ARTICLE IN PRESS

The Crop Journal xxx (xxxx) xxx





The Crop Journal



journal homepage: www.keaipublishing.com/en/journals/the-crop-journal/

Sequence analysis of *Erianthus arundinaceus* chromosome 1 isolated by flow sorting after genomic *in situ* hybridization in suspension

Shan Yang ^{a,d}, Petr Cápal ^c, Jaroslav Doležel ^c, Xueting Li ^a, Wang Qian ^a, Zhiqiang Wang ^b, Kai Zeng ^a, Peiting Li ^a, Hongkai Zhou ^d, Rui Xia ^e, Muqing Zhang ^{a,b,*}, Zuhu Deng ^{a,b,*}

^a National Engineering Research Center for Sugarcane & State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

^b State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources & Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, Guangxi, China

^c Institute of Experimental Botany of the Czech Academy of Sciences, Center of the Region Haná for Biotechnological and Agricultural Research, Olomouc CZ-77900, Czech Republic ^d College of Coastal Agricultural Sciences, Guangdong Ocean University, Zhanjiang 524088, Guangdong, China

e State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, South China Agricultural University, Guangzhou 510642, Guangdong, China

ARTICLE INFO

Article history: Received 22 September 2021 Revised 22 December 2021 Accepted 5 February 2022 Available online xxxx

Keywords: Flow cytometry GISHIS Chromosome sorting Genome analysis Sugarcane

ABSTRACT

Erianthus arundinaceus is a wild relative of sugarcane (Saccharum officinarum L.) with many desirable agronomic traits for sugarcane genetic improvement. However, limited knowledge of the complex genome of hexaploid E. arundinaceus has impeded the development of required molecular tools. Dissecting complex genomes into single chromosomes can simplify analyses. The flow-cytometric sorting of a single chromosome of *E. arundinaceus* in a Saccharum–Erianthus introgression line is reported. A novel approach called genomic in situ hybridization in suspension was used to discriminate the alien chromosome from sugarcane chromosomes at the same size. A total of 218,000 E. arundinaceus chromosome 1 (EaC1) were sorted to >97% purity and amplified DNA was sequenced using Illumina and PacBio technologies. The resulting assembly had a 70.93 Mb contig sequence with an N50 of 19.62 kb. A total of 56.69 Mb repeat sequences were predicted, accounting for 79.1% of the chromosome and 2646 genes having a total length of 1.84 Mb that represented 2.59% of the chromosome. Of these genes, 1877 (70.9%) genes were functionally annotated. The phylogenetic relationship of *E. arundinaceus* with other species using the chromosome 1 sequence revealed that E. arundinaceus was distantly related to Oryza sativa and Zea mays, followed by Sorghum bicolor, and was closely related to S. spontaneum and Saccharum spp. hybrids. This study provides the first insights into the characteristics of EaC1, and the results will provide tools to support molecular improvement and alien introgression breeding of sugarcane.

© 2022 Crop Science Society of China and Institute of Crop Science, CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Erianthus arundinaceus Retz. Jeswiet is a perennial grass native to southern China and comprises diploid, tetraploid, and hexaploid cytotypes with the basic chromosome number x = 10 [1]. *E. arundinaceus* is a wild relative of sugarcane (*Saccharum officinarum* L.), which produces over 85% of the global sugar supply and 40% of the bioethanol [2]. Modern sugarcane cultivars, highly heterozygous poly-aneuploids with somatic chromosome number 2n = 100-130, originated from a few interspecific crosses between octoploid *S. officinarum* (2n = 8x = 80), and wild relatives

(2n = 4x = 32 to 16x = 128). *S. officinarum* with high sugar content was domesticated about 8000 years ago, whereas wild *S. spontaneum* of varying ploidies contributes to disease resistance, vigor, and adaptability. Nobilization breeding gave rise to modern cultivars with complex genomes comprising a majority of chromosomes from *S. officinarum*, 10%–20% chromosomes from its wild relatives, and about 10% recombinant chromosomes [3,4]. Thus, modern sugarcane cultivars are vegetatively propagated and have a narrow genetic base.

A promising approach to introduce new genes and alleles to sugarcane is intergeneric crosses with wild relatives from the so-called "Saccharum complex", which comprises four genera: *Erianthus, Miscanthus, Narenga*, and *Sclerostachya* [5]. Of these, *E. arundinaceus* is the most attractive gene donor owing to its high

2214-5141/© 2022 Crop Science Society of China and Institute of Crop Science, CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: S. Yang, P. Cápal, J. Doležel et al., Sequence analysis of *Erianthus arundinaceus* chromosome 1 isolated by flow sorting after genomic *in situ* hybridization in suspension, The Crop Journal, https://doi.org/10.1016/j.cj.2022.02.001

^{*} Corresponding authors.

E-mail addresses: dengzuhu@163.com (Z. Deng), zmuqing@163.com (M. Zhang).

https://doi.org/10.1016/j.cj.2022.02.001

biomass productivity, superior ratooning ability, and tolerance to biotic and abiotic stresses [6]. However, attempts to transfer favorable genes from *E. arundinaceus* have been hampered by cross incompatibilities, low seed setting, and irregular chromosome transmission [7,8]. Thus, F₁ hybrids between *S. officinarum* and *E. arundinaceus* have 68 to 69 chromosomes rather than the expected 70, with 40 *S. officinarum* and 28–29 *E. arundinaceus* chromosomes [9,10]. Similar chromosome transmission irregularities were observed in hybrids between *Saccharum* spp. and *E. arundinaceus* [11].

Although the absence of some sugarcane chromosomes in hybrids could compromise the performance of newly developed lines, the loss of *E. arundinaceus* chromosomes should not be problematic, as the aim is to introduce only small regions of the E. arundinaceus genome. Elimination of unexpected E. arundinaceus chromosomes can be achieved by backcrossing F₁ hybrids to Saccharum spp. as demonstrated by Wu et al. [10] and Huang et al. [12]. The resulting plants in the BC_4 progeny carried only 1–6 *E*. arundinaceus chromosomes [13]. Intergeneric chromosome translocations were detected in the backcross progenies [12,13], indicating the occurrence of recombination events during meiosis that should facilitate introgression of small genome regions from *E. arundinaceus.* The appearance and agronomic traits of BC₄ progenies were similar to those of commercial sugarcane cultivars [14], confirming the potential of this strategy for sugarcane breeding and genetic improvement.

The development of chromosome introgression lines requires methods for identifying alien chromosomes and their segments during the breeding process. Genomic in situ hybridization (GISH) was performed to discriminate chromosomes from Saccharum spp. and E. arundinaceus and detect a sizeable intergeneric chromosome translocation [11,12,15]. However, GISH is slow and laborious, and higher-throughput methods are needed to screen larger populations. PCR with primers for 5S rDNA spacer and microsatellite loci demonstrated their suitability for identifying S. officinarum \times *E. arundinaceus* F₁ hybrids and the presence of *E. arundinaceus* chromatin in backcross (BC₁) progeny derived from crosses between selected F₁ clones and Saccharum spp. [8,16]. However, the presence of only certain chromosomes or chromosome regions of E. arundinaceus could be verified owing to the small number of genome loci amplified. More single-copy markers are needed, but their development has been slowed by the poor availability of genomic information for Saccharum spp. and E. arundinaceus.

Significant efforts have been made to sequence the sugarcane genome [17]. It is challenging to assemble and annotate the large and complex genome [11], comprising a mixture of homo(eo)logous chromosomes in different numbers [18]. Alignment of sugarcane genomic BAC library clones to the sorghum genome revealed 4660 clones corresponding to a "mosaic" monoploid with minimum tiling path [18]. The clones were sequenced, assembled, and annotated to yield a 382 Mb sequence with 25,316 predicted protein-coding genes. Soon afterward, a whole-genome shotgun approach was employed to characterize the sugarcane gene space [19]. After assembly of a 4.26-Gb sequence, 373,869 putative genes and promoter regions were predicted. However, to date, no reference genome of modern sugarcane has been published.

Genome complexity may also explain why a genome sequence is unavailable for the main parent of octoploid *S. officinarum* (2n = 8x = 80) with an average genome size of 7.88 Gb [20]. Autopolyploid *S. spontaneum* is the second main progenitor of modern sugarcane cultivars. A haploid plant AP85-441 from the octoploid cytotype SE208 (2n = 8x = 64) has been sequenced and assembled into chromosomes [21]. The haploid had 2n = 4x = 32and a genome size of 3.36 Gb [20]. A total of 35,156 BAC clones were sequenced using Illumina technology and genomic DNA by both PacBio and Illumina technologies. The resulting assembly of 32 pseudochromosomes comprises eight homologous chromosome groups of four chromosomes each, anchoring 2.9 Gb of the genome and defining 35,525 genes with alleles. Nascimento et al. [22] identified more genes (39,234) using a reference-guided transcript analysis of six different tissues.

Despite the high potential of *E. arundinaceus* for alien introgression breeding of sugarcane, knowledge of its genome remains poor. The only genome sequence currently available was produced recently by Zeng et al. [23]. A genome sequence of hexaploid line HN-92-77 (2n = 6x = 60) used in sugarcane breeding was surveyed, whose genome size was estimated by k-mer analysis to be 3.23 Gb, which is less than the 3.67 Gb [11] and 3.57 Gb [24] estimated using flow cytometry. The genome was sequenced to ~52-fold coverage by Illumina HiSeq 2100 platform (Illumina, Inc., San Diego, CA, USA) and assembled into 3156 Mb in 15,238,738 scaffolds with an N50 length of 216 bp [23]. A total of 36,616 microsatellite motifs were screened from the scaffolds and used for identifying *E. arundinaceus* chromatin introgressed into sugarcane.

Flow cytometric sorting can simplify complex genomes by dissecting them into single chromosomes [25]. This approach has been beneficial for mapping and cloning genes from alien introgressed chromosomes or their segments [26–28]. Flow cytometry requires samples in the form of single-chromosome suspensions prepared from plant materials enriched for mitotic metaphase cells [29,30]. Metcalfe et al. [31] and Yang et al. [32] demonstrated the feasibility of inducing a high degree of mitotic synchrony in sugarcane root tips cells, allowing for isolation of large quantities of chromosomes. Five groups of chromosomes were identified by flow cytometric analysis based on the relative DNA content of sugarcane chromosomes, with each group containing one or more homo(eo)logs of specific chromosomes [31]. Each group represented an ancestral subgenome, or more likely, subgenomes that have undergone full-genome duplication. A study by Metcalfe et al. [31] indicated that the sugarcane genome could be dissected into small parts but not into a single chromosome. A similar situation has been observed for other species and was overcome by sorting single chromosome copies [33,34] or by fluorescent labeling of DNA repeats on chromosomes prior to flow cytometry by fluorescence *in situ* hybridization in suspension [35,36].

Despite its potential for genome analysis, no attempts have been made to flow-sort *E. arundinaceus* chromosomes. The research aimed to develop a novel approach for discriminating the alien chromosome from the *Erianthus*-sugarcane introgression line carrying a single chromosome from *E. arundinaceus* based on GISHIS. The single chromosome of *E. arundinaceus* was flowsorted, purified, sequenced, and assembled to identify and annotate genes on the chromosome from *E. arundinaceus* and establish its synteny with related species.

2. Materials and methods

2.1. Plant materials and cell-cycle synchronization

The clone 1679-33 used in the study is the BC_5 progeny of a cross between *Saccharum* spp. and *Erianthus arundinaceus*. This clone contains 114 *Saccharum* spp. chromosomes and one *E. arundinaceus* chromosome and has no translocations between *E. arundinaceus* and *Saccharum* chromosomes. The maternal and paternal parents of the 1679-33 clone were LC03-1137 and YCE07-71, respectively, with the latter containing five *E. arundinaceus* chromosomes. The 1679-33 clone was planted in the sugarcane breeding field of Fujian Agriculture and Forestry University. Stems of adult plants were cut into single-bud segments, cleaned and soaked in 0.5% carbendazim solution for 24 h, placed in a plastic tray, covered with wet perlite, and incubated at 25 °C in the dark.

Cell-cycle synchronization and accumulation of metaphases were performed following Doležel et al. [37]. The roots were about 1.5 cm long; the segments were washed in ddH₂O, then transferred to a plastic tray filled with 150 mL 0.1× Hoagland Solution (0.1× HS) containing 2 mmol L⁻¹ hydroxyurea (HU), and incubated at 25 °C for 18 h in the dark. After the recovery treatment, the roots were immersed in 2.5 µmol L⁻¹ amiprophos-methyl (APM) solution and incubated for 3 h at 25 °C, then thoroughly rinsed in ddH₂-O, transferred to HU-free 0.1× HS, and incubated at 25 °C for 5 h with aeration.

2.2. Preparation of chromosome suspension

Chromosome suspensions were prepared with modifications following Doležel et al. [37]. Synchronized root tips were cut off and rinsed in ddH₂O, fixed in 2% (v/v) formaldehvde (10 mmol L⁻¹ Tris, 10 mmol L⁻¹ Na₂EDTA, 100 mmol L⁻¹ NaCl, 0.1% Triton X-100, 2% formaldehyde stock solution (37%, v/v), pH 7.5) in Tris buffer (10 mmol L⁻¹ Tris, 10 mmol L⁻¹ Na₂EDTA, 100 mmol L⁻¹ NaCl, pH 9.0) for 20 min at 4 °C, and then washed three times with Tris buffer for 5 min each at 4 °C. The terminal 1-1.5 mm of 30 root tips was cut off with a sterile scalpel, collected in 0.5 mL LB01 buffer (15 mmol L⁻¹ Tris, 2 mmol L⁻¹ Na₂EDTA·2H₂O, 0.5 mmol L⁻¹ spermine·4HCl, 80 mmol L⁻¹ KCl, 20 mmol L⁻¹ NaCl, 15 mmol L⁻¹ HOCH₂CH₂SH, 0.1% (v/v) Triton X-100, pH 7.5) [38], and homogenized with a Polytron PT1300 homogenizer (Kinematica AG, Litau, Switzerland) at 18,000 r min⁻¹ for 15 s to release chromosomes. The chromosome suspension was filtered through 50 µm nylon mesh and stored at 4 °C. The quality of the suspensions was checked microscopically after DAPI staining.

2.3. GISHIS

For preparing the probe, 1 µg *E. arundinacesus* gDNA was added to 6 µL dNTP mix with fluorescein-12-dUTP (FITC, Thermo Fisher Scientific, Waltham, MA, USA), 4 µL nick translation mix, and ddH_2O in a total volume of 20 μ L, which was then incubated at 15 °C for 5 h. The correct probe size (<800 bp) was confirmed by gel electrophoresis. The hybridization solution contained 50 µL formamide, 10 μ L 20 \times saline sodium citrate (SSCs), and 20 μ L *E. arun*dinacesus gDNA probe, in a total volume of 100 µL. The probe was denatured at 99 °C for 5 min and then cooled on ice before 200 µL formamide was added to 500 µL of filtered chromosomal suspension and mixed thoroughly. The samples were denatured at 99 °C for 5 min and immediately put into ice-water for 10 min. Then, 100 µL denatured hybridization solution was added to the sample tube, which was incubated for 5 h at 37 °C in a Hula Mixer sample mixer (Thermo Fisher Scientific) that rotates and shakes continually. After hybridization, the samples were centrifuged at 5000 r min^{-1} for 8 min, and the pellet was resuspended in 500 μ L LB01 buffer. The effectiveness of chromosome labeling was assessed by fluorescence microscopy. Chromosome suspensions were stored at 4 °C prior to flow sorting.

2.4. Flow cytometric analysis and sorting

Chromosome suspensions were filtered through a 20 μ m nylon mesh and stained with DAPI (2 μ g mL⁻¹). Chromosomal populations were gated on FSC-A vs. DAPI-A parameters, with subsequent dependent gates representing *E. arundinaceus* chromosomes drawn in a DAPI-A vs. FITC-A (log) scatter plot. The samples were separated in a FACSAria II SORP flow cytometer (BD Biosciences, San Jose, CA, USA) at a rate of 3000–4000 events per second with *E. arundinaceus* chromosomes sorted into ddH₂O water. The purity of the sorted chromosomes was evaluated microscopically from 1000 to 2000 chromosomes sorted onto a slide (repeat 3 slides, 6 screens/slide). *E. arundinaceus* chromosomes were identified by their FITC fluorescence, whereas *Saccharum* spp. chromosomes carried only the DAPI signal.

2.5. DNA sequencing and genome assembly

The sorted E. arundinaceus chromosomes were treated with proteinase and their DNA was purified and amplified using a Genomi-Phi V2 DNA Amplification Kit (GE Healthcare, Chalfont St. Giles, UK) following Šimková et al. [39]. Two sequencing libraries were constructed from the amplified DNA. The gDNA was randomly broken into 350-bp fragments by sonication for the Illumina library and into 11-kb fragments using g-TUBE (https://www.pacb.com/ wp-content/uploads/2015/09/Covaris-g-TUBE-protocol.pdf) for the PacBio library. A short-read Illumina library was prepared with a NEBNext Ultra II DNA Library Prep Kit (Illumina, Inc.), and a longread PacBio sequencing library was prepared with PacBio's SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). The short-read library was sequenced on an Illumina NovaSeq 6000 instrument (Illumina, Inc., San Diego, CA, USA) in 150-bp paired-end mode, and the long-read library was sequenced on a PacBio Sequel II instrument (Pacific Biosciences) in Circular Consensus Sequencing (CCS) mode. Short-read sequences of the single chromosome from *E. arundinaceus* were checked with FastQC (https://www.bioinformatics.babraham.ac. uk/projects/fastqc/), trimmed with TrimGalore (https://github.com/FelixKrueger/TrimGalore) and aligned to the Sorghum bicolor genome using bwa [40] before conversion to BAM format using Samtools [41]. Coverage depth was characterized using bedtools [42] and plotted using a Python script. The PacBio CCS reads were quality-filtered using SMRTlink 8.0 (https://www.pacb.com/wpcontent/uploads/SMRT_Link_Troubleshooting_Guide_v80.pdf) before being subjected to single-chromosome genome assembly with HiFiASM software (https://github.com/chhylp123/hifiasm).

2.6. Gene prediction and annotation

2.6.1. DNA repeat prediction

Two methods were used to predict DNA repeats. (i) Homologous prediction: the DNA repeat sequence databases RepBase [43], RepeatMasker [44], and RepeatProteinMask [44] were used to predict sequences similar to known repeat sequences. (ii) *Ab initio* structure prediction method: RepeatModeler [45] was used to establish a repeat sequence library for *ab initio* structure prediction, and then RepeatMasker was used to make the prediction, and TRF [46] software was used to find tandem repeat sequences in the genome.

2.6.2. Gene prediction

Three methods were used for gene prediction. (i) The *ab initio* structure prediction method used Augustus 2.4 [47] and SNAP [48] software to make gene predictions. (ii) Homologous prediction: BLAST [49] and GeneWise [50] were used for gene prediction based on homologous species. (iii) PASA 2.0.2 [51] was used to predict genes based on transcriptome unigene sequences. Then, EVM 1.1.1 [52] was used to remove redundancy and integrate the predicted results.

2.6.3. Non-coding RNA prediction

The genome-wide alignment against Rfam [53] was performed with BLAST to identify microRNA, rRNA, snRNA, and snoRNA. tRNAscan-SE [54] was used to identify tRNA.

2.6.4. Gene function annotation

The predicted genes were aligned with BLAST 2.2.31 [49] against protein and gene databases including SwissProt [55], NT

(https://www.ncbi.nlm.nih.gov/nucleotide/), NR [56], PFAM [57], eggnog [58], KOG [59], GO [60], and KEGG [61], with the e-value threshold set to 1e-5 for gene function annotation.

2.7. Identification of syntenic genome regions in Sorghum bicolor and *S. spontaneum*

All 2646 gene sequences predicted from the scaffolds of *E. arundinaceus* chromosome 1 (EaC1) were compared to identify syntenic regions in the two genomes by BLAST against the coding sequences (CDS) from *Sorghum bicolor* and *S. spontaneum*. The following filtering criteria were applied: the first BLAST hits showing at least 70.0% identity and a minimum alignment of 200 bp were considered to be homologous. Then, the recurring gene events were analyzed using Multiple Collinear Scanning (MCScanX) toolkits [62]. The results were visualized with TBtools software [63].

2.8. Analysis of evolutionary relationships

The CDS of chromosome 1 from *E. arundinaceus, Sorghum bicolor, S. spontaneum, Saccharum* hybrids, *Zea mays,* and *Oryza sativa* were aligned using MAFFT v7.205 [64]. Genome regions with low alignment scores were removed using Gblocks v0.91b [65] (parameter: -b5 = h), and "supergenes" were obtained by connecting all matching gene family sequences of each species. The JTT + 1 model was used for evolutionary tree construction with the ModelFinder detection tool [66] from the IQ-Tree 1.6.11 package [67], and the maximum likelihood method was used to build an evolutionary tree with a bootstrap value of 1000.

3. Results

3.1. GISHIS

No individual chromosome peaks were resolvable in univariate histograms obtained after flow cytometric analysis of DAPI-stained chromosomes isolated from the 1679-33 clone (Fig. S1A). Each group of 1000 particles was sorted to verify the quality of chromosome suspension from 1 to 4 sort windows (Fig. S1B). Particles in these sort windows were chromosomes, indicating that the chromosome suspension could be used for subsequent experiments (Fig. S1C–F). The chromosome population of *E. arundinaceus* could be identified on bivariate flow karyotypes by DAPI-A vs. FITC-A fluorescence after GISHIS (Fig. 1). The chromosome morphology was retained intact after GISHIS, as represented by a similar univariate histogram (Fig. 2A vs. Fig. S1A).

3.2. Chromosome flow sorting after GISHIS

A total of 1200 particles were sorted from sort windows P2 and P3 (Fig. 2B) onto a microscopic slide and evaluated by fluorescence microscopy. All chromosomes sorted from the P2 sort window carried FITC fluorescence, and the purity in the sorted fraction was 97.85% (Fig. 3). The particles sorted from the P3 sort window contained a mixture of *Saccharum* chromosomes and *E. arundinaceus* chromosomes, with approximately 75% of chromosomes carrying FITC signals (Fig. S2). Based on these observations, a total of 218,000 chromosomes were flow-sorted from the P2 sort window and used for DNA amplification.

3.3. Multiple displacement amplification (MDA)

The chromosomes sorted from the P2 sort window were deproteinized and purified, and their DNA was amplified to yield 29.63 μ g DNA. The size ranged from 3 to 45 kb, with the central band of about 45 kb (Fig. S3). The DNA fragments obtained after the amplification was checked by 1% agarose gel electrophoresis.

3.4. Chromosome sequencing and identification

The amplified DNA was initially sequenced with 150-bp pairedend reads with an Illumina NovaSeq 6000, yielding 10.45 Gb raw data with 91.29% of Q30. The reads were aligned to the *Sorghum bicolor* genome (BioProject: PRJNA13876) to identify the homologous group to which the sorted *E. arundinaceus* chromosome belonged. The percentages of mappable and unique reads were 41.68% and 58.32%, respectively, with the reads mapping mainly to chromosome 1 of *Sorghum bicolor* (Fig. 4), indicating that the sorted chromosomes were *E. arundinaceus* chromosome 1 (EaC1).

3.5. DNA sequencing and assembly

PacBio sequencing of the amplified chromosomal DNA yielded 72.43 Gb reads with an N50 length of 179,114 bp (Table S1). Removal of sequencing adapters and low-quality reads yielded 58.60 Gb subreads with an N50 length of 9819 bp (Table S1). Final CCS yielded 1.33 Gb, with an N50 of 10,499 bp (Table S1). The CCS assembly of EaC1 had a total length of 70.93 Mb consisting of 3696 contigs, with a contig N50 of 19,624 bp and 44.90% GC content. Although the total length of the assembled chromosome was lower than its predicted length (~119 Mb), the contig N50 was low given that the N50 of CCS reads was 10.50 kb. CCS reads were aligned against *Sorghum* chromosome 1, and the resulting mean length of



Fig. 1. Microscopic examination of a crude chromosome suspension. The suspension was stained with DAPI (blue) and subjected to GISHIS to label *E. arundinaceus* chromosomes (green). (A) All particles show DAPI fluorescence. (B) Only *E. arundinaceus* chromosomes are labeled by FITC. Red arrows indicate chromosomes with a green signal. Scale bars, 5 μm.



Fig. 2. Flow karyotypes obtained by separating chromosomes isolated from the *Saccharum* spp. hybrid-*E. arundinaceus* chromosome introgression line 1679-33. (A) Monovariate plot of DAPI-A fluorescence. (B) Bivariate plot of DAPI-A vs. FITC-A. The positions of the sort are shown in windows P2 and P3.



Fig. 3. Microscopic examination of a chromosome population sorted from the P2 sort window. (A) and (C) show green signals. (B) and (D) show green and blue signals together. Scale bars, 5 µm.



Fig. 4. Mapping depth of next-generation sequence reads on the Sorghum bicolor genome.

mapped fragments was 1725 bp, explaining the discontinuity of the assembly.

3.6. DNA sequence annotation

A total of 2646 genes were predicted, with a total length of 1.84 Mb, accounting for 2.59% of the chromosome assembly (Table 1). Of these, 1877 genes were annotated, representing 70.94% of the total number of predicted genes (Table S2). For non-coding RNAs, 861 with a total length of 109,391 bp were predicted, accounting for 0.153% of the assembly (Table S3). A total of 56.44 Mb repeat sequences were detected in the assembly of EaC1 using different prediction methods, accounting for 79.58% of the assembly (Table S4). Long terminal repeat (LTR) retrotransposons, DNA transposons, long interspersed repetitive elements (LINE), and short interspersed repetitive elements (SINE) occupied respectively 36.26, 9.36, 4.54, and 0.065 Mb in the EaC1 assembly, representing 51.11%, 13.20%, 6.40%, and 0.09%, of the chromosome (Table S5).

3.7. Collinearity analysis

All 2646 predicted gene sequences from EaC1 scaffolds were used to identify syntenic regions in *Sorghum bicolor* and *S. spontaneum* genomes. Total 2666 and 5542 genes were annotated from the EaC1 scaffolds in *Sorghum bicolor* and *S. spontaneum*. After filtering, 2161 out of 2646 identified homologous genes (81.67%) were shared between *Sorghum* and *S. spontaneum*, suggesting the presence of collinear regions among *E. arundinaceus*, *Sorghum*, and *S. spontaneum*. The 2161 homologous genes of EaC1 were plotted according to their positions on the chromosomes of their respective species, and clear syntenic regions between *Sorghum* and *S. spontaneum* were observed (Fig. 5). The EaC1 syntenic regions in *Sorghum* and *S. spontaneum* were distributed on chromosome 1. There were syntenic regions on other chromosomes of the two species, but the gene density was low.

3.8. Evolutionary relationships

As expected, *E. arundinaceus* showed the closest evolutionary relationship with *S. spontaneum* and *Saccharum* hybrids. *Sorghum bicolor* branched in the second tier, marking it more evolutionary distance (Fig. 6). The most significant evolutionary distance was found between *E. arundinaceus*, *O. sativa*, and *Zea mays* (Fig. 6).

4. Discussion

Alien introgression breeding is an attractive approach to improving crops by introducing desired genes from wild relatives [68]. Sugarcane would benefit significantly from the introgression of genes from its wild relatives. One of them is *E. arundinaceus*, a potential donor of genes for disease resistance, drought resistance,

Table 1	
---------	--

Predictive	gene	information	statistics	for	EaC1.

and high biomass [6]. Unfortunately, knowledge of its genome remains limited [23]. Examination and sequencing of complex plant genomes can be simplified by dissecting the genomes into individual chromosomes using flow cytometric sorting [69]. Chromosome sorting technology has been developed for 29 species, including sugarcane [25,31]. Metcalfe et al. [31] demonstrated the feasibility of dissecting the sugarcane genome into small parts, though not single chromosomes. Flow cytometric classification of chromosomes according to DNA content alone resulted in flow karyotypes having composite peaks representing more than one chromosome of similar sizes. This problem was encountered for several species [29,70]. One approach to overcoming this problem was labeling DNA repeats on chromosomes prior to flow cytometry using fluorescence *in situ* hybridization in suspension [35,36].

The result presented here represents the first demonstration that DNA of *E. arundinaceus* chromosomes in suspension can be fluorescently labeled using a genomic probe and that GISHIS permits purification of alien introgressed chromosomes by flow cytometry. As sugarcane chromosome introgression lines carrying chromosomes of *E. arundinaceus* were available [13], GISH in suspension was used to discriminate the introgressed chromosomes in analogy with FISH in suspension. If the introgression lines carry different alien chromosomes, this approach should make it possible to sequence single chromosomes of E. arundinaceus. The same strategy could be used to sequence other *E. arundinaceus* chromosomes introgressed into sugarcane. This method should generally be applicable for any species and, in principle, could also facilitate the sorting of translocated chromosomes if the translocated segment of an alien chromosome is sufficiently large to allow detection.

The short reads from the sorted chromosome 1 (EaC1) were sequenced and mapped to *Sorghum* chromosome 1. EaC1 syntenic regions were distributed on chromosome 1 from *Sorghum* and *S. spontaneum*. The long-range organization of chromosomes was conserved relative to other species, and no exchange or translocation occurred during evolution [21]. We inferred that EaC1 was relatively conserved based on its high collinearity with *S. spontaneum* and *Sorghum*.

The phylogenetic relationship of *E. arundinaceus* confirmed its evolutionary distance from *O. sativa* and *Zea mays* and that it was close to *Sorghum* and closest to *Saccharum* hybrids and *S. spontaneum*. Thus, although *E. arundinaceus*, *Sorghum*, *Saccharum* hybrids, and *S. spontaneum* had a common ancestor, their divergence times were different. Tsuruta et al. [71] showed that *Sorghum* was more closely related to *S. hybrids* than *E. arundinaceus* and estimated that *E. arundinaceus* diverged from the subtribe *Sorghinae* before the divergence of *Sorghum bicolor*. The difference between our results and those of Tsuruta et al. [71] could be due to the analysis of different genotypes, and its explanation invites further study.

Considering that the genome size of hexaploid *E. arundinaceus* is \sim 3.57 Gb [24], the expected average size of one chromosome should be \sim 119 Mb, a value greater than the PacBio assembly length of 70.93 Mb. The PacBio assembly comprised 3696 contigs

Software	Species	Number of genes	Mean gene length (bp)	Mean CDs length (bp)	Mean exon length (bp)	Mean intron length (bp)
Augustus		2274	533.51	515.04	413.12	75.84
SNAP		1801	417.40	356.74	293.11	280.40
PASA		1210	1076.72	550.09	574.80	777.54
BLAST, GeneWise	Sorghum	2578	1070.28	506.08	186.81	331.12
	O. sativa	2197	986.19	468.02	195.11	371.44
	S. hybrids	2732	784.25	427.80	187.48	279.09
	S. spontaneum	3398	893.34	421.16	176.54	341.77
	Zea mays	2492	1078.05	500.42	187.27	346.44
EVM		2646	695.81	503.84	321.94	340.74



Fig. 5. Syntenic regions for *E. arundinaceus* chromosome 1 annotated genes in the Sorghum bicolor and S. spontaneum genome. Blue-green chromosomes indicate S. spontaneum chromosomes. Gray chromosomes indicate Sorghum bicolor chromosomes.



Fig. 6. Diagram of evolutionary relationships among six species.

with an N50 of 19,624 bp. The assembly quality was probably affected by sequencing amplified chromosomal DNA [28,72–74]. To our knowledge, our first report of PacBio sequencing of chromosomal DNA and the assembly is better than those of chromosome assemblies obtained after sequencing using the Illumina platform [28,74]. We expect that this high-quality assembly of *E. arundinaceus* will facilitate gene cloning, the development of DNA markers, and the development of chromosome typing markers in this species.

Our analysis predicted 56.44 Mb repeat sequences on EaC1, accounting for 79.58% of its sequence, of which 51.11% were LTR repeats, 13.20% were DNA TEs, 6.40% were LINEs, and 0.09% were SINES. The proportion of repeat sequences is consistent with various DNA repeats in the *Sorghum, Zea mays, O. sativa*, and *S. spontaneum* genomes [21,75–77]. Total 2646 encoding genes in EaC1 were predicted, with a total length of 1.84 Mb accounting for 2.59% of the chromosome. The functions of 1877 (70.94%) of these genes were annotated. The low number of coding genes in EaC1 might be due to the poor assembly of regions around the centromeres. Zhang et al. [21] identified 35,525 putative coding genes in *S. spontaneum*, and the genome of diploid *Sorghum bicolor* contains 34,129 genes [78].

To conclude, this new approach allowed purification of a single *E. arundinaceus* chromosome by flow-cytometric sorting, and the DNA sequence of this chromosome was determined. Genes on the chromosome were identified and annotated from the assembled sequences, and synteny with related species was established. This study provides initial insights into the characteristics of EaC1, and the results will support molecular improvement and alien introgression breeding of sugarcane.

Data availability

The EaC1 genome assembly has been deposited in the NCBI database under BioProject number PRJNA715955 and BioSample number SAMN18388831.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Shan Yang: Visualization, Writing – original draft. Petr Cápal: Visualization, Data curation, Writing – review & editing. Jaroslav Doležel: Funding acquisition, Project administration, Writing – review & editing. Xueting Li: Methodology, Software. Wang Qian: Investigation. Zhiqiang Wang: Methodology. Kai Zeng: Resources. Peiting Li: Visualization, Data curation. Hongkai Zhou: Formal analysis. Rui Xia: Funding acquisition. Muqing Zhang: Methodology, Resources, Supervision, Funding acquisition, Writing – review & editing. Zuhu Deng: Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (31771863), Science and Technology Major Project of

the Fujian Province of China (2015NZ0002-2), Special Fund for Scientific and Technological Innovation of the Fujian Agriculture and Forestry University (KFA17168A), Doctoral Students of Fujian Agriculture and Forestry University Going Abroad to Cooperative Research (324-112110082), Key Laboratory of Conservation and Utilization of Subtropical Agricultural Biological Resources (SKLCUSA-a201912). Jaroslav Doležel and Petr Cápal were supported by ERDF Project "Plants as a tool for sustainable global development" (CZ.02.1.01/0.0/16_019/0000827). None of the funding institutions played a role in the study design, data collection, analysis, or manuscript writing. We thank the Key Laboratory of Conservation and Utilization of Subtropical Agricultural Biological Resources for providing the flow cytometer used in this study. We thank Annoroad Gene Technology company (Beijing, China) for Illumina and PacBio sequencing.

Appendix A. Supplementary data

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2022.02.001.

References

- Q. Cai, J.C. Wen, Y.H. Fan, L.P. Wang, L. Ma, Chromosome analysis of Saccharum L. and related plants, Southwest China, J. Agric. Sci. 18 (2002) 16–19.
- [2] P.H. Moore, F.C. Botha, Sugarcane: Physiology, Biochemistry, and Functional Biology, John Wiley & Sons Inc, Hoboken, NJ, USA, 2013.
- [3] G. Piperidis, N. Piperidis, A. D'Hont, Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane, Mol. Genet. Genomics 284 (2010) 65–73.
- [4] A. D'Hont, L. Grivet, P. Feldmann, J.C. Glaszmann, S. Rao, N. Berding, Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics, Mol. Gen. Genet. 250 (1996) 405–413.
- [5] C. Grassl, Taxonomy of Saccharum relatives: Sclerostachya, Narenga, and Erianthus, in: Proceedings of 14th Congress of the International Society of Sugar Cane Technologists (ISSCT), New Orleans, LA, USA, 1972, pp. 240–248.
- [6] V.A. Amalraj, N. Balasundaram, On the taxonomy of the members of *'Saccharum* complex, Genet. Resour. Crop Evol. 53 (2006) 35–41.
- [7] G. Piperidis, M.J. Christopher, B.J. Carroll, N. Berding, A. D'Hont, Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species *Erianthus arundinaceus*, Genome 43 (2000) 1033–1037.
 [8] Q. Cai, K. Aitken, H.H. Deng, X.W. Chen, C. Fu, P.A. Jackson, C.L. McIntyre,
- [8] Q. Cai, K. Aitken, H.H. Deng, X.W. Chen, C. Fu, P.A. Jackson, C.L. McIntyre, Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers, Plant Breed. 124 (2005) 322–328.
- [9] A.D. Hont, P. Rao, P. Feldmann, L. Grivet, N. Islam-Faridi, P. Taylor, J.C. Glaszmann, Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum × Erianthus arundinaceus*, with molecular markers and DNA *in situ* hybridisation, Theor. Appl. Genet. 91 (1995) 320–326.
- [10] J. Wu, Y. Huang, Y. Lin, C. Fu, S. Liu, Z. Deng, Q. Li, Z. Huang, R. Chen, M. Zhang, Unexpected inheritance pattern of *Erianthus arundinaceus* chromosomes in the intergeneric progeny between *Saccharum* spp. and *Erianthus arundinaceus*, PLoS ONE 9 (2014) e110390.
- [11] B. Pachakkil, Y. Terajima, N. Ohmido, M. Ebina, S. Irei, H. Hayashi, H. Takagi, Cytogenetic and agronomic characterization of intergeneric hybrids between Saccharum spp. hybrid and Erianthus arundinaceus, Sci. Rep. 9 (2019) 1748.
- [12] Y. Huang, J. Wu, P. Wang, Y. Lin, C. Fu, Z. Deng, Q. Wang, Q. Li, R. Chen, M. Zhang, Characterization of chromosome inheritance of the intergeneric BC₂ and BC₃ progeny between *Saccharum* spp. and *Erianthus arundinaceus*, PLoS ONE 10 (2015) e0133722.
- [13] S. Yang, K. Zeng, K. Chen, J. Wu, Q. Wang, X. Li, Z. Deng, Y. Huang, F. Huang, R. Chen, Chromosome transmission in BC₄ progenies of intergeneric hybrids between *Saccharum* spp. and *Erianthus arundinaceus* (Retz.) Jeswiet, Sci. Rep. 9 (2019) 2528.
- [14] Q. Wnag, J. Xie, C. Zhang, Q. Wu, J. Wu, C. Fu, S. Liu, Analysis on breeding values of sugarcane with consanguinity of *E. arundinaceus* and their crosses, Chin. J. Trop. Crops 38 (2017) 1274–1279.
- [15] J. Shore, N. Piperidis, J.W. Chen, H.H. Deng, L.P. Wang, P. Jackson, G. Piperidis, GISH characterization of *Erianthus arundinaceus* chromosomes in three generations of sugarcane intergeneric hybrids, Genome 53 (2010) 331–336.
- [16] S. Fukuhara, Y. Terajima, S. Irei, T. Sakaigaichi, K. Ujihara, A. Sugimoto, M. Matsuoka, Identification and characterization of intergeneric hybrid of commercial sugarcane (*Saccharum* spp. hybrid) and *Erianthus arundinaceus* (Retz.) Jeswiet, Euphytica 189 (2013) 321–327.
- [17] P.P. Thirugnanasambandam, N.V. Hoang, R.J. Henry, The challenge of analyzing the sugarcane genome, Front. Plant Sci. 9 (2018) 616–616.
- [18] O. Garsmeur, G. Droc, R. Antonise, J. Grimwood, B. Potier, K. Aitken, J. Jenkins, G. Martin, C. Charron, C. Hervouet, L. Costet, N. Yahiaoui, A. Healey, D. Sims, Y. Cherukuri, A. Sreedasyam, A. Kilian, A. Chan, M.A. Van Sluys, K. Swaminathan,

C. Town, H. Bergès, B. Simmons, J.C. Glaszmann, E. van der Vossen, R. Henry, J. Schmutz, A.D. Hont, A mosaic monoploid reference sequence for the highly complex genome of sugarcane, Nat. Commun. 9 (2018) 2638.

- [19] G.M. Souza, M.A. Van Sluys, C.G. Lembke, H. Lee, G.R.A. Margarido, C.T. Hotta, J. W. Gaiarsa, A.L. Diniz, M.M. Oliveira, S.S. Ferreira, M.Y. Nishiyama, F. Ten-Caten, G.T. Ragagnin, P.M. Andrade, R.F. de Souza, G.G. Nicastro, R. Pandya, C. Kim, H. Guo, A.M. Durham, M.S. Carneiro, J. Zhang, X. Zhang, Q. Zhang, R. Ming, M.C. Schatz, B. Davidson, A.H. Paterson, D. Heckerman, Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop, GigaScience 8 (2019) giz129.
- [20] J. Zhang, C. Nagai, Q. Yu, Y.B. Pan, T. Ayala-Silva, R.J. Schnell, J.C. Comstock, A.K. Arumuganathan, R. Ming, Genome size variation in three *Saccharum* species, Euphytica 185 (2012) 511–519.
- [21] J. Zhang, X. Zhang, H. Tang, Q. Zhang, X. Hua, X. Ma, F. Zhu, T. Jones, X. Zhu, J. Bowers, Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L, Nat. Genet. 50 (2018) 1565.
- [22] L.C. Nascimento, K. Yanagui, J. Jose, E.L.O. Camargo, M.C.B. Grassi, C.P. Cunha, J. A. Bressiani, G.M.A. Carvalho, C.R. Carvalho, P.F. Prado, P. Mieczkowski, G.A.G. Pereira, M.F. Carazzolle, Unraveling the complex genome of *Saccharum spontaneum* using polyploid gene assembler, DNA Res. 26 (2019) 205–216.
- [23] Q. Zeng, Z. Huang, Q. Wang, J. Wu, X. Feng, Y. Qi, Genome survey sequence and the development of simple sequence repeat (SSR) markers in *Erianthus* arundinaceus, Sugar Tech 23 (2021) 77–85.
- [24] J. Yan, J. Zhang, K. Sun, D. Chang, S. Bai, Y. Shen, L. Huang, J. Zhang, Y. Zhang, Y. Dong, Ploidy level and DNA content of *Erianthus arundinaceus* as determined by flow cytometry and the association with biological characteristics, PLoS ONE 11 (2016) e0151948.
- [25] J. Zwyrtková, H. Šimková, J. Doležel, Chromosome genomics uncovers plant genome organization and function, Biotechnol. Adv. 46 (2021) 107659.
- [26] M. Bansal, N.M. Adamski, P.I. Toor, S. Kaur, I. Molnár, K. Holušová, J. Vrána, J. Doležel, M. Valárik, C. Uauy, P. Chhuneja, Aegilops umbellulata introgression carrying leaf rust and stripe rust resistance genes *Lr76* and *Yr70* located to 9.47-Mb region on 5DS telomeric end through a combination of chromosome sorting and sequencing, Theor. Appl. Genet. 133 (2020) 903–915.
- [27] A.K. Thind, T. Wicker, H. Šimková, D. Fossati, O. Moullet, C. Brabant, J. Vrána, J. Doležel, S.G. Krattinger, Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly, Nat. Biotechnol. 35 (2017) 793–796.
- [28] J. Xiao, K. Dai, L. Fu, J. Vrána, M. Kubaláková, W. Wan, H. Sun, J. Zhao, C. Yu, Y. Wu, M. Abrouk, H. Wang, J. Doležel, X. Wang, Sequencing flow-sorted short arm of *Haynaldia villosa* chromosome 4V provides insights into its molecular structure and virtual gene order, BMC Genomics 18 (2017) 791.
- [29] J. Doležel, J. Greilhuber, J. Suda, Flow cytometry with plant cells: analysis of genes, chromosomes and genomes, in: J. Doležel, J. Greilhuber, J. Suda (Eds.), Flow Cytometry with Plants: an Overview, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2007, pp. 41–65.
- [30] J. Doležel, J. Vrána, J. Šafář, J. Bartoš, M. Kubaláková, H. Šimková, Chromosomes in the flow to simplify genome analysis, Funct. Integr. Genomics 12 (2012) 397–416.
- [31] C.J. Metcalfe, J. Li, D. Giorgi, J. Doležel, N. Piperidis, K.S. Aitken, Flow cytometric characterisation of the complex polyploid genome of *Saccharum officinarum* and modern sugarcane cultivars, Sci. Rep. 9 (2019) 19362.
- [32] S. Yang, K. Zeng, L. Luo, W. Qian, Z. Wang, J. Doležel, M. Zhang, X. Gao, Z. Deng, A flow cytometry-based analysis to establish a cell cycle synchronization protocol for Saccharum spp., Sci. Rep. 10 (2020) 5016.
- [33] P. Cápal, N. Blavet, J. Vrána, M. Kubaláková, J. Doležel, Multiple displacement amplification of the DNA from single flow-sorted plant chromosome, Plant J. 84 (2015) 838–844.
- [34] J. Kreplak, M.A. Madoui, P. Cápal, P. Novák, K. Labadie, G. Aubert, P.E. Bayer, K. K. Gali, R.A. Syme, D. Main, A. Klein, A. Bérard, I. Vrbová, C. Fournier, L. d'Agata, C. Belser, W. Berrabah, H. Toegelová, Z. Milec, J. Vrána, H. Lee, A. Kougbeadjo, M. Térézol, C. Huneau, C.J. Turo, N. Mohellibi, P. Neumann, M. Falque, K. Gallardo, R. McGee, B. Tar'an, A. Bendahmane, J.M. Aury, J. Batley, M.C. Le Paslier, N. Ellis, T.D. Warkentin, C.J. Coyne, J. Salse, D. Edwards, J. Lichtenzveig, J. Macas, J. Doležel, P. Wincker, J. Burstin, A reference genome for pea provides insight into legume genome evolution, Nat. Genet. 51 (2019) 1411–1422.
- [35] D. Giorgi, A. Farina, V. Grosso, A. Gennaro, C. Ceoloni, S. Lucretti, FISHIS: fluorescence in situ hybridization in suspension and chromosome flow sorting made easy, PLoS ONE 8 (2013) e57994.
- [36] I. Molnár, J. Vrána, V. Burešová, P. Cápal, A. Farkas, É. Darkó, A. Cseh, M. Kubaláková, M. Molnár-Láng, J. Doležel, Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their synteny with wheat, Plant J. 88 (2016) 452–467.
- [37] J. Doležel, J. Číhalíková, S. Lucretti, A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L., Planta 188 (1992) 93– 98.
- [38] J. Loureiro, E. Rodriguez, J. Dolezel, C. Santos, Comparison of four nuclear isolation buffers for plant DNA flow cytometry, Ann. Bot. 98 (2006) 679–689.
- [39] H. Simková, J.T. Svensson, P. Condamine, E. Hribová, P. Suchánková, P.R. Bhat, J. Bartos, J. Safár, T.J. Close, J. Dolezel, Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley, BMC Genomics 9 (2008) 294.
- [40] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760.

- [41] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, Genome project data processing S: the sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
- [42] A.R. Quinlan, I.M. Hall, BEDTools, Bioinformatics 26 (2010) 841-842.
- [43] J. Jurka, V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase update, a database of eukaryotic repetitive elements, Cytogenet. Genome Res. 110 (2005) 462–467.
- [44] N. Chen, Using repeatmasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics 5 (2004) 4.10.1–4.10.14.
- [45] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, A.F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 9451–9457.
- [46] I. Mender, J. Shay, Telomere restriction fragment (TRF) analysis, Bio-protocol 5 (2015) e1658.
- [47] M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics 19 (2003) ii215-ii225.
- [48] I. Korf, Gene finding in novel genomes, BMC Bioinformatics 5 (2004) 59.
- [49] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment
- search tool, J. Mol. Biol. 215 (1990) 403–410.
 [50] E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise, Genome Res. 14 (2004) 988–995.
- [51] M. Campbell, B. Haas, J. Hamilton, S. Mount, C. Buell, Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*, BMC Genomics 7 (2006) 327.
- [52] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C.R. Buell, J.R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments, Genome Biol. 9 (2008) R7.
- [53] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S.R. Eddy, A. Bateman, Rfam: annotating non-coding RNAs in complete genomes, Nucleic Acids Res. 33 (2005) D121–D124.
- [54] T. Lowe, S. Eddy, T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
- [55] A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, Nucleic Acids Res. 28 (2000) 45–48.
- [56] A. Marchler-Bauer, S. Lu, J. Anderson, F. Chitsaz, M. Derbyshire, C. DeWeese-Scott, J. Fong, L. Geer, R. Geer, N. Gonzales, M. Gwadz, D. Hurwitz, J. Jackson, Z. Ke, C. Lanczycki, F. Lu, G. Marchler, M. Mullokandov, M. Omelchenko, S. Bryant, CDD: a Conserved Domain Database for the functional annotation of proteins, Nucleic Acids Res. 39 (2011) D225–D229.
- [57] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucleic Acids Res. 42 (2013) D222–D230.
- [58] L.J. Jensen, P. Julien, M. Kuhn, C. von Mering, J. Muller, T. Doerks, P. Bork, eggNOG: automated construction and annotation of orthologous groups of genes, Nucleic Acids Res. 36 (2007) D250–D254.
- [59] E.V. Koonin, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, D.M. Krylov, K.S. Makarova, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, I.B. Rogozin, S. Smirnov, A.V. Sorokin, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes, Genome Biol. 5 (2004) R7.
- [60] E. Dimmer, R. Huntley, Y. Alam-Faruque, T. Sawford, C. O'Donovan, M. Martin, B. Bely, P. Browne, W. Chan, R. Eberhardt, M. Gardner, K. Laiho, D. Legge, M. Magrane, K. Pichler, D. Poggioli, H. Sehra, A. Auchincloss, K. Axelsen, R. Apweiler, The UniProt-GO annotation database in 2011, Nucleic Acids Res. 40 (2011) D565–D570.
- [61] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 27 (1999) 29–34.
- [62] Y. Wang, H. Tang, J.D. Debarry, X. Tan, J. Li, X. Wang, T.H. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (2012) e49.

- [63] C. Chen, H. Chen, Y. Zhang, H.R. Thomas, M.H. Frank, Y. He, R. Xia, TBtools: an integrative toolkit developed for interactive analyses of big biological data, Mol. Plant 13 (2020) 1194–1202.
- [64] K. Katoh, K. Misawa, K.i. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066.
- [65] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.
- [66] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589.
- [67] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274.
- [68] M. Hao, L. Zhang, S. Ning, L. Huang, Z. Yuan, B. Wu, Z. Yan, S. Dai, B. Jiang, Y. Zheng, D. Liu, The resurgence of introgression breeding, as exemplified in wheat improvement, Front. Plant Sci. 11 (2020) 252.
- [69] J. Dolezel, M. Kubaláková, E. Paux, J. Bartos, C. Feuillet, Chromosome-based genomics in the cereals, Chromosome Res. 15 (2007) 51–66.
- [70] J. Doležel, J. Vrána, P. Cápal, M. Kubaláková, V. Burešová, H. Simková, Advances in plant chromosome genomics, Biotechnol. Adv. 32 (2014) 122–136.
- [71] S.I. Tsuruta, M. Ebina, M. Kobayashi, W. Takahashi, Complete chloroplast genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: comparative genomics and evolution of the *Saccharum* complex, PLoS ONE 12 (2017) e0169992.
- [72] J. Vrána, P. Cápal, J. Číhalíková, M. Kubaláková, J. Doležel, Flow Sorting Plant Chromosomes, Springer, New York, NY, USA, 2016.
- [73] D. Kopecký, M. Martis, J. Číhalíková, E. Hřibová, J. Vrána, J. Bartoš, J. Kopecká, F. Cattonaro, Š. Stočes, P. Novák, P. Neumann, J. Macas, H. Šimková, B. Studer, T. Asp, J.H. Baird, P. Navrátil, M. Karafiátová, M. Kubaláková, J. Šafář, K. Mayer, J. Doležel, Flow sorting and sequencing *Meadow fescue* chromosome 4F, Plant Physiol. 163 (2013) 1323.
- [74] B.A. Akpinar, S.J. Lucas, J. Vrána, J. Doležel, H. Budak, Sequencing chromosome 5D of Aegilops tauschii and comparison with its allopolyploid descendant bread wheat (*Triticum aestivum*), Plant Biotechnol. J. 13 (2015) 740–752.
- [75] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D. G. Peterson, R. Mehboob ur, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, D.S. Rokhsar, The Sorghum bicolor genome and the diversification of grasses, Nature 457 (2009) 551–556.
- [76] S.I. Wright, I.V. Bi, S.G. Schroeder, M. Yamasaki, J.F. Doebley, M.D. McMullen, B. S. Gaut, The effects of artificial selection on the maize genome, Science 308 (2005) 1310–1314.
- [77] S.A. Goff, D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchison, C. Martin, F. Katagiri, B. M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W.L. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant, S. Briggs, A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), Science 296 (2002) 92–100.
- [78] E.A. Cooper, Z.W. Brenton, B.S. Flinn, J. Jenkins, S.Q. Shu, D. Flowers, F. Luo, Y.S. Wang, P. Xia, K. Barry, C. Daum, A. Lipzen, Y. Yoshinaga, J. Schmutz, C. Saski, W. Vermerris, S. Kresovich, A new reference genome for *Sorghum bicolor* reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism, BMC Genomics 20 (2019) 13.