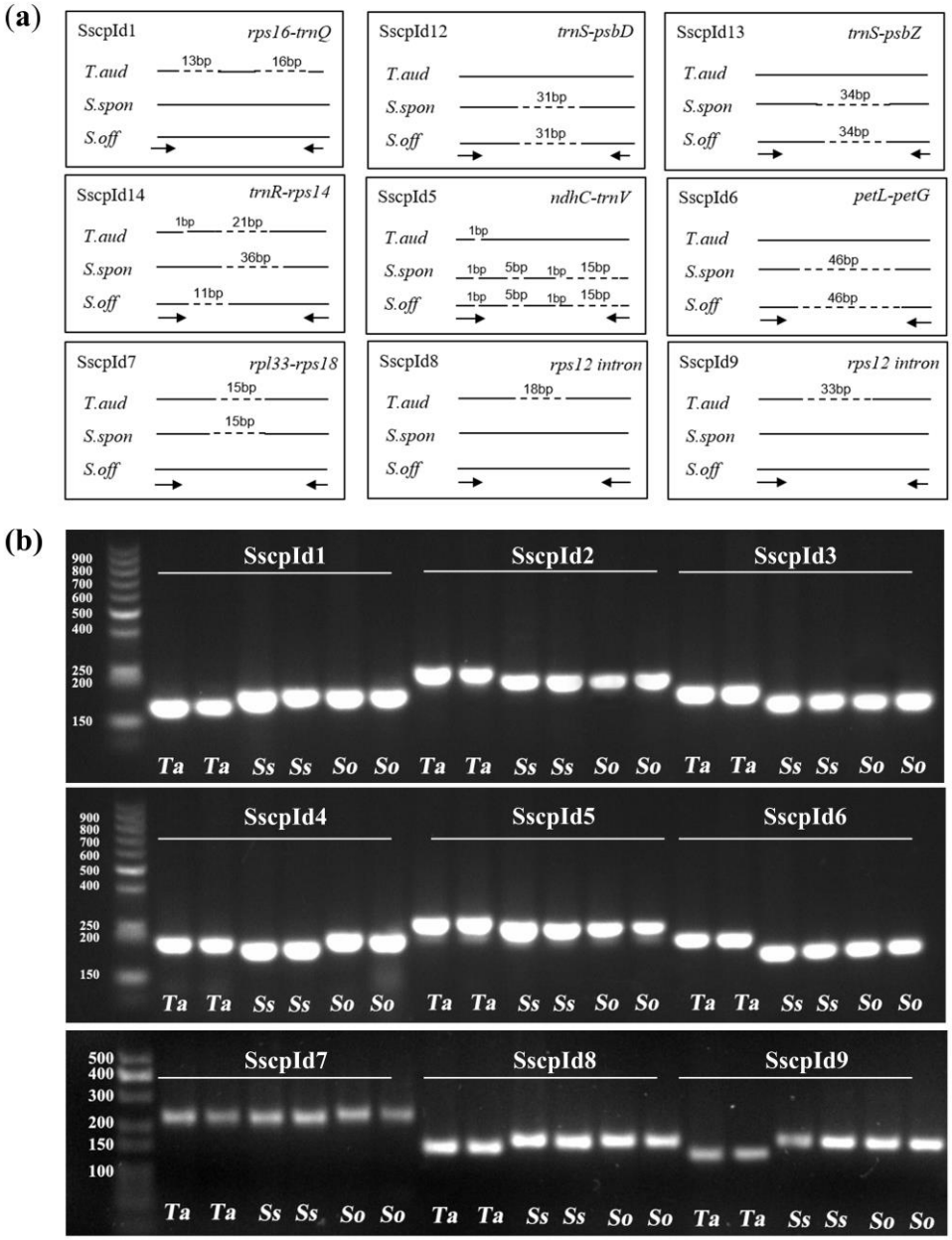


Figure 6. The gel electrophoresis results of the validation of nine molecular markers derived from InDels regions of three chloroplast genomes of *T. arundinaceum*, *S. spontaneum*, and *S. officinarum*. (a) The dashed line indicated deleted sequences. Left and right black arrows indicated forward and reverse primers, respectively. (b) The gel electrophoresis results from the amplification of InDels markers using designed primers. Abbreviated species names were shown on schematic diagrams, *Ta*, *T. arundinaceum*; *Ss*, *S. spontaneum*; *So*, *S. officinarum*.

Table 3. dCAPS makers design based on SNPs between *S. robustrum* and *S. officinarum*.

Primer	Forward Reverse	Region	Digestion enzyme	PCR product size(bp)/Target SNP	
				S. rob	*S. off*
Sscpd1	**CGTACTCCTAATCGAATTTGTATTTTC** **AGTGGTAAAAGTGTGATTCGTTC**	***rps16-trnQ***	*Taq*I	226/A	199/G
Sscpd2	CGTACTCCTAATCGAATTTGTATTCTA TGGTAAAAGTGTGATTCGTTCTATT	*trnS-psbD*	*Xba*I	224/A	197/G
Sscpd3	AACAAAAGAATAAATCCAGGGATTCCA CGGGTCCTTACTCCCCTTTA	*trnS-psbZ*	*Taq*I	158/C	174/G
Sscpd4	TCTTCTTTATTCTTCGAATTGATTGC TGCGTATTTGATTCCATTATCGT	*trnR-rps14*	*Taq*I	216/A	201/G

[1] *S. rob: S. robustrum; S. off: S. officinarum*

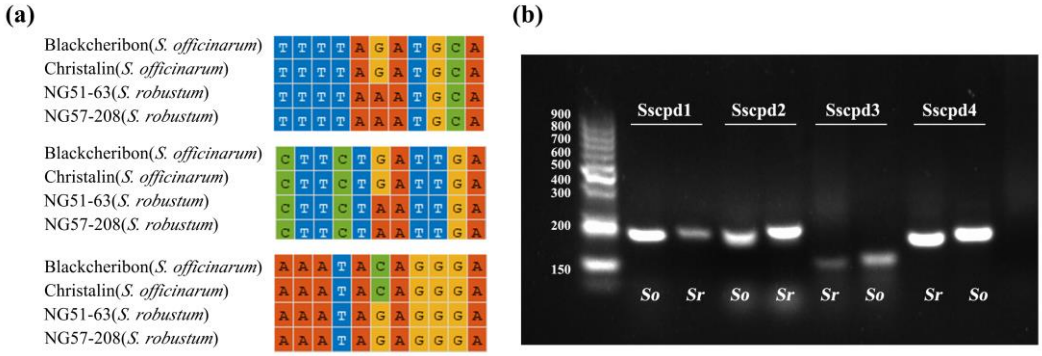


Figure 7. dCAPS makers design based on SNPs between *S. robustrum* and *S. officinarum*. (a) The multiple alignments of the chloroplast genomes by MAFFT, (b) SNP analysis with dCAPS primers, Sscpd1 and Sscpd2, designed for the SNP site in *trnG-trnG* region, with *TaqI* and *XbaI* restriction sites, respectively. Sscpd3 and Sscpd4, were designed for SNPs in *petA-psbJ* and *rbcL-psaI* intergenic region, with *Taq*I restriction sites. Abbreviated species names were shown on schematic diagrams, Ss, *S. spontaneum*, Sr, *S. robustum*, So, *S. officinarum*.

The species-specific SNPs between *S. robustum* and *S. officinarum* were used to develop dCAPS markers to generate the mismatches near SNPs for the recognition site of restriction endonuclease (Table 3). One SNP in *trnG-trnG* intergenic regions ("G" for *S. officinarum*, and "A" for *S. robustum*) was used to design two dCAPS markers Sscpd1 and Sscpd2 (Figure 7a), with *TaqI* and *XbaI* digestion of the amplicon produced a digested fragment except in *S. officinarum*. Sscpd3 and Sscpd4, designed for SNPs in *petA-psbJ* ("G" for *S. officinarum*, and "A" for *S. robustum*) and *rbcL-psaI* ("C" for *S. officinarum*, and "G" for *S. robustum*) intergenic region, with *TaqI* digestion of the amplicon produced a digested fragment except in *S. robustum*. The validation showed the markers could separate *S. officinarum* (Badila) and *S. robustum* (NG57-208) clearly by fragment length difference (Figure 7b), indicating that the dCAPS markers worked for discrimination of *S. robustum* and *S. officinarum*.

Table 4. dCAPS maker design based on SNPs between *S. sinense*/*S. barberi* and other *Saccharum* spp.

Primer	Forward Reverse	Region	Digestion enzyme	PCR product size(bp)/Target SNP	
				S. sin/ S. bar	*Other S. spp.*
Sscpd5	TCTTCTTTATTCTTCGAATTGATTGC TGCGTATTTGATTCCATTATCGTTCAT	*rbcL- psaI*	*Taq*I	201/A	216/G

[1]*S. sin: S. sinense*; *S. bar: S. barberi*; *S.* spp. : *Saccharum* spp. *(included Sh, S. hybrid; So, S. officinarum; Sr, S. robustum;* Ss, *S. spontaneum.)*

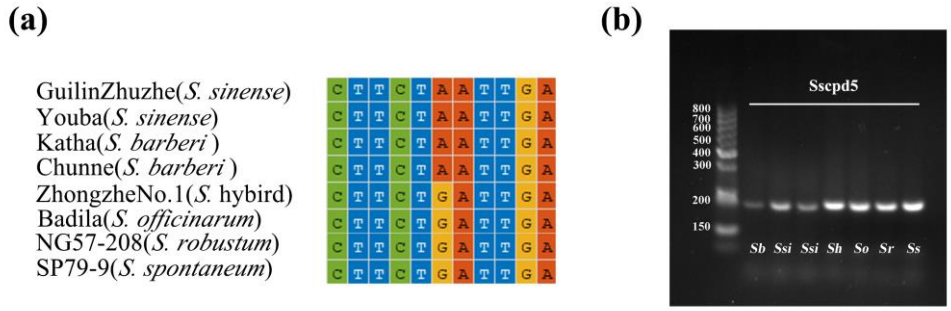


Figure 8. The gel electrophoresis results of the amplification of DNA barcodes using designed primers. (a) The multiple alignments of the chloroplast genomes by MAFFT, (b) SNP analysis with dCAPS primers, Sscpd5 designed for the SNP site in *rbcL-psaI* region with *TaqI* restriction sites. Abbreviated species names were shown on schematic diagrams, *Sb, S. barberi; Ssi, S. sinense; Sh, S. hybrid; So, S. officinarum; Sr, S. robustum;* Ss, *S. spontaneum.*

The species-specific SNPs between *S. sinense*, *S. barberi* and other *Saccharum* spp. (including *S.* hybrid, *S. officinarum*, *S. robustum* and *S. spontaneum* which is closely related to *S. sinense* and *S. barberi*) were used to develop dCAPS marker (Table 4). Sscpd5 designed for SNPs in *rbcL-psaI* ("A" for *S. sinense* and *S. barberi*, and "G" for other *Saccharum* spp.) intergenic region, with *TaqI* digestion of the amplicon produced a digested fragment except in *S. sinense* and *S. barberi*. The validation showed the markers could separate ancient cultivars *S. sinense* and *S. barberi* from modern sugarcane hybrids, *S. officinarum*, and *S. robustum* clearly by fragment length difference (Figure 8b), indicating that the dCAPS marker worked for discrimination of *S. sinense*/*S. barberi* with other *Saccharum* spp..

3. Discussion

We firstly sequenced and assembled *Saccharum* species *S. barberi* (Katha), *S. sinense* (Youba and GuilinZhuzhe), and *S. robustum* (NG57-208 and NG51-63). A total of 23 cp genomes of Saccharinae subfamily were presented, including 21 *Saccharum* species, two *Saccharum*-related species, *N. porphyrocoma* (HebawangNo1) and *T. arundinaceum* (BM87-36), which enriched and improved the chloroplast database of Saccharinae. The chloroplast sequences of *Saccharum* species showed high variation in their intergenic regions. We took advantage of these sequence polymorphisms in the chloroplast genomes, and nine InDels markers and five dCAPS markers were successfully developed. Those markers not only enriched the marker for species discrimination in Saccharinae subfamily, and also had great potential for use in distinguishing progeny of crosses generated in sugarcane breeding program.

3.1. Differences in gene and structure of chloroplast genomes among Saccharinae species

For the first time, the chloroplast genomes of the representative six species in *Saccharum* genera and relative species *T. arundinaceum* and *N. porphyrocoma* were described, which provided important insights into the characteristics of plastid genomes of the members of Saccharinea. Our results showed that the chloroplast genomes of *Saccharum* species were extremely similar, and the total genome size was similar to those of most plant cp genomes reported previously[14,17-20], indicating their high level of conservation, especially in terms of chloroplast genome organization, gene content, and SSRs. The expansion and contraction of IR regions were the main cause of variation in cp genome size[21]. We had compared the cp IR boundary regions of reported plastomes to those of closely related species, and found slight difference between closely related species. The six chloroplast sequences of *Saccharum* species showed higher variation in non-coding regions than in their coding regions. Some of the mutation hotspot regions, including *rps16-trnQ*, *trnG-trnM*, *psbM-petN*, *petN-trnC*, *ndhC-trnV*, *petA-psbJ*, *psbJ-rpl33*, *ndhF-rpl32*, and *trnL-ccsA*, were detected from the cp genomes, which were potential regions for marker development for species discrimination.

3.2. Plastid view on the evolutionary relationships within Saccharinae

Within genera *Saccharum*, *S. spontaneum* diverged from the other *Saccharum* species firstly, and *S. officinarum* diverged from a common ancestor with *S. robustum*. *S. sinense* and *S. barberi* were clustered as a monophyletic group with *S. officinarum* and *S. robustum* linage, suggesting that their maternal origin might be the common ancestors. *S. sinense* and *S. barberi* were distant to *S. robustum* than *S. officinarum* and modern cultivar sugarcane. As expected, high polymorphisms within species observed in *S. spontaneum* and *S. robustum* clade, maybe due to they were wild species and had widely ecological distributions. Low polymorphisms were observed in branches *S. sinense* and *S. barberi*, indicating their limited maternal origin.

The existing consensus within *Saccharum* genera was that *S. officinarum* originated from *S. robustum*, while *S. sinense* and *S. barberi* were believed to have arisen by local/natural hybridization of *S. officinarum* with *S. spontaneum*. Recent studies[22] based on homologous haplotypes suggested that *S. offcinarum* was presumed to be derived from *S. robustum* 0.64 Mya ago, and *S. officinarum* and *S. robustum* were derived from interspecific hybridization between two unknown ancestors[13]. The phylogram presented in current study confirmed that *S. officinarum* shared a common ancestor with *S. robustum*, but the true natural evolutionary processes were still in the mist. Our results demonstrated that *S. sinense* and *S. barberi* might be from a single-origin natural hybridization, and they might have arisen earlier than previously inferred (4000BP)[13], because their cytoplasmic origin was much ancient than the *S. offcinarum*/*S. robustum* linage. However, more evidences were required to corroborate this conclusion.

Within the '*Saccharum* complex', *T. arundinaceum* and *N. porphyrocoma* were considered as excellent gene resources for sugarcane variety improvement. Many closely related species have been employed within sugarcane breeding programs, but the information on the phylogeny of the core '*Saccharum* complex' is still rather limited, especially these species (like *Erianthus rockii* Keng.) closely relative to be used in introgressing process with sugarcane. Previous studies had suggested that *Tripidium* and *Saccharum* were divergent eight million years ago, last sharing a common ancestor 12 million years ago[15]. However, the phylogram in this study showed that *T. arundinaceum* was sister to *E. aurea*, which was closely related to *Miscanthus*. The distant phylogenetic relationship found between sugarcane and *Tripidium* suggested that these taxa may be less cross-compatible with sugarcane than the species of *Erianthus* and *Miscanthus*, therefore not ideal for genetic crosses with sugarcane[19]. Recently, comparison of chloroplast and ITS phylogenies revealed that *Erianthus* was monophyletic with *Narenga* and *Saccharum*[17]. Therefore, these accessions should be the preferred classified as genus '*Tripidium*' instead of its traditional genus

'*Erianthus*'[16]. Phylogenomic analysis had good branch support that *N. porphyrocoma* typically emerges as sister to *Saccharum*, demonstrating that *Narenga* is more closely related to *Saccharum* than *Tripidium*.

3.3. Development of markers at the inter- and intra-species level for Saccharum

The cytoplasm plays a significant role in the inheritance and realization of agronomically important traits. To date, many examples of the influence of the cytoplasmic background on numerous plant traits, including those of great biological and economic importance, have been documented[23]. The identification of correct parental germplasm and verifying the cytoplasm origin of their progeny is crucial to sugarcane breeding. Matrilineal inheritance of chloroplasts makes it possible to detect maternal-specific SNPs, thus enabling the identification of maternal parents[24].

Previously, researchers have attempted to utilize cytogenetic tools such as GISH[25] or the detection of the progeny of hybrid populations from intra- or inter-species cross[26-28], and identification of *Saccharum* germplasm materials. But GISH is very expensive, laborious, and time-consuming. Instead, species-specific molecular markers were favored to be utilized. At the beginning, instance ribosomal restriction fragment length polymorphism (RFLP) markers were used to distinguish *S. sinense* and *S. barberi* accessions, but its detectable variations were limited within species[29]. Then, PCR-based markers such as Random Amplification of Polymorphic DNA (RAPD) markers and simple sequence repeats (SSR) or microsatellite markers were applied. RAPD markers were used to identify "*Saccharum officinarum* × *Erianthus fulvus*" true F1 hybrids[30], but it proved to be less reproducibility across labs. SSR markers are considered one of the most efficient markers, which had been utilized greatly to identify true hybrids of sugarcane[31-36]. Nevertheless, SSRs have some dis-advantages over SNPs/InDels-based marker such as low mutation rate.

The large and polyploid genome of sugarcane is a challenge for markers development, but the cp genome provides shortcut to be a useful target for developing DNA markers to verify species. For the development of DNA markers using sequence variations between species with similar genetic backgrounds, it would be efficient to use sequence variations, such as SNPs and InDels. dCAPS markers had been proved to have important applications in the analyses of genetic polymorphism and phylogeny in closely related species[24,37]. To date, no SNPs or InDels markers for species discrimination have been developed in the sugarcane cp genome. In this study, we successfully develop 14 valuable and unique polymorphic markers to distinguish *Saccharum* species. Most of these markers were derived from the intergenic regions of cp genomes and showed high interspecific polymorphisms. In the intra-species level, four SNP-based dCAPS makers discriminated *S. officinarum* from *S. robustum*, one dCAPS markers for differentiated *S. barberi* and *S. sinense* from other *Saccharum* species, and two InDels based makers discriminated *S. officinarum* from *S. spontaneum*. In the inter-species level for *Saccharum*, eight InDels-based makers discriminated *Saccharum* spp. from *T. arundinaceum*. As a result, those markers were fragment-length-difference-based (InDels) or SNP-transformed dCAPs, which were easy and cheap to run for sample discrimination. Besides, these were very fast to generate, had a good repeatability, and were easy to transfer between populations or laboratories. These markers were not only possible to utilize for germplasm evaluation, but they also had great potential for use in distinguishing progeny of reciprocal crosses in sugarcane breeding.

4. Materials and Methods

4.1. DNA extraction, library construction, and sequencing

Fresh leaves of 16 samples were collected from Guangxi University (Nanning, Guangxi, China) (108°19′ E, 22°51′ N, 530,023) (Table S1) and stored in liquid nitrogen. After the DNA extraction from fresh leaf tissues, its quantification was validated using

Agarose gel electrophoresis and Nanodrop concentration 500 bp randomly interrupted by the Covaris ultrasonic breaker for library construction. DNA library and sequenced by Illumina NovoSeq 6000 sequencing platform (Illumina, San Diego, CA). Approximately, 10.0 GB of raw data were generated with 150 bp paired-end read lengths.

4.2. Plastome assemblies and annotation

Fastp (v0.21.0)[38] was used for quality control, adapter trimming, quality filtering, and per-read quality pruning. The trimmed reads were subjected to GetOrganelle (v1.7.5)[39] to generate plastome assemblies as suggested for Embryophyta plant plastome sequences (-R 10; -F embplant_pt). The SPAdes[40] kmer settings were set to -k 21, 45, 65, 85, or 105. The contig coverage information and other graph characteristics are used by GetOrganelle to construct the final assembly graphs, which were plotted and visually assessed using Bandage (v0.8.1)[41]. The completed chloroplast assembly was re-aligned to *S. officinarum* cultivar badila (GenBank: MG685915.1) to confirm the quadripartite structure representing - LSC, SSC, and two IRs.

Genome annotation and visualization were performed using web server CPGAVAS2[42], CDS and rRNA annotation based on BLAST search, while tRNAscan-SE v2.0.2[43] was used to find and annotate tRNA genes. Then manually corrected using the Apollo[44]. Last, Chloroplot[45] was used to generate a corrected cp circular map.

4.3. Phylogenetic Analyses

For phylogenetics analysis for Saccharinae species, 14 published plastome sequences were achieved. *Sorghum bicolor* (NC_008602.1) was used as outgroups. Fasta files with the nucleotide sequences for all 74 PCGs shared from were extracted from the GenBank files using PhyloSuite(v1.2.2)[46]. All the genes were aligned in batches with MAFFT (v7.313)[47] integrated into PhyloSuite, using normal-alignment mode. The best partitioning scheme and evolutionary models for 74 pre-defined partitions were selected using PartitionFinder2[48], with greedy algorithm and AICc criterion, which were divided into 27 subsets. PhyloSuite was then used to concatenate these alignments into a single alignment and generate phylip and nexus format files for the phylogenetic analysis, conducted using maximum likelihood (ML) methods. The selection of the best-fit partition strategy and models was carried out using PartitionFinder2. Maximum likelihood phylogenies were inferred using IQ-TREE[49] under Edge-unlinked partition model for 50,000 ultrafast[50] bootstraps, as well as the Shimodaira-Hasegawa-like approximate likelihood-ratio test[51]. Images were drawn with iTOL[52] and Adobe Illustrator.

4.4. Plastome polymorphism

To explore the sequence divergence and hotspot regions, DnaSP(v6.0)[53] was used for sliding window analysis for computing nucleotide diversity (pi) among the chloroplast genome sequences, with 500 bp windows size and 250 bp step size.

For variation analysis, clean reads were aligned to *S. officinarum* cultivar badila (GenBank: MG685915.1) cp genome v3.0 using BWA-mem[54] with default settings. For variant calling, we used the GATK HaplotypeCaller v.4.1.8.1[55] software with the '-ploidy 1' parameter to compare SNPs and small indels from BAM files. We filtered out low-quality variants ("QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0") in GATK.

The microsatellites were determined by perl script MIcroSAtellite (MISA)[56] with search parameters: 1-10; 2-5; 3-5; 4-4; 5-3; 6-3 where 1, 2, 3, 4, 5, and 6 indicate the mono-di-, tri-, tetra-, penta-, and hexa-nucleotide repeats.

4.5. Comparative genome alignment

Aligned sequences and annotations for assembled chloroplast genomes were used to construct sequence conservation plots in the program mVISTA[57] and were compared

using mVISTA software in Shuffle-LAGAN mode with the *S. bicolor* as a reference. IR-scope[58] was used to detect IR border expansion and contraction.

4.6. Development and validation of InDel and dCAPS markers

To discriminate among the Sacchariane species, specific primers were designed using Primer 3 based on the mutational hotspot regions found in *Saccharum* and *Tripidium arundinaceum* chloroplast genomes. The specific primers for InDel regions and dCAPS primers for SNP sites were designed using dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html) and the Primer 3 program (http://bioinfo.ut.ee/primer3-0.4.0/), respectively.

Polymerase chain reaction (PCR) was carried out in a 20 μl reaction mixture containing 10 μl of 10× reaction buffer, 5 pmol of each primer, 1.25 units of Taq DNA polymerase, and 20 ng of DNA template. The PCR reaction was performed in thermocyclers using the following cycling parameters: 94°C (5 min); 35 cycles of 94°C (30 s), 56–62°C (30 s); 72°C (30 s), then 72°C (7 min). The PCR products of each dCAPS marker were treated with a restriction enzyme at 37°C (*Xba*I) or 65 °C (*Taq*I) for 1 h. PCR products were visualized on agarose gels (2.0–3.0%) containing safe gel stain.

5. Conclusions

We sequenced and assembled the chloroplast sequences of six species in the genera *Saccharum* and relative species *T. arundinaceum* and *N. porphyrocoma*. Our results enriched, updated, and improved the database genomics of Saccharinae, furthermore confirming the phylogeny of Saccharinae based on available cp genomes. Those results can be used to elucidate evolutionary history such as the origin of Saccharinea, allowing for a deeper understanding of their parental lineages.

In addition, we developed 14 robust molecular markers, including nine InDels markers for discrimination of *S. spontaneum*, *S. officinarum*, *T. arundinaceum*, five dCAPS markers for *S. officinarum* and *S. robustum*, and one dCAPS markers for differentiated *S. sinense* and *S. barberi* from other *Saccharum* species. These markers had been demonstrated to have high discriminatory power in the *Saccharum* genera. Taken together, the results obtained from our study will contribute to the understanding of Saccharinae classification, and plastome evolution at inter- and intra-species levels of *Saccharum* genera. Molecular markers developed in this study were inexpensive and efficient, and thus were helpful to accelerate the process of marker-assisted breeding in sugarcane for parental selection and cross verification.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1

Author Contributions: Conceptualization, SL and XY; validation, MF, NW, and KB; data curation, SL, MF, and JZ; writing—original draft preparation, SL, JZ and XY; writing—review and editing, XY; funding acquisition, XY, BC, and DW. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 'One Hundred Person' Project of Guangxi Province, Science, Guangxi Natural Science Foundation (GK AD20297064), Technology Major Project of Guangxi Province (GK2018-266-Z01), National Natural Science Foundation of China (31901510) and National Key Research and Development Program (2021YFD1200204-05).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. James, G.; Tate, B. *Sugarcane*; Blackwell Science: Oxford, 2004.
2. Yu, F.; Chai, J.; Li, X.; Yu, Z.; Yang, R.; Ding, X.; Wang, Q.; Wu, J.; Yang, X.; Deng, Z. Chromosomal Characterization of *Tripidium arundinaceum* Revealed by Oligo-FISH. *Int J Mol Sci* **2021**, *22*, 8539, doi:10.3390/ijms22168539.
3. Chang, H.; Wang, Q.; Qiu, Y.; Qin, Y.; Li, X.; Wu, Q.; He, W.; Guo, Y.; Zhang, W.; Chen, J.; et al. Production, identification and characterization of *Erianthus rockii* × *Narenga porphyrocoma* intergeneric hybrids as a new germplasm for sugarcane breeding and genetic research. *Sugar Tech* **2020**, *22*, 389-395, doi:10.1007/s12355-020-00804-3.
4. Liu, X.; Zhang, R.; Ou, H.; Gui, Y.; Wei, J.; Zhou, H.; Tan, H.; Li, Y. Comprehensive transcriptome analysis reveals genes in response to water deficit in the leaves of *Saccharum narenga* (Nees ex Steud.) hack. *BMC Plant Biol* **2018**, *18*, 250, doi:10.1186/s12870-018-1428-9.
5. Qi, Y.; Gao, X.; Zeng, Q.; Zheng, Z.; Wu, C.; Yang, R.; Feng, X.; Wu, Z.; Fan, L.; Huang, Z. Sugarcane breeding, germplasm development and related molecular research in China. *Sugar Tech* **2022**, *24*, 73-85, doi:10.1007/s12355-021-01055-6.
6. A DNA barcode for land plants. *Proc Natl Acad Sci U S A* **2009**, *106*, 12794-12797, doi:10.1073/pnas.0905845106.
7. Kim, T.-H.; Yoon, Y.-M.; Park, J.-C.; Park, J.-H.; Kim, K.-H.; Kim, Y.-K.; Son, J.-H.; Park, T.-I. *De Novo* Transcriptome Assembly and SNP Discovery for the Development of dCAPS Markers in Oat. *Agronomy* **2022**, *12*, 184.
8. Nishijima, R.; Yoshida, K.; Sakaguchi, K.; Yoshimura, S.I.; Sato, K.; Takumi, S. RRNA sequencing-based bulked segregant analysis facilitates efficient D-genome marker development for a specific chromosomal region of synthetic hexaploid wheat. *Int J Mol Sci* **2018**, *19*, doi:10.3390/ijms19123749.
9. Zhao, B.; Wang, B.; Li, Z.; Guo, T.; Zhao, J.; Guan, Z.; Liu, K. Identification and characterization of a new dwarf locus DS-4 encoding an Aux/IAA7 protein in *Brassica napus*. *Theoretical and Applied Genetics* **2019**, *132*, 1435-1449, doi:10.1007/s00122-019-03290-8.
10. Asano, T.; Tsudzuki, T.; Takahashi, S.; Shimada, H.; Kadowaki, K.-i. Complete Nucleotide Sequence of the Sugarcane (*Saccharum Officinarum*) Chloroplast Genome: A Comparative Analysis of Four Monocot Chloroplast Genomes. *DNA research : an international journal for rapid publication of reports on genes and genomes* **2004**, *11*, 93-99, doi:10.1093/dnares/11.2.93.
11. Calsa Júnior, T.; Carraro, D.M.; Benatti, M.R.; Barbosa, A.C.; Kitajima, J.P.; Carrer, H. Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Current Genetics* **2004**, *46*, 366-373, doi:10.1007/s00294-004-0542-4.
12. Vidigal, P.M.; Coelho, A.S.; Novaes, E.; Barbosa, M.H.; Peternelli, L.A. Complete Chloroplast Genome Sequence and Annotation of the *Saccharum* Hybrid Cultivar RB867515. *Genome Announc* **2016**, *4*, doi:10.1128/genomeA.01157-16.
13. Evans, D.L.; Joshi, S.V. Complete chloroplast genomes of *Saccharum spontaneum*, *Saccharum officinarum* and *Miscanthus floridulus* (Panicoideae: Andropogoneae) reveal the plastid view on sugarcane origins. *Systematics and Biodiversity* **2016**, *14*, 548-571, doi:10.1080/14772000.2016.1197336.
14. Xu, F.; He, L.; Gao, S.; Su, Y.; Li, F.; Xu, L. Comparative analysis of two sugarcane ancestors Saccharum officinarum and S. spontaneum based on complete chloroplast genome sequences and photosynthetic ability in cold stress. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20153828.
15. Lloyd Evans, D.; Joshi, S.V.; Wang, J. Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoideae: Andropogoneae) to sugarcane. *BMC evolutionary biology* **2019**, *19*, 33, doi:10.1186/s12862-019-1356-9.
16. Tsuruta, S.I.; Ebina, M.; Kobayashi, M.; Takahashi, W. Complete Chloroplast Genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative Genomics and Evolution of the *Saccharum* Complex. *PLoS One* **2017**, *12*, e0169992, doi:10.1371/journal.pone.0169992.

17. Evans, D.L.; Hughes, B. Complete Chloroplast Genomes of *Saccharum giganteum*, *Saccharum longisetosum*, *Cleistachne sorghoides*, *Sarga timorense*, *Narenga porphyrocoma* and *Tripsacum dactyloides*. Comparisons with ITS phylogeny and Placement within *Saccharum*. *bioRxiv* **2020**, 2020.2006.2012.149476, doi:10.1101/2020.06.12.149476.
18. Lloyd Evans, D.; Hlongwane, T.T.; Joshi, S.V.; Riaño Pachón, D.M. The sugarcane mitochondrial genome: assembly, phylogenetics and transcriptomics. *PeerJ* **2019**, *7*, e7558, doi:10.7717/peerj.7558.
19. Welker, C.A.D.; McKain, M.R.; Vorontsova, M.S.; Peichoto, M.C.; Kellogg, E.A. Plastome phylogenomics of sugarcane and relatives confirms the segregation of the genus *Tripidium* (Poaceae: Andropogoneae). *TAXON* **2019**, *68*, 246-267.
20. Lloyd Evans, D.; Hughes, B.; Joshi, V. Comparative whole plastome and low copy number phylogenetics of the core Saccharinae and Sorghinae. *bioRxiv* **2022**, doi:10.1101/2022.01.04.474895.
21. Huang, H.; Shi, C.; Liu, Y.; Mao, S.-Y.; Gao, L.-Z. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC evolutionary biology* **2014**, *14*, 151, doi:10.1186/1471-2148-14-151.
22. Pompidor, N.; Charron, C.; Hervouet, C.; Bocs, S.; Droc, G.; Rivallan, R.; Manez, A.; Mitros, T.; Swaminathan, K.; Glaszmann, J.-C.; et al. Three founding ancestral genomes involved in the origin of sugarcane. *Annals of botany* **2021**, *127*, 827-840, doi:10.1093/aob/mcab008.
23. Elkonin, L.; Kibalnik, O.; Zavalishina, A.; Gerashchenkov, G.; Rozhnova, N. Genetic functions of cytoplasm in plants with special emphasis on sorghum. 2018; pp. 97-154.
24. Zhang, Y.; Fang, Z.; Wang, Q.; Liu, Y.; Yang, L.; Zhuang, M.; Sun, P. Chloroplast subspecies-specific SNP detection and its maternal inheritance in *Brassica oleracea* L. by using a dCAPS marker. *Journal of Heredity* **2012**, *103*, 606-611, doi:10.1093/jhered/ess006.
25. Li, X.; Guo, Y.; Huang, F.; Wang, Q.; Chai, J.; Yu, F.; Wu, J.; Zhang, M.; Deng, Z. Authenticity Identification of *Saccharum officinarum* and *Saccharum spontaneum* Germplasm Materials. *Agronomy* **2022**, *12*, 819.
26. D'Hont, A.; Rao, P.S.; Feldmann, P.; Grivet, L.; Islam-Faridi, N.; Taylor, P.; Glaszmann, J.C. Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* × *Erianthus arundinaceus*, with molecular markers and DNA *in situ* hybridisation. *Theor Appl Genet* **1995**, *91*, 320-326, doi:10.1007/bf00220894.
27. Pachakkil, B.; Terajima, Y.; Ohmido, N.; Ebina, M.; Irei, S.; Hayashi, H.; Takagi, H. Cytogenetic and agronomic characterization of intergeneric hybrids between *Saccharum* spp. hybrid and *Erianthus arundinaceus*. *Sci Rep* **2019**, *9*, 1748-1748, doi:10.1038/s41598-018-38316-6.
28. Huang, Y.; Wu, J.; Wang, P.; Lin, Y.; Fu, C.; Deng, Z.; Wang, Q.; Li, Q.; Chen, R.; Zhang, M. Characterization of Chromosome Inheritance of the Intergeneric BC_2 and BC_3 Progeny between *Saccharum* spp. and *Erianthus arundinaceus*. *PLoS One* **2015**, *10*, e0133722, doi:10.1371/journal.pone.0133722.
29. Wood, B. Restriction fragment length polymorphism distinguishes members of *Saccharum sinense* and *S. barberi*. *Plant Physiol (suppl)* **1987**, *83*, 15.
30. Zhang, H.; Li, F.; He, L.; Zhong, H.; Yang, Q.; He, S. Identification of sugarcane interspecies hybrids with RAPDs. *African Journal of Biotechnology* **2008**, *7*.
31. Manigbas, N.L.; Villegas, L.C. Microsatellite markers in hybridity tests to identify true hybrids of sugarcane. *Philippine Journal of Crop Science* **2004**, *29*, 23-32.
32. Todd, J.; Pan, Y.-B.; Boykin, D. Fidelity of sugarcane crosses assessed with SSR markers. *Agronomy* **2020**, *10*, 386.
33. Ahmad, A.; Wang, J.-D.; Pan, Y.-B.; Sharif, R.; Gao, S.-J. Development and use of simple sequence repeats (SSRs) markers for sugarcane breeding and genetic studies. *Agronomy* **2018**, *8*, 260.
34. Santos, J.; Barbosa, G.; Neto, C.; Almeida, C. Efficiency of biparental crossing in sugarcane analyzed by SSR markers. *Crop Breeding and Applied Biotechnology* **2014**, *14*, 102-107, doi:10.1590/1984-70332014v14n2a18.

35. Xavier, M.A.; Pinto, L.R.; Fávero, T.M.; Perecin, D.; Carlini-Garcia, L.A.; Landell, M.G. Paternity identification in sugarcane polycrosses by using microsatellite markers. *Genet Mol Res* **2014**, *13*, 2268-2277, doi:10.4238/2014.March.31.7.
36. Liu, X.-H.; Fang, F.-X.; Zhang, R.-H.; Song, H.-Z.; Yang, R.-Z.; Gao, Y.-J.; Ou, H.-P.; Lei, J.-C.; Luo, T.; Duan, W.-X.; et al. Identification of progenies from sugarcane × *Narenga porphyrocoma*(Hance) Bor. by SSR marker. *Southwest China Journal of Agricultural Sciences* **2012**, *25*, 38-43.
37. Shavrukov, Y.N. CAPS markers in plant biology. *Russian Journal of Genetics: Applied Research* **2016**, *6*, 279-287, doi:10.1134/S2079059716030114.
38. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884-i890, doi:10.1093/bioinformatics/bty560.
39. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* **2020**, *21*, 241, doi:10.1186/s13059-020-02154-5.
40. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012**, *19*, 455-477, doi:10.1089/cmb.2012.0021.
41. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350-3352, doi:10.1093/bioinformatics/btv383.
42. Shi, L.; Chen, H.; Jiang, M.; Wang, L.; Wu, X.; Huang, L.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* **2019**, *47*, W65-w73, doi:10.1093/nar/gkz345.
43. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* **2021**, *49*, 9077-9096, doi:10.1093/nar/gkab688.
44. Lewis, S.E.; Searle, S.M.J.; Harris, N.; Gibson, M.; Lyer, V.; Richter, J.; Wiel, C.; Bayraktaroglu, L.; Birney, E.; Crosby, M.A.; et al. Apollo: a sequence annotation editor. *Genome biology* **2002**, *3*, RESEARCH0082-RESEARCH0082, doi:10.1186/gb-2002-3-12-research0082.
45. Zheng, S.; Poczai, P.; Hyvönen, J.; Tang, J.; Amiryousefi, A. Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Frontiers in genetics* **2020**, *11*, 576124, doi:10.3389/fgene.2020.576124.
46. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.; Wang, G. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* **2020**, *20*, 348-355, doi:10.1111/1755-0998.13096.
47. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **2013**, *30*, 772-780, doi:10.1093/molbev/mst010.
48. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular biology and evolution* **2017**, *34*, 772-773, doi:10.1093/molbev/msw260.
49. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **2015**, *32*, 268-274, doi:10.1093/molbev/msu300.
50. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution* **2013**, *30*, 1188-1195, doi:10.1093/molbev/mst024.
51. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **2010**, *59*, 307-321, doi:10.1093/sysbio/syq010.
52. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research* **2021**, *49*, W293-W296, doi:10.1093/nar/gkab301.

53. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution* **2017**, *34*, 3299-3302, doi:10.1093/molbev/msx248.
54. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* **2013**, *1303*.
55. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* **2013**, *43*, 11.10.11-11.10.33, doi:10.1002/0471250953.bi1110s43.
56. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics (Oxford, England)* **2017**, *33*, 2583-2585, doi:10.1093/bioinformatics/btx198.
57. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic acids research* **2004**, *32*, W273-W279, doi:10.1093/nar/gkh458.
58. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030-3031, doi:10.1093/bioinformatics/bty220.