

RESEARCH ARTICLE

Comparative genetic analysis of the 45S rDNA intergenic spacers from three *Saccharum* species

Yongji Huang¹, Fan Yu¹, Xueting Li¹, Ling Luo¹, Jiayun Wu^{2,3}, Yongqing Yang¹, Zuhu Deng^{1,4*}, Rukai Chen¹, Muqing Zhang⁴

1 Key Lab of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China, **2** Guangdong Key Laboratory of Sugarcane Improvement and Biorefinery, Guangzhou, China, **3** Guangdong Provincial Bioengineering Institute, Guangzhou Sugarcane Industry Research Institute, Guangzhou, China, **4** Guangxi Collaborative Innovation Center of Sugar Industries, Guangxi University, Nanning, China

* dengzuhu@163.com



OPEN ACCESS

Citation: Huang Y, Yu F, Li X, Luo L, Wu J, Yang Y, et al. (2017) Comparative genetic analysis of the 45S rDNA intergenic spacers from three *Saccharum* species. PLoS ONE 12(8): e0183447. <https://doi.org/10.1371/journal.pone.0183447>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: December 31, 2016

Accepted: August 6, 2017

Published: August 17, 2017

Copyright: © 2017 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. The IGS sequences are deposited in the Genbank database under accession numbers KX254591-KX254605.

Funding: This work was funded by National Natural Science Foundation of China (31571730, <https://isisn.nsf.gov.cn/egrantweb/>; 31401440, <https://isisn.nsf.gov.cn/egrantweb/>) and supported by the Scientific Research Foundation of Graduate School of Fujian Agriculture and Forestry University (324-1122yb021). This project is also supported

Abstract

The 45S ribosomal DNA (rDNA) units are separated by an intergenic spacer (IGS) containing the signals for transcription and processing of rRNAs. For the first time, we sequenced and analyzed the entire IGS region from three original species within the genus *Saccharum*, including *S. spontaneum*, *S. robustum*, and *S. officinarum* in this study. We have compared the IGS organization within three original species of the genus *Saccharum*. The IGS of these three original species showed similar overall organizations comprised of putative functional elements needed for rRNA gene activity as well as a non-transcribed spacer (NTS), a promoter region, and an external transcribed spacer (ETS). The variability in length of the IGS sequences was assessed at the individual, intraspecies, and interspecies levels of the genus *Saccharum*, including *S. spontaneum*, *S. robustum*, and *S. officinarum*. The ETS had greater similarity than the NTS across species, but nevertheless exhibited variation in length. Within the IGS of the *Saccharum* species, base substitutions and copy number variation of sub-repeat were causes of the divergence in IGS sequences. We also identified a significant number of methylation sites. Furthermore, fluorescent *in situ* hybridization (FISH) co-localization of IGS and pTa71 probes was detected on all representative species of the genus *Saccharum* tested. Taken together, the results of this study provide a better insight into the structure and organization of the IGS in the genus *Saccharum*.

Introduction

Sugarcane, belonging to the genus *Saccharum* in the family *Poaceae* and the tribe *Andropogoneae*, is mainly cultivated in tropical and subtropical regions of both hemispheres [1]. Earlier classification of species within this genus has been based on morphological, cytological, and biochemical characteristics [2]. However, the taxonomy within the genus *Saccharum* remains controversial. Over the past two decades, molecular cytogenetics and markers have

by the Natural Science Foundation of Fujian Province of China (2016J01094, <http://xmgl.fjkjt.gov.cn/>) and Natural Science Foundation of Guangxi (2014GXNSFFA118002, <https://isisn.nsf.gov.cn/egrantweb/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

contributed to inference of the phylogenetic relationship of the species within the genus *Saccharum* [3–7], and therefore the classification of the species within this genus has gradually been elucidated. The currently accepted taxonomy for the genus *Saccharum* comprises six species, including *Saccharum officinarum*, *Saccharum robustum*, *Saccharum spontaneum*, *Saccharum sinense*, *Saccharum barberi*, and *Saccharum edule* [8]. Of these, *S. spontaneum* and *S. robustum* are considered to be wild species [9]. *S. spontaneum* is widely recognized as the most primitive species within the genus *Saccharum*, whereas *S. robustum* has been postulated to be the progenitor of the high sugar content species, *S. officinarum* [10]. These three species represent the original species within the genus *Saccharum*. The chromosome number of *S. spontaneum* ranges from $2n = 40–128$ with a basic chromosome number $x = 8$, while that of *S. officinarum* and *S. robustum* is constant at $2n = 80$ with a basic chromosome number $x = 10$ [6]. The remaining three species are thought to have interspecific or intergeneric origins. *S. sinense* and *S. barberi* are believed to be derived from naturally interspecific hybridization between *S. officinarum* and *S. spontaneum* [7]. *S. edule*, which is characterized by its abortive flowers, is thought to have arisen from intergeneric crosses between *S. officinarum* or *S. robustum* and a related genus, or derived from *S. robustum* [9, 11–14].

In higher plants, nuclear ribosomal DNA (rDNA) belongs to two universal multiple tandem repeats gene families, the minor 5S rDNA and major 45S rDNA [15]. The minor 5S rDNA units consist of a 5S rRNA gene region transcribed by RNA polymerase III, a conserved transcribing region of 120 bp, and a variable non-transcribed spacer (5S rDNA NTS, hereafter referring to as the non-transcribed region in the 5S rDNA unit). The 5S rDNA consists of multiple tandemly arrayed repeats at one or several chromosomal locations throughout the genome. On the other hand, the major 45S rDNA units are located at the nucleolar organizer regions (NORs) with a high copy number and are arranged in tandem at one or several sites per haploid genome. Each 45S rDNA repeated unit is composed of the coding region for the 18S, 5.8S, and 25S ribosomal RNA (rRNA) genes separated by two internal transcribed spacers (ITS1 and ITS2). Moreover, each transcribed unit is separated by a large non-transcribed intergenic spacer (IGS). The IGS can be subdivided into the non-transcribed region (NTS, hereafter referring to the non-transcribed region in 45S rDNA unit, but not the non-transcribed region in 5S rDNA unit), and the external transcribed spacer (ETS) (Fig 1) [15]. The 45S rRNA genes, which are transcribed by RNA polymerase I, are highly conserved throughout plants, animals, and fungi. However, the functional role of the IGS region is conserved and contains the RNA polymerase I promoter and other essential regulatory elements, such as sub-repeat (SR) elements, a transcription initiation site (TIS), a transcription termination site (TTS), and enhancers.

The 45S rDNA undergoes a peculiar pattern of evolution referred to as concerted evolution [16, 17]. This is an evolution process that homogenizes all of the 45S rDNA repeat units within the array and fixes the 45S rDNA repeat units during chromosomal recombination within a sexually reproducing species. Despite the presence of concerted evolution leading to the slowly evolving 45S rRNA gene, the noncoding IGS sequence changes rapidly. Variations in both length and organization of the IGS region have been commonly observed among individuals as well as intraspecies and interspecies of several species. Notably, it has been recently reported that IGS regions in *Medicago arborea* showed heterogeneity in a single 45S rDNA locus [18]. Variability in the IGS region is thought to be the result of diversity in the numbers of SRs generated by unequal crossing over [19]. Moreover, in higher eukaryotes, DNA is methylated exclusively at 5-methyl-cytosine to guanosine in the cytosine-phosphate-guanosine (CpG) dinucleotides [20]. This modification is one of the major events causing gene repression, especially when involving the areas of the GC-rich base composition and high density of CpG

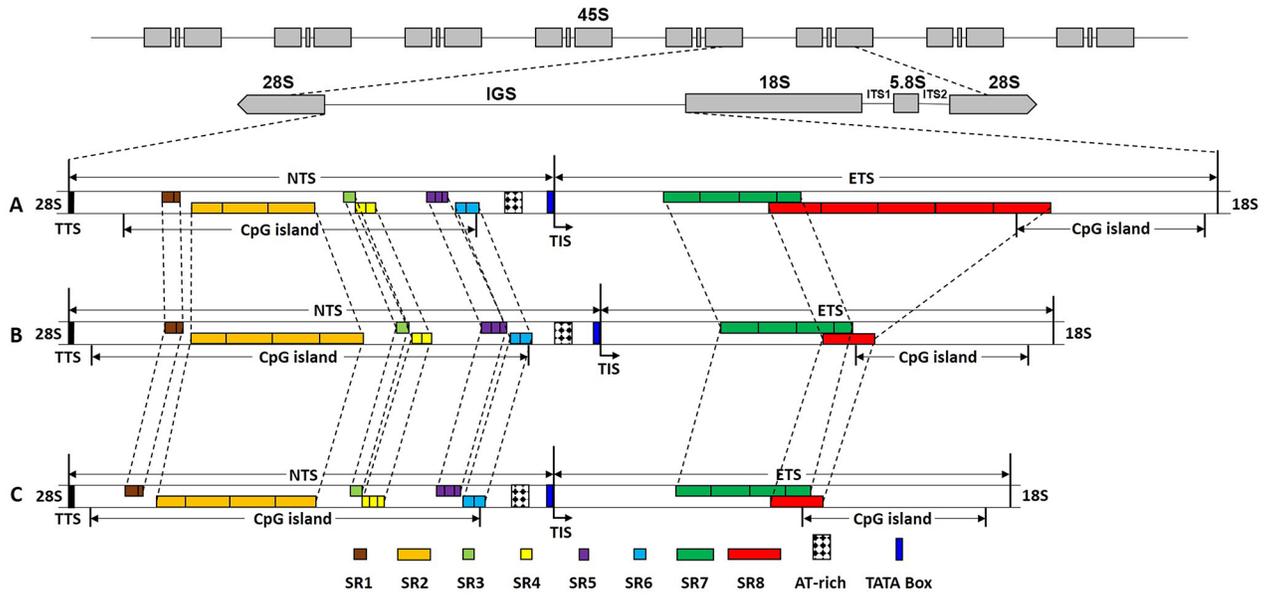


Fig 1. Canonical structural organization of the IGS regions in (A) Yunnan82-114 (*S. spontaneum*), (B) 51NG3 (*S. robustum*) and (C) Luohanzhe (*S. officinarum*). IGS: intergenic spacer; TTS: transcription termination site; NTS: non-transcribed spacer; TIS: transcription initiation site; ETS: external transcribed sequence; SR: sub-repeat. CpG island: cytosine-guanine island.

<https://doi.org/10.1371/journal.pone.0183447.g001>

dinucleotides known as CpG islands. In fact, the 5S rDNA, 45S rDNA, and IGS in eukaryotes seem to be also highly methylated [21–23].

In contrast to the evolutionary conservation in 45S and 5S rRNA genes, the spacers (ITS, IGS, and 5S rDNA NTS) are highly variable both in length and sequence [24]. These spacers are valuable for phylogenetic inference at low taxonomic levels because they show high nucleotide substitution rates and low intragenomic sequence heterogeneity [25, 26]. Indeed, the spacers have been widely used for resolving phylogenetic relationships over the past three decades [27–30]. The vast majority of studies using rDNA sequences for phylogeny and taxonomy have focused on the 45S rDNA ITS or 5S rDNA NTS. Conversely, the IGS region has been poorly investigated in the taxonomy [25, 26]. Owing to the apparent lack of selective constraints, the IGS region is considered to be the most rapidly evolving spacer region within the rDNA array and displays the highest variability. Therefore, it is suitable for phylogenetic analyses of closely related species.

IGS region plays a vital role in the regulation of 45S rDNA transcription because it is a complex modular unit containing consensus molecular structures for regulatory and processing signals of 45S rRNAs [24]. In fact, characterization of the structure and organization of IGS region could provide insight into the control of gene expression of 45S rRNA genes. So far, the sequence data of IGS region have been available for several plants of different families, including *Poaceae*, *Brassicaceae*, *Cucurbitaceae*, *Solanaceae* and *Fabaceae* [25–30]. However, to our knowledge, the comparative analysis of the IGS sequence, including the structural and functional importance of its constituents, has never been attempted among three original species of the genus *Saccharum*, including *S. spontaneum*, *S. robustum*, and *S. officinarum*. In the present study, we characterized the complete sequence of the IGS region in the genus *Saccharum* for the first time. The objectives of the present study were as follows: (1) to understand the organization of the IGS sequence among the three original species within the genus *Saccharum*; (2) to assess the individual, intraspecies, and interspecies heterogeneity in the IGS sequence; and

(3) to investigate the chromosomal location between 45S rDNA and IGS in the three original species using fluorescence in situ hybridization (FISH).

Materials and methods

Plant material and DNA extraction

All plant materials used in this study were grown in the nursery from Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences as stated in Table 1. At present, these plant materials were conserved in the China National Nursery of Sugarcane Germplasm Resources, which was built by China's Ministry of Agriculture in Kaiyuan city, Yunan province in 1995. We selected five genotypes within each original species, including *S. spontaneum*, *S. robustum*, and *S. officinarum*. In total, 15 genotypes of the genus *Saccharum* were selected. Leaf tissues of the above materials were ground in liquid nitrogen and stored at -80°C. Total genomic DNA was extracted from young leaves following CTAB methodology [31].

TA cloning and DNA sequencing

Polymerase chain reaction (PCR) amplification was carried out using a pair of universal plant primers, IGS P3 (TTGCTGCCACGATCCACTGAG) and IGS P4 (CTACTGGCAGGATCAACCAGG) [32], which were designed according to the the conserved regions of 18S rDNA and 25S rDNA, on a Veriti 96-Well Thermal Cycler. A 50 µl reaction was performed by adding 50 ng of template DNA, 25 µl of 2 × GC Buffer I (5 mM Mg²⁺ plus), 8 µl of dNTP Mixture (2.5 mM), 20 µM of each primer, and 2.5 U of LA Taq polymerase (TaKaRa LA Taq™, Takara Biotechnology Inc.). PCR amplification conditions were as follows: one cycle of denaturing at 95°C for 15 min, 35 cycles of denaturing at 95°C for 1 min, annealing at 58°C for 1 min, and extension at 72°C for 3 min, and final extension at 72°C for 10 min and storing at 4°C. The PCR products within a single band were purified using the QIAquick Gel Extraction Kit (Qiagen Inc.) and ligated to the pMD19-T-vector (Qiagen Inc.). The ligated product was transformed in *E. coli* DH 5α and plated onto the Amp/LB/X-gal/IPTG plates. The positive white colonies were further grown for plasmid isolation. Plasmid DNA was purified using a Plasmid Mini kit I (200) (OMEGA) and quantified. In each plant material of the 15 genotypes, eight positive

Table 1. The 15 genotypes of the genus *Saccharum* used in this study.

No.	Accession	Species	Ploidy level	Chromosome number
1	Yunnan82-215	<i>S. spontaneum</i>	Octaploid (8x)	2n = 64
2	Yunnan83-201	<i>S. spontaneum</i>	Nonaploid (9x)	2n = 72
3	Yunnan82-114	<i>S. spontaneum</i>	Decaploid (10x)	2n = 80
4	Fujian Huian	<i>S. spontaneum</i>	Dodecaploid (12x)	2n = 96
5	Fujian89-1-19	<i>S. spontaneum</i>	Tridecaploid (13x)	2n = 104
6	51NG3	<i>S. robustum</i>	Octaploid (8x)	2n = 80
7	57NG208	<i>S. robustum</i>	Octaploid (8x)	2n = 80
8	Daye	<i>S. robustum</i>	Octaploid (8x)	2n = 80
9	51NG63	<i>S. robustum</i>	Octaploid (8x)	2n = 80
10	NG77-004	<i>S. robustum</i>	Octaploid (8x)	2n = 80
11	Badila	<i>S. officinarum</i>	Octaploid (8x)	2n = 80
12	Nanjian Guozhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80
13	Luohanzhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80
14	Crystallina	<i>S. officinarum</i>	Octaploid (8x)	2n = 80
15	Vietnam Niuzhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80

<https://doi.org/10.1371/journal.pone.0183447.t001>

plasmids from the single bacterial transformed colony were further analyzed by PCR using the specific primers of IGS sequence (IGS P3 and IGS P4). The PCR product showed the presence of DNA fragment of the expected size. In 15 genotypes of the genus *Saccharum*, eight positive plasmids for each genotype were sequenced using the Sanger sequencing method by Beijing Genomics Institute Co. Ltd (Shenzhen, China). DNA sequencing was performed using a primer walking strategy, whereby vector-specific (M13F and M13R) and internal primers were used to generate partial sequences (S1 Table). This was repeated until sequences were obtained that overlapped in the forward and reverse directions. The IGS sequences are deposited in the Genbank database under accession numbers KX254591-KX254605.

Sequence analysis

The IGS sequence alignment was performed using the program ClustalW2 followed by manual adjustments [33, 34]. Sequence data were edited to remove the sequences at the 3' end of 18S rDNA and the 5' end of 25S rDNA. Alignment of the IGS sequences for all 15 plant materials using Geneious R7 [35]. Potential TATA box was found by the value of score computed by the TATA box weight matrix using Proscan 1.7 [36]. The multi-dot-matrix plot (MDMP) analysis was performed using the LBDot [37]. We used the Tandem Repeat Finder (TRF) program, which identified the SRs, and made manual adjustments using the BioEdit Sequence Alignment Editor [38]. Direct repeats and inverted repeats were detected and dot matrix analysis was performed using Unipro UGENE [39]. CpG islands were identified by CpGPlot [40].

The genetic distances among three different species were calculated according to Kimura 2-Parameter model using MEGA6.06 software. Phylogenetic analyses were performed by the maximum-likelihood (ML) and the Neighbor-joining (NJ) method according to the Kimura 2-Parameter model using molecular evolutionary genetics analysis program MEGA6, and all branches were evaluated with 1000 bootstrap replications [48].

FISH

The chromosomal preparations and the FISH technique were conducted as described in D'Hont et al [6]. Root tips were obtained from Yunnan82-114 (*S. spontaneum*), 51NG3 (*S. robustum*), and Luohanzhe (*S. officinarum*). The IGS labeled by nick translation with biotin-dUTP (Roche, Switzerland) was simultaneously hybridized with pTa71, and a 9-kb fragment of rDNA from wheat was also labeled by nick translation with digoxigenin on metaphase chromosomes of the above three materials [41]. High stringency conditions of post-hybridization washes were carried out with $2 \times$ SSC for 5 min at 42°C, a second wash in 50% formamide, $2 \times$ SSC, pH 7.0, for 3×5 min at 42°C, followed by a rinse in $2 \times$ SSC for 5 min at room temperature and a final wash in $0.1 \times$ SSC for 3×5 min at 55°C. Low stringency conditions were performed in $1 \times$ SSC for 30 min at 37°C. Chromosomes was counterstained with 4', 6-diamidino-2-phenylindole (DAPI) in a Vectashield anti-fade solution (Vector Laboratories, Burlingame, CA). Measurements of the IGS and pTa71 fluorescent signal intensity were performed in at least in ten cells (from two different individuals) using the AxioVision measurement module of the AxioScope A1 Imager fluorescent microscope (Zeiss, Germany).

Results

IGS variants within individual, intraspecies, and interspecies heterogeneity

To characterize IGS sequences, IGSs from the three original species, including *S. spontaneum*, *S. robustum*, and *S. officinarum*, were cloned and sequenced (Table 1). The sequencing results

of the entire IGS sequence revealed that variable sizes of the IGS were present in each genotype (S2 Table), which represented individual, intraspecific, and interspecific polymorphisms in the length of the IGS sequence. In eight sequences per genotype, the longest sequence had the conserved sequence to the other seven sequences, so the longest sequences per genotype was just deposited into the GenBank database. Comparisons of the IGS sequences from each level demonstrated that the individual level had the highest similarity and the interspecies level had the lowest similarity. At the intraspecies level of the genus *Saccharum*, comparisons of the IGS sequence of *S. robustum* and *S. officinarum* revealed a mean value of 87.86% of similarity and a mean GC content of 65.8%. The sizes of the IGS sequences ranged from 2592–2862 bp and 2590–2737 bp in *S. robustum* and *S. officinarum*, respectively. This result suggested that the IGS sequences of *S. robustum* and *S. officinarum* had few differences in length and a relatively low level of sequence heterogeneity (S2 Table). On the other hand, five *S. spontaneum* IGS clones showed some heterogeneity in length (from 2783 to 3340 bp), a mean value of 78.70% in similarity, and a mean GC content of 64.4%. This result demonstrated the high level of sequence divergence in the five *S. spontaneum*.

Due to heterogeneity in the length of IGSs in the three original species at the interspecies level of the genus *Saccharum*, the IGS sequences from Yunnan82-114 (*S. spontaneum*), 51NG3 (*S. robustum*), and Luohanzhe (*S. officinarum*) were the longest sequences in each species studied. Therefore, we chose the longest IGS sequence as a representative to outline the canonical structural organization of the IGSs from these three species of the genus *Saccharum* (S2 Table). The detailed analysis of the molecular structure of IGS sequences of these three species contains six distinct regions: NTS, ETS, SR, TTS, TIS, and a CpG island (Fig 1).

We found that all the different IGS variants from each genotype shared similar structural, regulatory and functional motifs, indicating all the different IGS variants had the necessary key elements to be functional. In addition, the differences between the different IGS variants were due to the presence or absence of sub-repeat motifs. Although it did not contradict the fact that a higher IGS diversity was present in the three species, eight sequenced clones within each genotype were enough to explore the sequence heterogeneity as well as the structure and organization of IGS region based on the two above mentioned facts.

Functional elements and domains

Beginnings and ends of transcripts. Unlike the ITS region, the IGS is a non-transcribed region of rDNA. Nevertheless, it contains several highly conserved *cis*-acting functional elements [42]. To investigate the IGS sequences for functional elements and domains, we aligned the IGSs from these three species with the transcription initiation regions from other species, including *Poaceae*, *Cucurbitaceae*, *Fabaceae*, *Brassicaceae*, *Solanaceae*, *Oleaceae*, *Fagaceae*, and *Punicaceae* [32, 43]. We found a putative TIS (TATATTAGGGG) for RNA polymerase I, which was in agreement with reports of TIS in other plants. The initiation site of RNA transcript at +1 position is the A residue at the middle position of the IGS region. Interestingly, adenine was present at the TIS in a majority of the higher plants, with the exception of *V. radiata* in the family *Fabaceae*. A putative TATA box and polyguanine sequence were generally present next to the TIS in most plants mentioned above (Fig 1). Similar to the families of *Poaceae*, *Cucurbitaceae*, *Fabaceae*, *Brassicaceae*, and *Solanaceae*, all of them shared a canonical TATA box (Fig 2). This is in agreement with most previous reports for other IGS sequences. However, *Miscanthus sinensis* (*M. sinensis*) in the family *Poaceae* lacked a typical TATA box [44]. In contrast to the presence of a TATA-containing sequence in the above plants, sequences lacking TATA adjacent to the TIS were flanked by pyrimidine bases in the families *Fagaceae* and *Punicaceae* [23, 43]. In addition, a unique AT-rich region that may function as an enhancer of

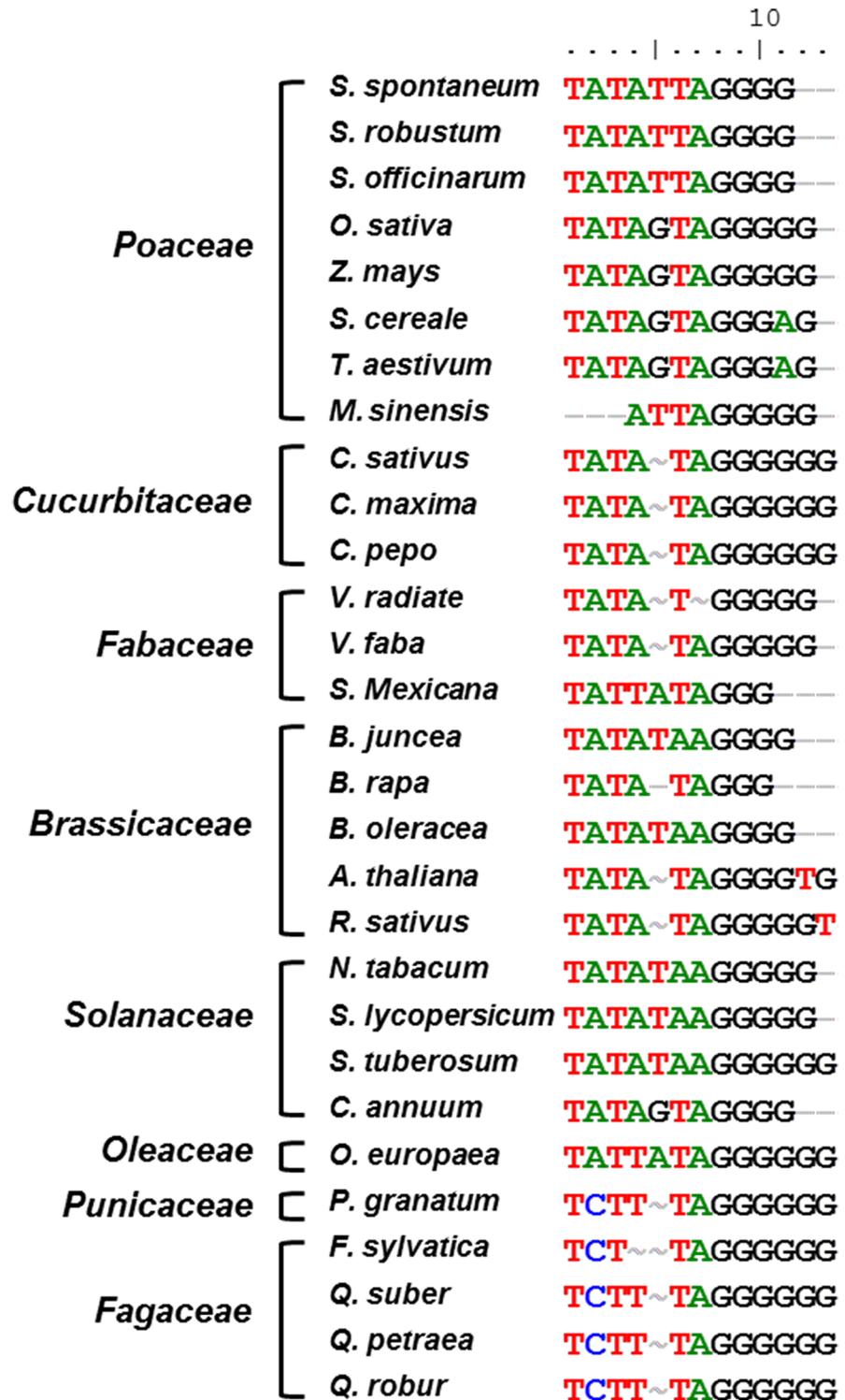


Fig 2. Comparison of putative TIS of different plants. The respective accession numbers: Yunnan82-114 (*S. spontaneum*) KX254601; 51NG3 (*S. robustum*) KX254591; Luohanzhe (*S. officinarum*) KX254598; *O. sativa* X54194; *Z. mays* X03990; *S. cereale* M37231; *T. aestivum* AJ315040; *M. sinensis* AJ238126; *C. sativus* X51542; *C. maxima* M28700; *C. pepo* X55960; *V. radiata* X17211; *V. faba* X16615; *S. mexicana* AJ489509; *B. juncea* X73032; *B. rapa* S78172; *B. oleracea* X60324; *A. thaliana* X52631; *R. sativus* Z11677; *N. tabacum* Y08422; *S. lycopersicum* AY366528; *S. tuberosum* AF464863; *C. annuum* HM352915; *O.*

europaea AJ865373; *P. granatum* JX121275; *F. sylvatica* KC700362; *Q. suber* AY428812; *Q. petraea* EU555524; *Q. robur* EF208969.

<https://doi.org/10.1371/journal.pone.0183447.g002>

transcription was present upstream of the TIS (Fig 1). Adjacent to the 5' end of the IGS, a typical pyrimidine rich motif of 14 bp (CCCTCCCCCTCC) was identified for RNA polymerase I as a putative TTS. This motif has been suggested to play a role in the finalization of rRNA transcription as a TTS.

NTS and ETS. The region between the 3' end of the 25S rRNA gene and the putative TIS represents the NTS, whereas the ETS was found to be located between the putative TIS and the beginning of the 18S rRNA gene (Fig 1). Due to lower functional constraints, the sequences of the NTS showed high divergence. In contrast, the ETS region has better conserved nucleotide sequences than the NTS despite having different lengths among the three species. The length differences of NTS and ETS regions among the three species were mainly due to insertions or deletions.

CpG islands and methylation sites in the IGS sequence

Among these three species, two CpG islands were detected in NTS and ETS, respectively. However, the length of the first CpG island was longer than that of the second one. For example, the first CpG island was 667–1112 bp in length downstream from the TTS, while the second one was 534–735 bp in length downstream from the TIS, which was located just after the putative TIS (Fig 1 and S3 Table). Most of the CpG islands lie further away from the TIS [45]. It is worth noting that the IGS sequence includes a large number of GCGC and CCGG motifs that were irregularly distributed along the entire sequence (Table 2). The presence of a large number of the methylatable motifs has also been reported in other studies [46–48].

Sub-repeat (SR) regions

To explore the nature of SR regions in the IGS in greater detail, we performed a bioinformatics analysis of the repeated DNA of IGS sequences using TRF and MDMP analysis. The TRF results showed a total of eight types of SRs in the IGS and different sizes in each SR (Fig 1; S4 and S5 Tables). Based on the sequence length, these SRs were classified into two categories: short SR and long SR. Five relatively short SRs (SR1, SR3, SR4, SR5, and SR6) were 24–32 bp in length, while the other three relatively long ones (SR2, SR7, and SR8) were 110–172 bp in length (S4 Table). Of these, the SR8 was the longest at 172 bp, whereas SR5 was the shortest with only 22 bp. Six SR blocks (SR1–SR6) occurred upstream of the TIS and two SR blocks (SR7 and SR8) were located downstream of the TIS (Fig 1). However, no SR element in *M. sinensis*, *Lycopersicon esculentum*, and *Zea mays* was present in the region downstream of the TIS, but the larger number of SRs was compensated upstream of the TIS [44, 49, 50]. At the intra- and the inter-species levels of the genus *Saccharum*, comparison results among the differently sized IGS variants revealed that the copy number of each types of SRs was similar, except for SR8, which contained a relatively higher copy number in *S. spontaneum* (Fig 1; S5 Table). These variants clearly occurred as a result of copy number variations in SR8. Among the three species, all SRs were highly conserved, and the length and sequence identity differed slightly among different SRs. Four SRs, including SR1 and SR2 at the beginning of the IGS and SR7 and SR8 at the end of the IGS, shared high similarity of at least 92.21%. With the exception of SR6, which shared comparatively low sequence similarity of 75%, the other three SRs (SR3, SR4, and SR5) had a mean similarity value of 88.71% (S4 Table). SR5 had the lowest GC content of 47.6%, but the other seven SRs exhibited high GC content at more than 60%, with SR7

Table 2. Methylation sites in the IGS sequence.

No.	Genotype	Species	Ploidy level	Chromosome number	No. of methylation sites	
					GCGC	CCGG
1	Yunnan82-215	<i>S. spontaneum</i>	Octaploid (8x)	2n = 64	21	26
2	Yunnan83-201	<i>S. spontaneum</i>	Nonaploid (9x)	2n = 72	19	26
3	Yunnan82-114	<i>S. spontaneum</i>	Decaploid (10x)	2n = 80	16	24
4	Fujian Huian	<i>S. spontaneum</i>	Dodecaploid (12x)	2n = 96	17	26
5	Fujian89-1-19	<i>S. spontaneum</i>	Tridecaploid (13x)	2n = 104	22	29
6	51NG3	<i>S. robustum</i>	Octaploid (8x)	2n = 80	25	33
7	57NG208	<i>S. robustum</i>	Octaploid (8x)	2n = 80	26	30
8	Daye	<i>S. robustum</i>	Octaploid (8x)	2n = 80	23	25
9	51NG63	<i>S. robustum</i>	Octaploid (8x)	2n = 80	23	26
10	NG77-004	<i>S. robustum</i>	Octaploid (8x)	2n = 80	26	30
11	Badila	<i>S. officinarum</i>	Octaploid (8x)	2n = 80	23	31
12	Nanjian Guozhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80	22	26
13	Luohanzhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80	22	27
14	Crystallina	<i>S. officinarum</i>	Octaploid (8x)	2n = 80	23	32
15	Vietnam Niuzhe	<i>S. officinarum</i>	Octaploid (8x)	2n = 80	23	31

<https://doi.org/10.1371/journal.pone.0183447.t002>

having a GC content of 75% (S4 Table). Interestingly, SR7 and SR8 shared a partial overlapping sequence motif (Fig 1). A MDMP analysis was applied by self-comparison of the IGS sequences using LBDot. Blue dots represented the homolog regions of the IGS sequences both in the direct orientation, and red dots compared the IGS sequences in the direct orientation with in the reverse orientation. That is, blue dots represented the direct repeats, and red dots were representative for the inverted repeats. The distribution pattern of the direct and inverted repeats was shown by the concentrated distribution areas in the IGS sequences. The more the distribution density was, the more the copy number of the direct and inverted repeats was. Self-comparison analysis of three IGSs using MDMP further revealed that each spacer was composed of two intensively repeated regions, as shown in the square boxes in Fig 3. A significant number of direct repeats were concentrated in both regions, and the inverted repeats were scattered without a distinct distribution pattern in the IGS sequence. It is noteworthy that in the light of the distribution pattern, Luohanzhe and 51NG3 shared a similar distribution pattern of the direct and inverted repeats, whereas Yunnan82-114 was different. In addition, we found the presence of several base substitutions and partially deleted copies of each SR both at the individual level (data not shown) and at the inter-species level (S1 File).

Phylogenetic relationship of the three original species based on analysis of IGS sequence

The phylogenetic trees were constructed using ML and NJ based on the IGS sequences. The cluster analysis revealed that the NJ tree was similar to the ML tree with seven major clusters for the three original species in the genus *Saccharum* and other species from the family *Poaceae*, including *Miscanthus*, *Zea*, *Oryza*, *Hordeum*, *Aegilops*, and *Triticum* (Fig 4). In the genus *Saccharum*, the five *S. spontaneum* fell into a single cluster (clade II), and all of them shared the same array of the SR8 element in terms of number and position. However, ten accessions of *S. robustum* and *S. officinarum* formed two discrete groups (clade I), and all of them were completely intermingled without any hint of distinct lineage between *S. robustum* and *S. officinarum*. This was similar to the species from the genus *Triticum* and *Aegilops* (clade VII).

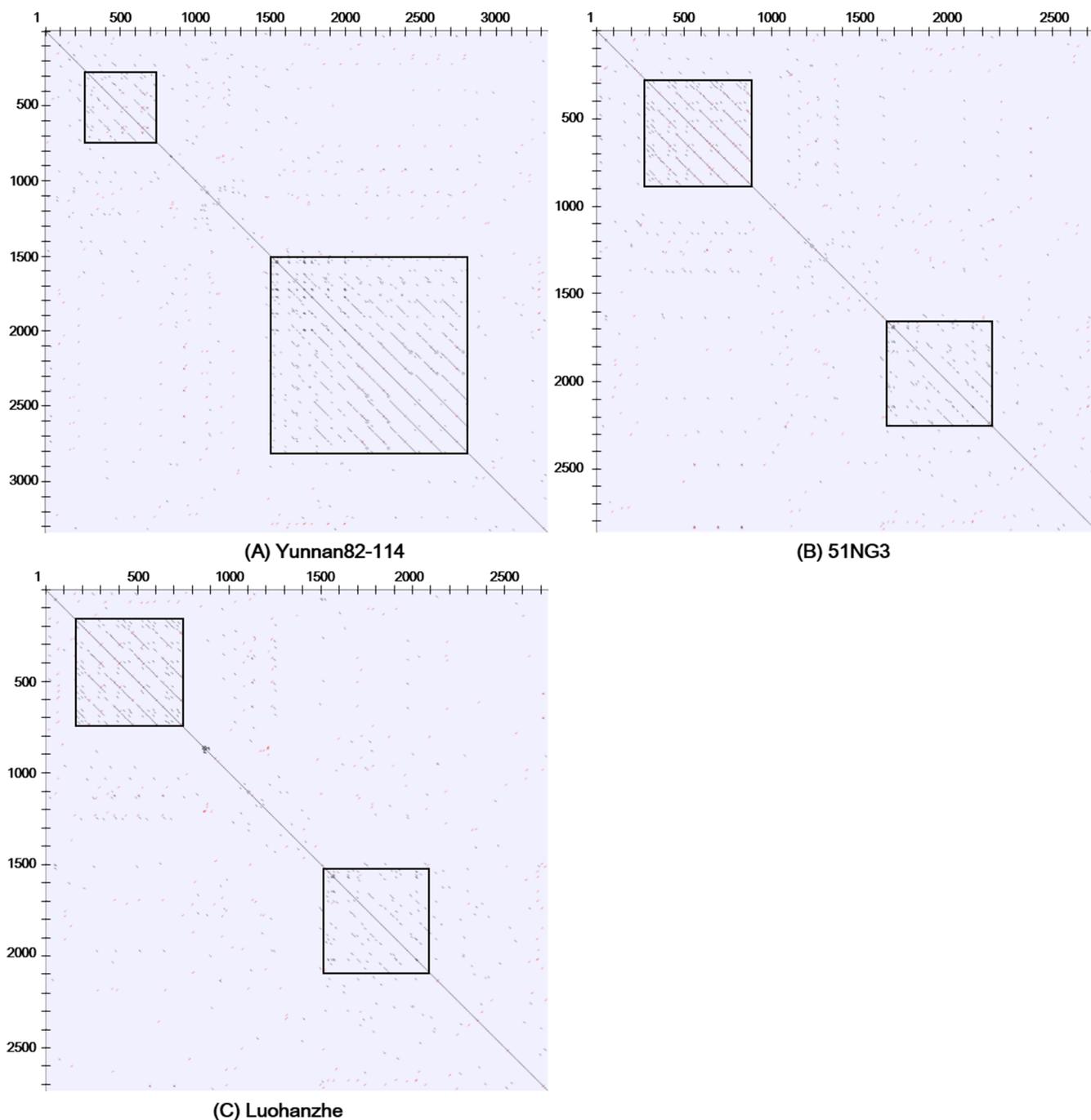


Fig 3. Dot matrix plot of IGS spacers in (A) Yunnan82-114, (B) 51NG3 and (C) Luohanzhe. IGS spacer was performed using a k-tuple of 8 with 100% identity.

<https://doi.org/10.1371/journal.pone.0183447.g003>

Physical mapping of IGS in *S. spontaneum*, *S. robustum*, and *S. officinarum*

To determine the chromosomal location between 45S rDNA and IGS within the three original species, including *S. spontaneum*, *S. robustum*, and *S. officinarum*, we selected one genotype in

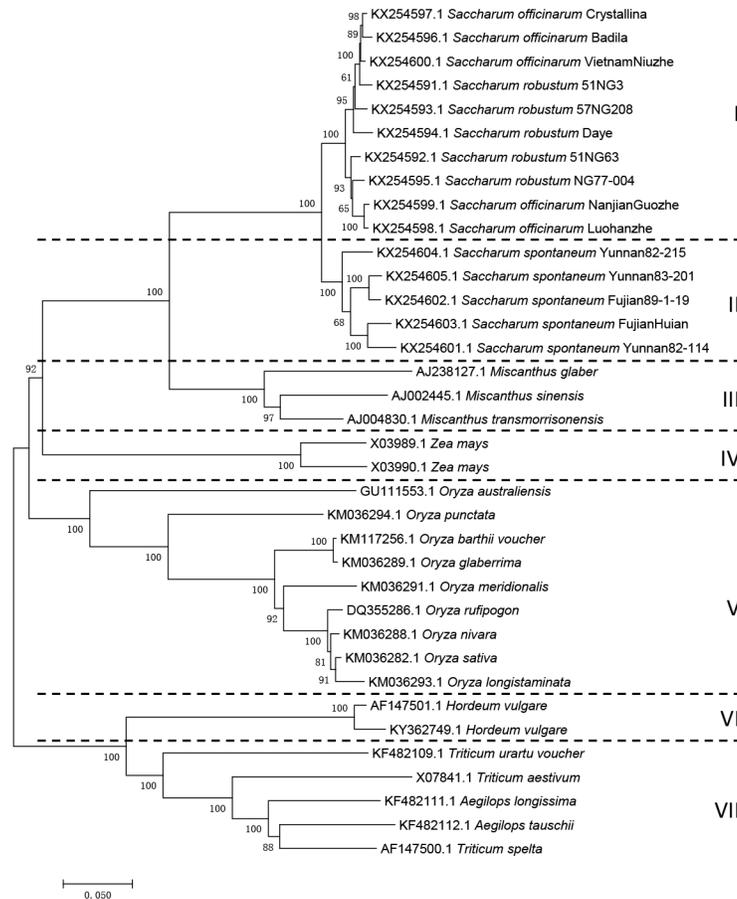


Fig 4. Phylogenetic relationships among 15 accessions of the genus *Saccharum*. Bootstrap values from 1000 tests are indicated at the nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The ML and NJ tree of the same data is topologically identical.

<https://doi.org/10.1371/journal.pone.0183447.g004>

each original species as representatives, including Yunnan82-114 (*S. spontaneum*), 51NG3 (*S. robustum*), and Luohanzhe (*S. officinarum*). FISH analysis was applied simultaneously using the pTa71 probe for the 45S rDNA locus and IGS probe for the IGSs from each species on the metaphase chromosome. As expected, the IGSs from each species co-localized with the 45S rDNA, and the number of IGS signals was in accordance with the number of the 45S rDNA signals in each species. However, the number of loci detected varied, as follows: five loci in Yunnan82-114, seven loci in Luohanzhe, and eight loci in 51NG3 (Fig 5). In addition, the location of both 45S rDNA and the IGS in *S. spontaneum* can differ from that in *S. robustum* and *S. officinarum*. In agreement with the report of D'Hont et al. [6], the interstitial signal was constantly observed on the 45S rDNA-bearing chromosome in *S. spontaneum*, whereas the terminal signal was in *S. robustum* and *S. officinarum* (Fig 5). Remarkably, some minor loci NORs were found among Luohanzhe, 51NG3, and Yunnan82-114, but the number of minor loci NORs varied slightly. Of these, 51NG3 and Yunnan82-114 bore three minor loci NORs, while Luohanzhe carried only one minor locus NOR.

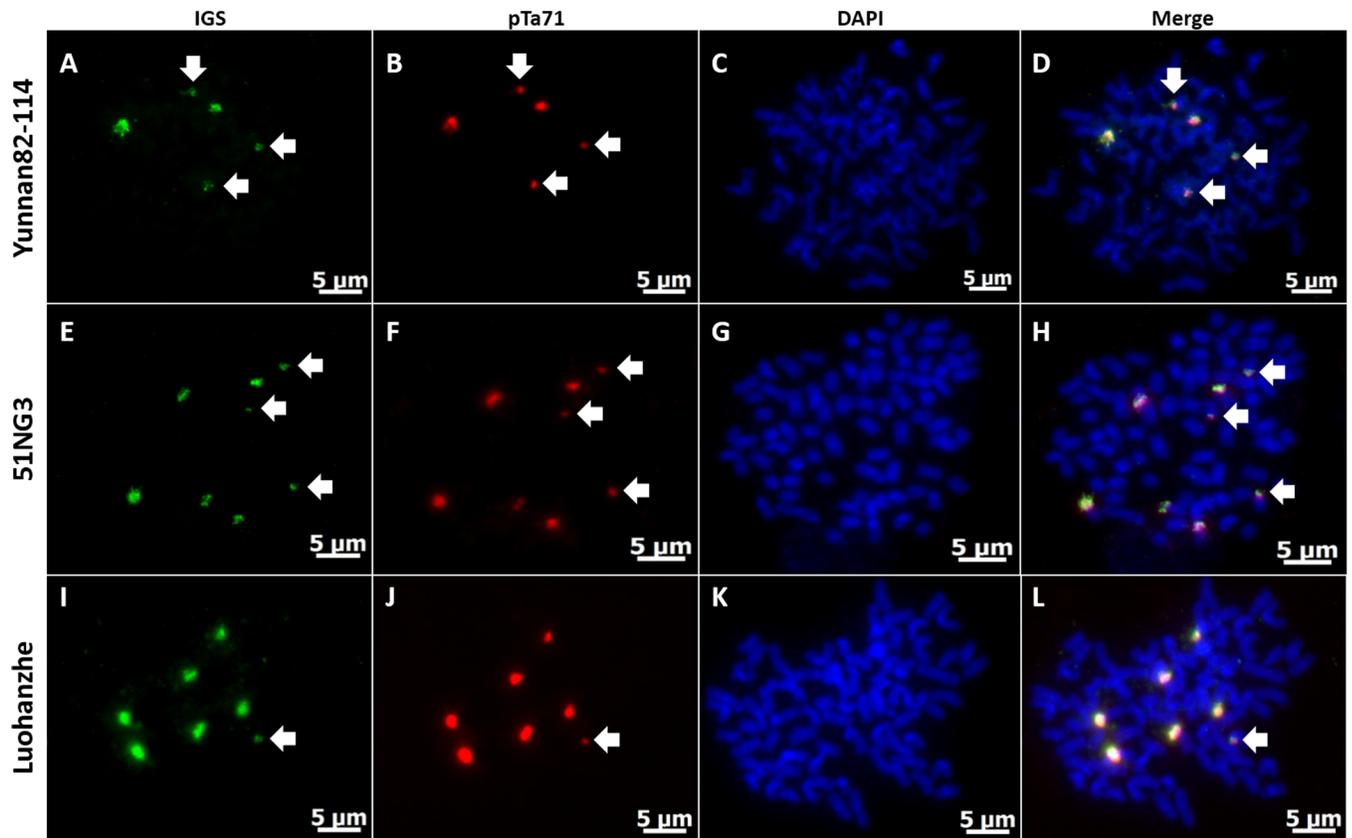


Fig 5. Physical mapping of IGS and pTa71 probe in Yunnan82-114 (*S. spontaneum*), 51NG3 (*S. robustum*) and Luohanzhe (*S. officinarum*). FISH with IGS (red—B, F, J), and wheat rDNA pTa71 probe (green—A, E, I) in meristematic root tip metaphase chromosomes of Yunnan82-114 (A–D), 51NG3 (E–H), and Luohanzhe (I–L). DNA is counterstained with DAPI (blue—C, G, K). The fourth column shows the merged images of both signals of IGS and wheat rDNA pTa71 probe and DAPI-stained chromosomes (D, H, L). Arrows indicate minor loci NORs (A, B, D; E, F, H; I, J, L). Scale bars = 5 μ m.

<https://doi.org/10.1371/journal.pone.0183447.g005>

Discussion

Length heterogeneity at the individual, intraspecies, and interspecies levels of the genus *Saccharum*

In this study, we observed variable length of the IGS sequence at the individual, intraspecies, and interspecies levels of the genus *Saccharum*, including *S. spontaneum*, *S. robustum*, and *S. officinarum*. The IGS is a rapidly evolving region of the 45S rDNA with several internal SRs. The length of polymorphisms has been observed among populations, species, and even in individuals in many plant species [24]. For instance, Jo et al. found that the length of the IGS sequence differed greatly at the inter-species and inter-generic level of the family *Solanaceae* [51]. In contrast, we detected a narrow range of length variability at the individual, intraspecies, and interspecies levels of the genus *Saccharum*. Notably, our sequence analysis showed that the IGS variants occur mainly due to duplications or deletions and base substitutions at these three levels. Except for the duplications or deletions, the degree of sequence similarity was considerably high. The common characteristic within the rDNA IGS is the presence of SR elements for both low and high eukaryotes. In fact, heterogeneity of IGS length is attributed to duplications or deletions of the SR region that can be present in different copy numbers in almost all species [51–55]. Similarly, in this study, the variation in IGS length was due to

variation in the copy number of one or several SR motifs generated by duplications or deletions. In other words, indels were restricted to SR-rich regions, suggesting that these elements are involved in the differences in length observed. The SR regions may be the hot-spot regions of the IGS during their evolutionary history. The polymorphism of IGS length is also believed to be the result of concerted evolution [15]. In plants, heterogeneity in the length of the IGS has been thought to be due to variations in the SR elements generated by unequal crossing over [19]. In this study, from the analysis between the different IGS variants in the intraindividual level, we found evidence that duplication and deletion events in the IGS region within the genus *Saccharum* gave rise to the differences in length as well as different copy numbers of SRs. The polymorphism of the IGS sequence may have arisen following seven indel events, except for the SR7 with the same copy number (3.7). As one of the major driving forces in the evolution of the rDNA units, this variability can be attributed to unequal crossing over between SR elements.

Structure of promoter and terminator sequences and ETS region

Based on sequence analyses, we found several functional elements within the IGS sequences. The functional elements with high nucleotide conservation are hypothesized to be associated with the crucial process of initiating and terminating rRNA transcription by RNA polymerase I, such as a promoter, enhancer elements, and the TATA box. Based on comparisons with the promoter regions from other species, we only detected a unique putative promoter in the IGS variants of the genus *Saccharum*, although several animal and plant species have more than one promoter and the functional significance has been related to their transcriptional activity. Similar to the IGS characterized for most vascular plant species, the promoter region appears to be located between the repetitive areas (between SR6 and SR7) in the genus *Saccharum*. The putative TIS sequence (TATATTAGGGG) in the genus *Saccharum* contained the core sequence around the highly conserved A residue. The putative TIS sequence showed resemblance to the hypothesized consensus promoter sequence reported for the plant promoters. In most families of *Angiosperms*, the TATA and GGGG boxes in the core promoter seem to be a pervasive feature of rRNA transcription [24]. This also suggests that, rather than the sequence itself, a structural code is the determinant for its function, probably by directing specific DNA-protein interactions involved in transcriptional control. Similar to the GGGG boxes in most families of *Angiosperms*, the hexanucleotide sequence GCGGG was previously found at similar regions in several eukaryotic promoters and was shown to be crucial for promoter activity [56]. In *P. granatum*, *M. sinensis*, *F. sylvatica*, *Q. suber*, *Q. petraea*, and *Q. robur*, the TATA is missing in the upstream of the conserved A residue of the putative TIS sequence [23, 43, 44]. Precise transcription factor interactions, dictated by the correct spacing of promoter elements, are key to promoter recognition, while TATA sequence may not be a direct binding site for transcription factors, but is conserved probably only for DNA to be easily melted [57]. In addition, it has been noted that in several plant species, the promoter sequences in IGS regions are usually duplicated or contain multiple repeats that enhance the polymerase activity via an increase in transcription initiation.

A unique AT-rich region upstream of the TIS in the genus *Saccharum* may represent potential binding protein regions involved in the TIS. The pyrimidine-rich motif sequence (CCCTCCCCCTCC) located just after the 3' terminus of the 25S rRNA gene likely acts as a proximal terminator. The putative terminator was highly similar to a portion of the conserved motif present at the beginning of the IGS of some plant species. In general, the AT-rich sequences in several plants are preceded by a region composed of the putative TIS. This region could form the domain of the promoter sequence known as the upstream binding element,

which confers species specificity to polymerase I [24, 43]. The presence of a short AT-rich region in the genus *Saccharum* has been shown to be associated with enhanced transcription.

In this study, the sequences of the NTS showed high divergence among *S. spontaneum*, *S. robustum*, and *S. officinarum*. In contrast to the lower level of similarity of the NTS region, the ETS region has better-conserved nucleotide sequences than the NTS region. Similar results were also observed in other grasses, such as *M. sinensis*, *Oryza sativa*, *Zea mays*, and *Triticum aestivum* [44, 49, 58, 59]. In addition, sequence alignment showed that the difference in ETS length among *S. spontaneum*, *S. robustum*, and *S. officinarum* was due to insertions of SR8, suggesting that SR8 might be an evolutionary hot-spot region in the IGS sequence. The result seemed to show that the length of IGS tended to reduce gradually from the most primitive species *S. spontaneum* to the domesticated species *S. officinarum* with increasing evolutionary relationship distance.

CpG islands and methylation sites

The average GC content of the entire IGS region was over 60% in the genus *Saccharum*. The regions with the greatest GC content shared the SR and ETS regions in *Poaceae*. Indeed, there were two CpG islands in the entire IGS. A relatively small CpG island was located between the TIS and 3' end of the 18S rRNA gene. As a rule, CpG islands have a GC content that is significantly higher than that of the genome average and are nonmethylated [60]. Cytosine is a DNA base that can be chemically modified, and cytosine methylation of CpG base pairs in plants is associated with the inhibition of rRNA gene transcription initiation [61]. Similar to *Avena*, *Olea*, and *Triticum* [46–48], the IGS of the genus *Saccharum* has a sufficient number of methylation sites that are irregularly distributed along the sequence. Indeed, methylation levels have been found to be associated with the transcription of rRNA genes [22, 62].

Phylogenetic relationship of the three original species within the genus *Saccharum*

The spacers (45S rDNA ITS, IGS, and 5S rDNA NTS) with the higher variation in length and sequence allow for us to resolve the phylogenetic relationships of species. Phylogenetic relationships of inter-species in the genus *Saccharum* have been previously investigated with regard to sequence spacing. Liu et al. reported that *S. spontaneum*, *S. robustum*, and *S. officinarum* were clustered together based on 45S rDNA-ITS haplotype data in *S. spontaneum* [3]. Pan et al. found that the two *S. officinarum* accessions did not form distinct clusters, but rather were interrelated within the *S. spontaneum* cluster based on the 5S rDNA NTS sequences [4].

Phylogeny based on 45S rDNA ITS and 5S rDNA NTS sequences is clearly not enough to reflect the relationship of species within the genus *Saccharum*. The most plausible explanation is the lack of sufficient abundant variances in the highly conserved length of 45S rDNA ITS and 5S rDNA NTS sequences in the species of the genus *Saccharum*. The IGS region is evolving rapidly compared to the 45S rDNA ITS and 5S rDNA NTS sequences. In most plants, due to the differences in the SR sequences (including copy number, sequence, and length) of the IGS region among different species, the IGS region may provide a more suitable marker for reconstructing phylogenetic relationships within both inter- and intra-species than the 45S rDNA ITS and 5S rDNA NTS.

In this study, *S. officinarum* shared the same clade with *S. robustum* in the NJ tree, supporting the close relationship between *S. officinarum* and *S. robustum*. On the contrary, *S. spontaneum* formed a highly supported separate clade, indicating its phylogenetic isolation from *S. officinarum* and *S. robustum*. The phylogenetic relationship of the IGS sequences closely

reflected the relationship of the species in the genus *Saccharum*. Many molecular marker studies have suggested that *S. robustum* is the progenitor of *S. officinarum*, while *S. spontaneum* is the most primitive species [6].

FISH mapping IGS sequences are only located at the NOR

In all of the species of the genus *Saccharum* that were analyzed, the IGS always co-localized in the same loci with 45S rDNA, suggesting that IGS sequences were only located at the NOR. However, Jo et al. reported that in addition to the co-localization of 45S rDNA and IGS at the same loci, an independent IGS-derived satellite DNA was located outside the 45S rDNA loci in the *Capsicum* species [51]. Additionally, Chester et al. reported that the IGS sequences containing a truncated retrotransposon sequence similar to members of the Copia superfamily was dispersed along both pairs of 45S rDNA loci [63]. Theoretically, the number of rDNA loci should be equal to the basic chromosome numbers in autopolyploid of the family *Poaceae*. For instance, the diploid species, such as *Zea mays* and *Sorghum bicolor*, have two 5S rDNA loci as well as two 45S rDNA loci [64, 65]. In this study, the decaploid species Yunnan82-114 (*S. spontaneum*) contained five loci, whereas the two octaploid species Luohanzhe (*S. officinarum*) and 51NG3 (*S. robustum*) bore seven and eight loci, respectively. Except for 51NG3 (*S. robustum*), the difference between the expected and observed number of 45S rDNA loci was present in Yunnan82-114 (*S. spontaneum*) and Luohanzhe (*S. officinarum*). This result suggests that the 45S rDNA loci in the genus *Saccharum* seems to experience a dynamic evolution and some 45S rDNA loci appear to have been eliminated. Indeed, chromosomal polymorphisms in the number and chromosomal position of the 45S rDNA loci were also found in other grasses, such as *Oryza* and *Paspalum* [66, 67]. In contrast to the 45S rDNA loci, the 5S rDNA loci are usually not syntenic to the 45S rDNA and are generally distributed in an interstitial position in the chromosomes among most species, suggesting that the location of the 5S rDNA sequence appears to be conserved. Such a conserved interstitial position within the 5S rDNA loci does not appear to result in the elimination of 5S rDNA. Thus, compared with 45S rDNA, 5S rDNA is more suitable for determining the basic chromosome numbers of species. For the three original species of the genus *Saccharum*, D'Hont et al. also provided evidence that the number of 5S rDNA loci corresponded well with the ploidy level, but the number of 45S rDNA sites was variable [6]. Notably, the distribution of co-localized loci was different among *S. spontaneum*, *S. robustum*, and *S. officinarum*. The co-localized loci were located in terminal regions of chromosomes in *S. officinarum* and *S. robustum*, whereas they were located more interstitially in *S. spontaneum*. A similar pattern was observed by D'Hont et al., who found that 45S rDNA sites were terminal for *S. officinarum* and *S. robustum* and interstitial for *S. spontaneum* [6]. Additionally, a BAC of 45S rDNA units (SCHRBa 039 D18) was obtained from the sugarcane hybrid R570, an amphidiploid derived from cross between *S. officinarum* and *S. spontaneum* [68]. Analyses of the sequence from the BAC of 45S rDNA units showed that there were 14 45S rDNA units with a portion of one unit in the reverse orientation to the other 13 and the sequence length of IGS is 3011 bp. The result of sequence alignment with our IGS sequences from the three species revealed that the IGS sequences from the BAC of 45S rDNA units are high similar to the IGS sequences from our genotypes of *S. spontaneum*. Thus, the BAC of 45S rDNA units was derived from *S. spontaneum*. This BAC was located to metaphases from root tips of the sugarcane cultivar R570. FISH result revealed that there were the less intense interstitial rDNA signals and the more intense terminal signals, implying 45S rDNA-bearing chromosomes are derived from *S. officinarum* and *S. spontaneum* chromosomes in this sugarcane modern cultivar R570.

Conclusion

In the present study, we performed for the first time a comparative analysis of the sequence organization of the IGS, and the chromosome locations of the IGS within these three original species. We found length variability of the IGS region in the three original species and a similar overall organization in the IGS as follows: (1) the presence of eight SR blocks (upstream and downstream from TIS); (2) a unique AT-rich region upstream of the TIS; (3) two CpG islands in NTS and ETS; (4) a highly homologous terminator and promoter regions with the same TTS and TIS sequences and other regulatory elements; and (5) an ETS with high similarity among species and an NTS with relatively higher variability. Moreover, the location of the IGS regions exclusively at the NORs was determined for different original species of the genus *Saccharum*. The results of this study provide important insights into the structure and organization of the IGS in the genus *Saccharum*.

Supporting information

S1 File. Sequence alignment of the sub-repeats of 15 genotypes at the inter-species level of the genus *Saccharum*.

(PDF)

S1 Table. The internal primer used for sequencing of IGS sequences.

(DOCX)

S2 Table. IGS characterization in 15 genotypes of the genus *Saccharum*.

(DOCX)

S3 Table. The CpG islands location in the IGS region.

(DOCX)

S4 Table. Sub-repeat in the genus *Saccharum*.

(DOCX)

S5 Table. The copy number of SR in the genus *Saccharum*.

(DOCX)

Acknowledgments

We thank Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences for providing the plant materials used in this study.

Author Contributions

Conceptualization: Yongji Huang, Zuhu Deng.

Formal analysis: Yongji Huang, Fan Yu.

Funding acquisition: Yongji Huang, Jiayun Wu, Yongqing Yang, Zuhu Deng, Muqing Zhang.

Investigation: Yongji Huang.

Methodology: Yongji Huang, Fan Yu, Xueting Li, Ling Luo, Zuhu Deng.

Project administration: Yongji Huang.

Resources: Zuhu Deng, Rukai Chen, Muqing Zhang.

Supervision: Zuhu Deng, Muqing Zhang.

Visualization: Yongji Huang, Fan Yu.

Writing – original draft: Yongji Huang.

Writing – review & editing: Zuhu Deng, Muqing Zhang.

References

1. Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS. Domestication to crop improvement: genetic resources for Sorghum and Saccharum (Andropogoneae). *Ann Bot.* 2007; 100(5): 975–989. <https://doi.org/10.1093/aob/mcm192> PMID: 17766842;
2. Bremer G. Problems in breeding and cytology of sugar cane. *Euphytica.* 1961; 11(12): 178–188. <https://doi.org/10.1007/BF00037206>
3. Liu X, Li X, Liu H, Xu C, Lin X, Li C, et al. Phylogenetic Analysis of Different Ploidy Saccharum spontaneum Based on rDNA-ITS Sequences. *PLoS One.* 2016; 11(3): e0151524. Epub 2016/03/18. <https://doi.org/10.1371/journal.pone.0151524> PMID: 26986847;
4. Pan YB, Burner DM, Legendre BL. An assessment of the phylogenetic relationship among sugarcane and related taxa based on the nucleotide sequence of 5S rRNA intergenic spacers. *Genetica.* 2000; 108(3): 285–295. <https://doi.org/10.1023/A:1004191625603> PMID: 11294615
5. Takahashi S, Furukawa T, Asano T, Terajima Y, Shimada H, Sugimoto A, et al. Very close relationship of the chloroplast genomes among *Saccharum* species. *Theor Appl Genet.* 2005; 110(8): 1523–1529. Epub 2005/04/09. <https://doi.org/10.1007/s00122-005-1990-z> PMID: 15818464.
6. D'Hont A. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome.* 1998; 41(2): 221–225. <https://doi.org/10.1139/g98-023>
7. D'Hont A, Paulet F, Glaszmann JC. Oligoclonal interspecific origin of 'North Indian' and 'Chinese' sugar-canes. *Chromosome Res.* 2002; 10(3): 253–262. <https://doi.org/10.1023/A:1015204424287> PMID: 12067214
8. de Souza Marislane Carvalho Paz S JN, Almeida Cícero. Differential detection of transposable elements between *Saccharum* species. *Genet Mol Biol.* 2013; 36(3): 408–412. <https://doi.org/10.1590/S1415-47572013005000030> PMID: 24130449
9. Guimarães CT, Sobral BWS. The *Saccharum* Complex: Relation to Other Andropogoneae. NCGR, Sante Fe, New Mexico 2010. 269–288 p.
10. Grivet L, Glaszmann JC, D'Hont A. Molecular evidence of sugarcane evolution and domestication. *Cirad.* 2011: 49–66.
11. Irvine JE. *Saccharum* species as horticultural classes. *Theor Appl Genet.* 1999; 98(2): 186–194. <https://doi.org/10.1007/s001220051057>
12. Ming R, Moore PH, Wu KK, Angélique DH, Glaszmann JC, Tew TL, et al. Sugarcane Improvement through Breeding and Biotechnology. *Plant breeding reviews.* 2005; 27: 15–118. <https://doi.org/10.1002/9780470650349.ch2>
13. Grivet L, Daniels C, Glaszmann JC, D'Hont A. A Review of Recent Molecular Genetics Evidence for Sugarcane Evolution and Domestication. *Ethnobotany Res Appl.* 2004; 2(1): 9–17. <https://doi.org/10.17348/era.2.0.9-17>
14. Amalraj VA, Balasundaram N. On the Taxonomy of the Members of 'Saccharum Complex'. *Genet Resour Crop Ev.* 2006; 53(1): 35–41. <https://doi.org/10.1007/s10722-004-0581-1>
15. Stage DE, Eickbush TH. Sequence variation within the rRNA gene loci of 12 *Drosophila* species. *Genome Res.* 2007; 17(12): 1888–1897. <https://doi.org/10.1101/gr.6376807> PMID: 17989256
16. Kovarik A, Matyasek R, Lim KY, Skalická K, Koukalová B, Knapp S, et al. Concerted evolution of 18–5.8-26S rDNA repeats in *Nicotiana* allotetraploids. *Biol J Linn Soc.* 2004; 82(4): 615–625. <https://doi.org/10.1111/j.1095-8312.2004.00345.x>
17. Ganley AR, Kobayashi T. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 2007; 17(2): 184–191. <https://doi.org/10.1101/gr.5457707> PMID: 17200233;
18. Galian JA, Rosato M, Rossello JA. Incomplete sequence homogenization in 45S rDNA multigene families: intermixed IGS heterogeneity within the single NOR locus of the polyploid species *Medicago arborea* (Fabaceae). *Ann Bot.* 2014; 114(2): 243–251. Epub 2014/06/14. <https://doi.org/10.1093/aob/mcu115> PMID: 24925322;
19. Ganley AR, Scott B. Extraordinary ribosomal spacer length heterogeneity in a *Neotyphodium* endophyte hybrid: implications for concerted evolution. *Genetics.* 1999; 150(4): 1625–1637.

20. Krieg AM. CpG motifs in bacterial DNA and their immune effects. *Annu Rev Immunol.* 2002; 20(1): 709–760. <https://doi.org/10.1146/annurev.immunol.20.100301.064842> PMID: 11861616
21. Fulneček J, Matyášek R, Kovařík A, Bezděk M. Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. *Mol Gen Genet.* 1998; 259(2): 133–141. <https://doi.org/10.1007/s004380050798> PMID: 9747704
22. Komarova NY, Grabe T, Huigen DJ, Hemleben V, Volkov RA. Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. *Plant Mol Biol.* 2004; 56(56): 439–463. <https://doi.org/10.1007/s11103-004-4678-x> PMID: 15604755
23. Parvaresh M, Talebi M. Molecular characterization of ribosomal DNA intergenic spacer (IGS) region in pomegranate (*Punica granatum* L.). *Plant Syst Evol.* 2014; 300(5): 899–908. <https://doi.org/10.1007/s00606-013-0928-1>
24. Poczai P, Hyvonen J. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol Biol Rep.* 2010; 37(4): 1897–1912. Epub 2009/07/25. <https://doi.org/10.1007/s11033-009-9630-3> PMID: 19626457.
25. Torres RA, Zentgraf U, Hemleben V. Species and genus specificity of the intergenic spacer (IGS) in the ribosomal RNA genes of Cucurbitaceae. *Z Naturforsch C.* 1989; 44(11–12): 1029–1034. Epub 1989/11/01. PMID: 2631730.
26. Kato A, Nakajima T, Yamashita J, Yakura K, Tanifuji S. The structure of the large spacer region of the rDNA in *Vicia faba* and *Pisum sativum*. *Plant Mol Biol.* 1990; 14(6): 983–993. Epub 1990/06/01. PMID: 2102880.
27. Maggini F, Tucci G, Demartis A, Gelati MT, Avanzi S. Ribosomal RNA genes of *Phaseolus coccineus*. I. *Plant Mol Biol.* 1992; 18(6): 1073–1082. Epub 1992/04/01. PMID: 1600143.
28. Bennett RI, Smith AG. The complete nucleotide sequence of the intergenic spacer region of an rDNA operon from *Brassica oleracea* and its comparison with other crucifers. *Plant Mol Biol.* 1991; 16(6): 1095–1098. Epub 1991/06/01. PMID: 1863764.
29. Lakshmikumaran M, Negi MS. Structural analysis of two length variants of the rDNA intergenic spacer from *Eruca sativa*. *Plant Mol Biol.* 1994; 24(6): 915–927. Epub 1994/03/01. PMID: 8204828.
30. Carvalho A, Guedes-Pinto H, Lima-Brito J. Intergenic spacer length variants in Old Portuguese bread wheat cultivars. *J Genet.* 2011; 90(2): 203–208. Epub 2011/08/27. PMID: 21869468.
31. Jr SC, Via LE. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques.* 1993; 14(5): 748–750. PMID: 8512694
32. Chang KD, Fang SA, Chang FC, Chung MC. Chromosomal conservation and sequence diversity of ribosomal RNA genes of two distant *Oryza* species. *Genomics.* 2010; 96(3): 181–190. <https://doi.org/10.1016/j.ygeno.2010.05.005> PMID: 20580815
33. Goujon M, McWilliam H, Li WZ, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 2010; 38(Web Server issue): 695–699. <https://doi.org/10.1093/nar/gkq313> PMID: 20439314
34. MA L, G B, NP B, R C, PA M, H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23(21): 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036
35. Olsen C, Qaadri K, Moir R, Kearse M, Buxton S, Cheung M, et al., editors. *Geneious R7: A Bioinformatics Platform for Biologists.* International Plant and Animal Genome Conference Xxii; 2014.
36. Dan SP. Predicting Pol II Promoter Sequences using Transcription Factor Binding Sites. *Journal of Molecular Biology.* 1995; 249(5): 923–932. <https://doi.org/10.1006/jmbi.1995.0349> PMID: 7791218
37. Huang Y, Zhang L. Rapid and sensitive dot-matrix methods for genome analysis. *Bioinformatics.* 2004; 20(4): 460–466. <https://doi.org/10.1093/bioinformatics/btg429> PMID: 14764561.
38. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999; 27(2): 573–580. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982;
39. Okonechnikov K, Golosova O, Fursov M, team U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012; 28(8): 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091> PMID: 22368248.
40. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000; 16(6): 276–277. [http://dx.doi.org/10.1016/S0168-9525\(00\)02024-2](http://dx.doi.org/10.1016/S0168-9525(00)02024-2). PMID: 10827456.
41. Gerlach WL, Bedbrook JR. Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* 1979; 7(7): 1869–1885. <https://doi.org/10.1093/nar/7.7.1869> PMID: 537913;
42. James SA, O’Kelly MJT, Carter DM, Davey RP, van Oudenaarden A, Roberts IN. Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole-genome resequencing. *Genome Res.* 2009; 19(4): 626–635. <https://doi.org/10.1101/gr.084517.108> PMID: 19141593

43. Inacio V, Rocheta M, Morais-Cecilio L. Molecular organization of the 25S-18S rDNA IGS of *Fagus sylvatica* and *Quercus suber*: a comparative analysis. *PLoS One*. 2014; 9(6): e98678. <https://doi.org/10.1371/journal.pone.0098678> PMID: 24893289;
44. Chou CH, Chiang YC, Chiang TY. Within- and between-individual length heterogeneity of the rDNA-IGS in *Miscanthus sinensis* var. *glaber* (Poaceae): phylogenetic analyses. *Genome*. 1999; 42(6): 1088–1093. <https://doi.org/10.1139/g99-057> PMID: 10659774.
45. Ioshikhes IP, Zhang MQ. Large-scale human promoter mapping using CpG islands. *Nat Genet*. 2000; 26(1): 61–63. <https://doi.org/10.1038/79189> PMID: 10973249
46. Maggini F, Gelati MT, Spolverini M, Frediani M. The intergenic spacer region of the rDNA in *Olea europaea* L. *Tree Genet Genomes*. 2008; 4(2): 293–298. <https://doi.org/10.1007/s11295-007-0109-x>
47. Sardana R, O'Dell M, Flavell R. Correlation between the size of the intergenic regulatory region, the status of cytosine methylation of rRNA genes and nucleolar expression in wheat. *Mol Gen Genet*. 1993; 236(2–3): 155–162. <https://doi.org/10.1007/BF00277107> PMID: 8437559
48. Polanco C, Perez de la Vega M. The structure of the rDNA intergenic spacer of *Avena sativa* L.: a comparative study. *Plant Mol Biol*. 1994; 25(4): 751–756. PMID: 8061326.
49. McMullen MD, Hunter B, Phillips RL, Rubenstein I. The structure of the maize ribosomal DNA spacer region. *Nucleic Acids Res*. 1986; 14(12): 4953–4968. <https://doi.org/10.1093/nar/14.12.4953> PMID: 3725589;
50. Borisjuk N, Hemleben V. Nucleotide sequence of the potato rDNA intergenic spacer. *Plant Mol Biol*. 1993; 21(2): 381–384. <https://doi.org/10.1007/BF00019953> PMID: 8425062.
51. Jo SH, Park HM, Kim SM, Kim HH, Hur CG, Choi D. Unraveling the sequence dynamics of the formation of genus-specific satellite DNAs in the family solanaceae. *Heredity*. 2011; 106(5): 876–885. <https://doi.org/10.1038/hdy.2010.131> PMID: 21063436;
52. Tautz D, Tautz C, Webb D, Dover GA. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species. Implications for molecular coevolution in multigene families. *J Mol Biol*. 1987; 195(3): 525–542. [https://doi.org/10.1016/0022-2836\(87\)90181-1](https://doi.org/10.1016/0022-2836(87)90181-1) PMID: 3116264.
53. Ambrose CD, Crease TJ. Evolution of the nuclear ribosomal DNA intergenic spacer in four species of the *Daphnia pulex* complex. *BMC Genet*. 2011; 12: 1–14. Epub 2011/01/26.
54. Ryu S, Do Y, Fitch DH, Kim W, Mishra B. Dropout alignment allows homology recognition and evolutionary analysis of rDNA intergenic spacers. *J Mol Evol*. 2008; 66(4): 368–383. Epub 2008/03/26. <https://doi.org/10.1007/s00239-008-9090-8> PMID: 18363028.
55. King K, Torres RA, Zentgraf U, Hemleben V. Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of cucurbitaceae. *J Mol Evol*. 1993; 36(2): 144–152. <https://doi.org/10.1007/BF00166250> PMID: 8433383.
56. Mcknight SL, Kingsbury RC, Spence A, Smith M. The distal transcription signals of the herpesvirus tk gene share a common hexanucleotide control sequence. *Cell*. 1984; 37(1): 253–262. [https://doi.org/10.1016/0092-8674\(84\)90321-0](https://doi.org/10.1016/0092-8674(84)90321-0) PMID: 6233005.
57. Doelling JH, Pikaard CS. Species-specificity of rRNA gene transcription in plants manifested as a switch in RNA polymerase specificity. *Nucleic Acids Res*. 1996; 24(23): 4725–4732. PMID: 8972859
58. Lassner M, Dvorak J. Preferential homogenization between adjacent and alternate subrepeats in wheat rDNA. *Nucleic Acids Res*. 1986; 14(13): 5499–5512. <https://doi.org/10.1093/nar/14.13.5499> PMID: 3737408
59. Takaiwa F, Kikuchi S, Oono K. The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in rice. *Plant Mol Biol*. 1990; 15(6): 255–259. <https://doi.org/10.1007/BF00039432>
60. Wang WX, Wei Z, Lam TW, Wang JW. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Sci Rep*. 2011; 1(2824): 55. Artn 55 <https://doi.org/10.1038/Srep00055> PMID: 22355574
61. Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007; 447(7143): 407–412. <https://doi.org/10.1038/nature05915> PMID: 17522673.
62. Santoro R, Li J, Grummt I. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet*. 2002; 32(3): 393–396. <https://doi.org/10.1038/ng1010> PMID: 12368916.
63. Chester M, Sykorova E, Fajkus J, Leitch AR. Single integration and spread of a Copia-like sequence nested in rDNA intergenic spacers of *Allium cernuum* (Alliaceae). *Cytogenet Genome Res*. 2010; 129(1–3): 35–46. Epub 2010/06/03. <https://doi.org/10.1159/000312959> PMID: 20516662.
64. Kato A, Lamb JC, Birchler JA. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci U S A*. 2004; 101(37): 13554–13559. <https://doi.org/10.1073/pnas.0403659101> PMID: 15342909;

65. Sang Y, Liang GH. Comparative physical mapping of the 18S-5.8S-26S rDNA in three sorghum species. *Genome*. 2000; 43(5): 918–922. <https://doi.org/10.1139/g00-063> PMID: 11081984.
66. Fukui K, Ohmido N, Khush GS. Variability in rDNA loci in the genus *Oryza* detected through fluorescence in situ hybridization. *Theor Appl Genet*. 1994; 87(8): 893–899. <https://doi.org/10.1007/BF00225782> PMID: 24190522
67. Vaio M, Speranza P, Valls JF, Guerra M, Mazzella C. Localization of the 5S and 45S rDNA sites and cpDNA sequence analysis in species of the Quadrifaria group of *Paspalum* (Poaceae, Paniceae). *Ann Bot*. 2005; 96(2): 191–200. <https://doi.org/10.1093/aob/mci168> PMID: 15911540;
68. Setta ND, Metcalfe CJ, Cruz GMQ, Bem LED, Vicentini R, Nogueira FTS, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics*. 2014; 15(1): 1–18.