

Advances in Sugarcane Genomics and Genetics

Tianyou Wang¹ · Jingping Fang² · Jisen Zhang¹ 

Received: 11 January 2021 / Accepted: 1 October 2021 / Published online: 21 October 2021
© The Author(s), under exclusive licence to Society for Sugar Research & Promotion 2021

Abstract Currently, genome analysis has become a routine component of molecular breeding of major crops. Today's commercial sugarcane hybrids are the products of breeding over one hundred years with the starting clones produced by crossing two founding species (*Saccharum officinarum* and *S. spontaneum*) in India and Indonesia. Current sugarcane varieties have a highly complex and large genome with 100–130 chromosomes. Despite the complexity and size of the genome, considerable progress has recently been made in sugarcane genomics, including the sequencing of a haploid *S. spontaneum*, AP85-441 and sugarcane cultivar hybrids R570 and SP80-3280. AP85-441 genome is assembled to chromosome level and allele-defined. Significant progress also has been made in genetic research of important agronomic traits. Here, we outline these advances in order to provide a reference for future sugarcane genomics and genetics research.

Keywords *Saccharum* · *Saccharum officinarum* · *Saccharum spontaneum* · Polyploidy · Genome evolution

Introduction

The successful sequencing of the whole genome sequence of the single-stranded DNA phage ϕ X174 marked the formal entry of human beings into the era of genomics (Sanger et al. 1977). The earliest model crop genomes were assembled with Sanger sequencing, such as rice (International Rice Genome Sequencing 2005), maize (Schnable et al. 2009), sorghum (Paterson et al. 2009), and grape (Jaillon et al. 2007). The first-generation sequencing technology has high accuracy and the reading length of the sequence can reach 1 KB, but its sequencing program is complex, high cost and low throughput. The rise of second-generation sequencing technologies (also known as high-throughput sequencing) represented by Roche 454, Illumina Solexa/HiSeq and ABI SOLiD technologies provide the foundation for large-scale genomics research (Hamilton and Robin Buell 2012). Although the second-generation sequencing read length is shorter, it has the advantages of high throughput, high speed, and low cost (Metzker 2010). It has greatly promoted the application and development of genome and transcriptome sequencing, and has become the main method of large-scale whole genome sequencing technology (Chan et al. 2010; Velasco et al. 2010). While this approach has generated more complete and contiguous assemblies of low-copy genic regions, the more repetitive, TE-rich regions of the genome have proven to be difficult to assemble with short reads, resulting in many gaps (due to the complex structure in the genome, the measured sequence cannot completely cover the whole genome. The middle missing sequence is called gap), resulting in partial assembly in these regions (Hert et al. 2008; Schatz et al. 2010; Alkan et al. 2011). In recent years, with the progress of third-generation sequencing technology, characterized by single molecule sequencing, genome sequencing

✉ Jisen Zhang
zjisen@126.com

¹ Guangxi Key Laboratory of Sugarcane Biology & Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, People's Republic of China

² College of Life Science, Center of Engineering Technology Research for Microalgae Germplasm Improvement of Fujian, Southern Institute of Oceanography, Fujian Normal University, Fuzhou 350117, Fujian, People's Republic of China

gradually matured and produced more continuous and complete assembly of some crop genomes (Zhang et al. 2016a; Jiao et al. 2017; Liu et al. 2020; VanBuren et al. 2020). It has been shown that the third generation of long sequencing reads are invaluable for obtaining high-quality assembly, as they span proportionally more of the repeats present in a genome (van Dijk et al. 2018; Della Coletta et al. 2021). With the ongoing advancement of sequencing technologies and the genome assembly algorithm, it is now possible to obtain the sequence information of many plant genomes (even the complex chromosomes of polyploid sugarcane) and important functional genes, facilitating studies on plant molecular evolution, gene composition and gene regulation at the molecular level, and genome editing tools to modify a target sequence to modify its expression (Zhang et al. 2018; Souza et al. 2019; Della Coletta et al. 2021).

Sugarcane is an important large perennial crop in tropical and subtropical regions, contributing 80% of the world's total sugar production. This C_4 species is a high efficiency converter of light energy into chemical energy (Hatch et al. 1967; Hatch and Slack 1998; Waclawovsky et al. 2010). Recent concerns about global warming and the search for alternatives to fossil fuels have focused global attention on sugarcane, which is a candidate crop for bioenergy and biomaterial production, providing 60% of biofuel globally (Yang et al. 2020a).

Despite the economic importance of sugarcane, research on the sugarcane genome has lagged behind due to its complex genome. In recent years, the development of new biotechnological methods provides opportunities to study the sugarcane genome and good progress has been made using genome sequencing strategies. The interspecific hybrid genome of sugarcane possesses genetic materials inherited from the two parental species unevenly, which makes the genome more complex than that of its progenitors (D'Hont et al. 1996). The basic chromosome number (also called monoploid number: the number of different chromosomes that make up a single complete set) is 10 in the sugarcane hybrid, and the total number of chromosomes in sugarcane varies from 100 to 130 (D'Hont et al. 1998; Zhang et al. 2014). It is estimated that there are 8–14 homoeologous copies of a given gene at a given locus in the sugarcane genome (Souza et al. 2011). However, the ratio of inherited chromosomes in hybrids may not be consistent, and each cross may result in a new genetic composition. Due to its high ploidy levels and distinct genetic compositions, creating a reference genome or a pan-genome (describes all genes and genetic variation within a species, including core genome that is the portion of the pan-genome common to all individuals in the species and dispensable genome that is the portion of the pan-genome that is only in a subset or unique to individuals.) in

sugarcane would be a real challenge. In addition, it has been found that sugarcane genomes have large proportions of highly repetitive DNAs and segmental duplications or whole genome duplication (WGD), which can cause problems in their genome assembly (Thirugnanasambandam et al. 2018).

In spite of the complexity, with the advance of sequencing technology and bioinformatic tools, great progress has been made in sugarcane genomics, such as sugarcane hybrid bacterial artificial chromosome (BAC) libraries of reference genomes and a high-quality reference genome of *S. spontaneum*. Now, researchers can easily obtain the sequence information of many important functional genes or regulatory regions and sequence. This provides a standard for sequence alignment and makes it possible to develop large-scale molecular markers.

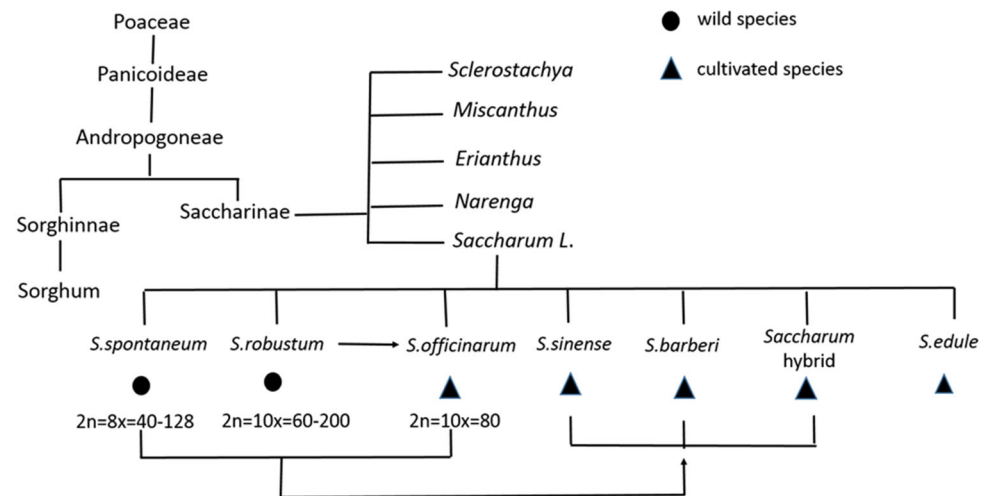
In recent years, sugarcane genomics and genetics have made great progress, including phylogenetics of *Saccharum*, evolution of genome size, basic chromosome number and polyploidization, sugarcane reference genome, centromere composition and genomic basis of important biological characteristics. Here, we discuss these new developments.

Origin and Classification of *Saccharum*

Sugarcane (*Saccharum L.*) refers to several species and hybrids of tall perennial grasses in the genus *Saccharum*, which are mainly used for sugar and biofuel production. It is native to the warm temperate of tropical regions of India, Southeast Asia and New Guinea (Daniels 1993). It belongs to the grass family Poaceae, sub-family Panicoideae, tribe Andropogoneae, and clade Saccharinae (Soreng et al. 2015). *Saccharum*, together with the genera *Narenga*, *Erianthus*, *Sclerostachya* and *Miscanthus*, form the “*Saccharum Complex*” (Irvine 1999; Amalraj and Balasundaram 2006), representing a significant reservoir of genetic variability that could be exploited and used in breeding (Fig. 1). Modern sugarcane cultivars are interspecific hybrids derived from crosses of *S. officinarum* × *S. spontaneum*, with repeated backcrossing to *S. officinarum* (Amalraj and Balasundaram 2006).

There are two theories for centers of domestication for sugarcane, one is the domestication of *Saccharum officinarum* by Papuans in New Guinea (Riaño-Pachón and Mattiello 2017), while the other is the domestication of *Saccharum sinense* by Austronesians in Taiwan and southern China (Daniels 1993). *Saccharum officinarum*, also known as ‘noble cane,’ is characterized by thick barrel-shaped internodes and is tall with juicy stalks which has a thin rind, high sugar content and low fiber content. It was thought to be evolved from wild species *Saccharum*

Fig. 1 Taxonomic position of *Saccharum* and related genera (only representative species commonly used in breeding are shown)



robustum as they share a common center of origin (New Guinea) and have the same basic chromosome number $x = 10$ (D'Hont et al. 1998). This hypothesis was confirmed by SSR markers that revealed *Saccharum officinarum* has the closest relationship to *Saccharum robustum* (Brown et al. 2007). *S. spontaneum* is a grass native to the Indian subcontinent and widely distributed in both tropical and subtropical regions, and is one of the important progenitors of modern cultivated sugarcane (Mukherjee 1957; Hodkinson et al. 2002; Yang et al. 2019).

Genome Sizes of *Saccharum* Species

Sugarcane genome size (nuclear DNA amounts) varies greatly due to varying ploidy levels in different *Saccharum* species. It is challenging to study the complex sugarcane genome due to the fact that there are 8 to 12 alleles in the euploid and aneuploid sugarcane genomes. The genome size of *S. officinarum* accessions ranged from 7.50 to 8.55 Gb with an average size of 7.88 Gb, and that of *S. robustum* ranged from 7.65 to 11.78 Gb (Zhang et al. 2012). The genome size of *S. spontaneum* varies widely, and it is reported to be 3.36 to 12.64 Gb (da Silva et al. 1995; Ha et al. 1999; Zhang et al. 2012). It was estimated that the average monoploid genome size of *S. officinarum* and *S. spontaneum* were 985 Mb and 843 Mb, respectively (Zhang et al. 2012). The haplotype size (~ 980 Mb) of sugarcane cultivar R570 is similar to the size of sorghum genome (~ 730 Mb) (Zhang et al. 2014). Genome size is of great significance for the study of plant taxonomy, evolutionary biology and ecological genetics (Greilhuber et al. 2005; Pellicer et al. 2018). Under a certain sequencing depth, genome size influences the amount of data required to provide a reference standard for future genome assembly and estimation of genome integrity (Dominguez Del Angel et al. 2018).

Synteny Between Sugarcane and Closely Related Species

In recent years, our knowledge on the evolutionary origin and genome structure of sugarcane has increased considerably. The sugarcane genome is highly collinear with that of some other grasses, such as sorghum, rice, and corn. Sorghum with a genome size of 730 Mb appears to be an excellent diploid organism for comparative genomics of sugarcane due to its limited divergence time (About 4.0–6.4 million years ago) (Zhang et al. 2021a, 2021b). DArT markers, SNP and EST-SSR markers were used to generate genetic maps for comparison with the sorghum genome to reveal the complex genome structure of sugarcane. Comparative mapping revealed that there is a good collinearity between sugarcane and sorghum in four of the eight homology groups (HGs) and four major chromosomal recombination events occurred between sugarcane and sorghum (Aitken et al. 2014a). Among them, two were condensations of chromosomes, which may account for the reduction of the basic chromosome number in sugarcane from $x = 10$ to $x = 8$, and the assembled AP85-441 genome verified this inference (discussed later). Twenty BACs from the sugarcane R570 hybrid, each corresponding to a sorghum chromosome arm, were sequenced to study the genome structure and organization (Wang et al. 2010). It was found that there is a high collinearity between the two species, and their coding regions have an average consensus sequence identity as high as 95.2%. Previously, the complete sorghum genome had been sequenced and yielded a well-annotated genome (the genome that identified the locations of genes and all of the coding regions in a genome and determining what those genes do), which laid an important foundation for the study of complex sugarcane genomes (Okura et al. 2016; Deschamps et al. 2018). Microcollinearity between sorghum and sugarcane could be

used to guide the assembly of the gene-rich region of a monoploid sugarcane genome. For instance, sugarcane hybrid R570 mosaic monoploid reference sequence was successfully assembled with sorghum genome as a reference to identify a minimum tiling path (MTP) of BACs covering the euchromatin of a monoploid sugarcane genome (Garsmeur et al. 2018).

Modern *Saccharum* Hybrid (Cultivated Sugarcane) Genome

Modern cultivated sugarcane is an aneuploid plant derived from *S. officinarum* and *S. spontaneum* through multiple interspecific crosses. Inter-specific hybridization provided a breakthrough in sugarcane breeding. It improved abiotic stress tolerance and disease resistance, and greatly helped to increase the yield. At one time, it was believed that there was no chromosomal recombination between the genomes (Price 1963). However, it was proved later that about 70–80% of the chromosomes of modern sugarcane cultivars were derived from *S. officinarum*, 10–23% from *S. spontaneum*, and 5%–17% came from the recombination of *S. officinarum* and *S. spontaneum* (D’Hont et al. 1996; Tu et al. 2009; Aitken et al. 2014b). RFLP and AFLP molecular markers were used to analyze the molecular diversity of cultivated sugarcane clones with more than 80 chromosomes, which showed that *S. officinarum* clones were modified through introgression with other members of the ‘*Saccharum complex*’ (Jannoo et al. 1999b; Aitken et al. 2006). In modern cultivars, all the *S. spontaneum* chromosomes came from *S. spontaneum* cytotypes with $x = 8$ and the vast majority of the chromosomes result from interspecific exchanges between homeologous chromosomes (Piperidis and D’Hont 2020). The development of technologies such as DNA molecular markers, genomic in situ hybridization (GISH) and fluorescence in situ hybridization (FISH) makes it possible to study the complex structure of the sugarcane genome.

Currently, two *Saccharum* hybrid reference genome sequences are available, one is the sugarcane hybrid R570 monoploid genome (Garsmeur et al. 2018), and the other is the unscaffolded genome (composed of contigs that is a contiguous length of genomic sequence in which the order of bases is known to a high confidence level) of Brazilian cultivar SP80-3280 (Souza et al. 2019). A BAC-based monoploid genome sequence of sugarcane cultivar R570 was obtained by exploiting the high level of collinearity with sorghum. A minimum tiling path (MTP) of 4660 sugarcane BAC clones that best covers the gene-rich part of the sorghum genome was selected for sequencing (Riaño-Pachón and Mattiello 2017; Garsmeur et al. 2018). Then, the MTP was trimmed to construct a 382-Mb single

tiling path (STP), which is a single copy in the sugarcane gene space (Gene space is the extended space including and around genic regions, encompassing coding as well as intergenic noncoding regions bounded by the farthest upstream and downstream conserved noncoding sequences associated with a gene), and a total of 25,316 gene models were predicted. This is the first assembly of the monoploid reference sequence, which provides an indispensable reference for aligning sequencing data, and offers an essential framework to help the whole genome sequencing projects to overcome difficulties in assembling such a complex genome.

The gene space of SP80-3280 has also been assembled with a total assembly size of 4.2 Gb, including 373,869 putative genes and their potential regulatory regions (Souza et al. 2019). This is the first release of an assembly of such a huge hybrid polyploid genome with part of putatively identified homeologs. This work, including a large collection of gene space homeolog diversity, is a fundamental step toward the whole genome assembly of a commercial sugarcane cultivar. These fragmented genomes are starting to bridge the long-standing knowledge gap in understanding the genome structure of sugarcane, providing some useful genomic tools that will be valuable to biotechnology and other molecular research in sugarcane, other crops and plants in general.

In recent years, scientists from Brazil, France, Australia, Thailand, China, and the USA have all tried to complete the genome sequencing of sugarcane cultivars, but progress is slow compared to other crops, which is attributed to the large size and its inherent genetic complexity of the sugarcane genome. A summary of currently available assembled genomic resources for sugarcane research as of October 2020 is presented (Table 1). With the advancement of genomic technologies and new methodologies, it is hoped that complete comparative and structural analyses among sugarcane species and cultivars become a reality in the near future.

The Genome of *S. spontaneum*

S. spontaneum is the only genome that has been assembled at the chromosome and allele-defined levels in *Saccharum*. AP85-441, a haploid *S. spontaneum* species ($1n = 4x = 32$), was sequenced and 32 pseudo-chromosomes were assembled, including 8 homologous groups of 4 members each (Zhang et al. 2018). The AP85-441 contig-level assembly integrated a variety of sequencing data, including BAC pools sequenced with Illumina HiSeq 2500 and whole genome shotgun sequencing with PacBio RS II, Hi-C reads and Illumina short reads. Each BAC pool was independently assembled using SPAdes, SOAPdenovo2

Table 1 Currently available assembled genomic resources of sugarcane

Global statistics	Cultivar SP80-3280 (gene space assembly)	Cultivar R570 (single tiling path)	<i>S. spontaneum</i> AP85-441 (haploid)
Total assembly size (Gb)	4.26	0.38	3.13
Number of contigs	450,609	5896	91,867
Contig N50/mean length (bp)	13,157	102,858 (mean length)	45,032
Predicted genes (alleles)	NA (373,869)	22,780 (NA)	35,525 (82,773)
Anchored contigs	NA	NA	76,131
Unanchored contigs	NA	NA	15,704
Gaps	NA	NA	76,009

Predicted genes (alleles), numbers outside parentheses means the total predicted genes only one allele was retained, while numbers in parentheses means total predicted genes including all alleles

NA, not applicable

and ALLPATHS-LG, and the PacBio RS II reads assembled by Canu to reduce fragmentation, yielding 3.13 Gb sequences with contig N50 of 45 kb. Chromosomal assembly was constructed using ALLHIC, which is designed for polyploid genome scaffolding. After two rounds of MAKER annotation (Genome annotation is the process of identifying functional elements along the DNA sequence of a genome, and it often focuses on genes) followed by manual annotation, 35,525 genes with alleles defined were annotated. This allele-defined genome represents a significant milestone for sugarcane biotechnology research, and it also provides a new perspective for solving many unresolved biological problems in sugarcane.

Centromere Composition of *Saccharum*

The centromere is composed of DNA and proteins, which is the necessary chromosome region for the precise separation and transmission of chromosomes in eukaryotes during mitosis and meiosis. Although the functions of centromere are very conservative, the centromeric DNA sequences of different species are highly variable (Talbert et al. 2002). In contrast, centromeric proteins are relatively conserved among species, and centromere-specific histone H3 (CENH3) plays a key role in kinetochore formation and centromere function. A repetitive sequence consisting of 140-bp repetitive units from sugarcane genomic DNA was cloned and designated as the centromeric tandem repeats (SCEN) family (Nagaki et al. 1998). In situ hybridization revealed that they were located on the centromeric region of almost all of the chromosomes and the amino acid sequence of CENH3 in sugarcane was very similar to that in rice and maize (Nagaki and Murata 2005). Therefore, rice CENH3 antibody was used for immunoprecipitation with CENH3 of sugarcane and sugarcane centromeric

retrotransposons (CRS) precipitated significantly, which means that CRS directly interacts with CENH3 in sugarcane centromeres.

However, the polyploidy, heterogeneity and aneuploidy of sugarcane genome as well as the widespread repetitive sequences, especially tandem repeats and retrotransposon repeats on centromere, have hindered centromere genome assembly in sugarcane genome sequencing program. Moreover, the gap between overlapping groups cannot be filled by automatic assembly. The centromeres mainly contained SCEN-like single satellite repeat (Ss1) and several Ty3/gypsy retrotransposon-related repeats (Ss166, Ss51, and Ss68). Ss1 exhibits a chromosome-specific enrichment in *S. spontaneum* and *S. robustum*, and it is dominant in the centromeric regions and spans up to 500 kb. In contrast, the Ty3/gypsy retrotransposon-related repetitive sequences are either clustered spanning over a short range, or scattered in the centromere regions (Zhang et al. 2017). The development of new cytological techniques and molecular biology will help to overcome the shortcomings in sugarcane molecular and cytology research on sugarcane centromere composition and accelerates the complete assembly of chromosome sequence.

The Evolution of *Saccharum*

Evolution of Basic Chromosomes

Clade Saccharinae (sugarcane, sorghum, miscanthus, and other related C4 species) includes a remarkable array of recently and independently derived polyploids that arose from a common diploid progenitor (Swaminathan et al. 2012). Sugarcane carries even multiples of a haploid complement of $x = 10$ or $x = 8$ chromosomes, while *Miscanthus* had a basic chromosome number of $n = 19$ ($2n = 38$) (Swaminathan et al. 2012; Mitros et al. 2020). It

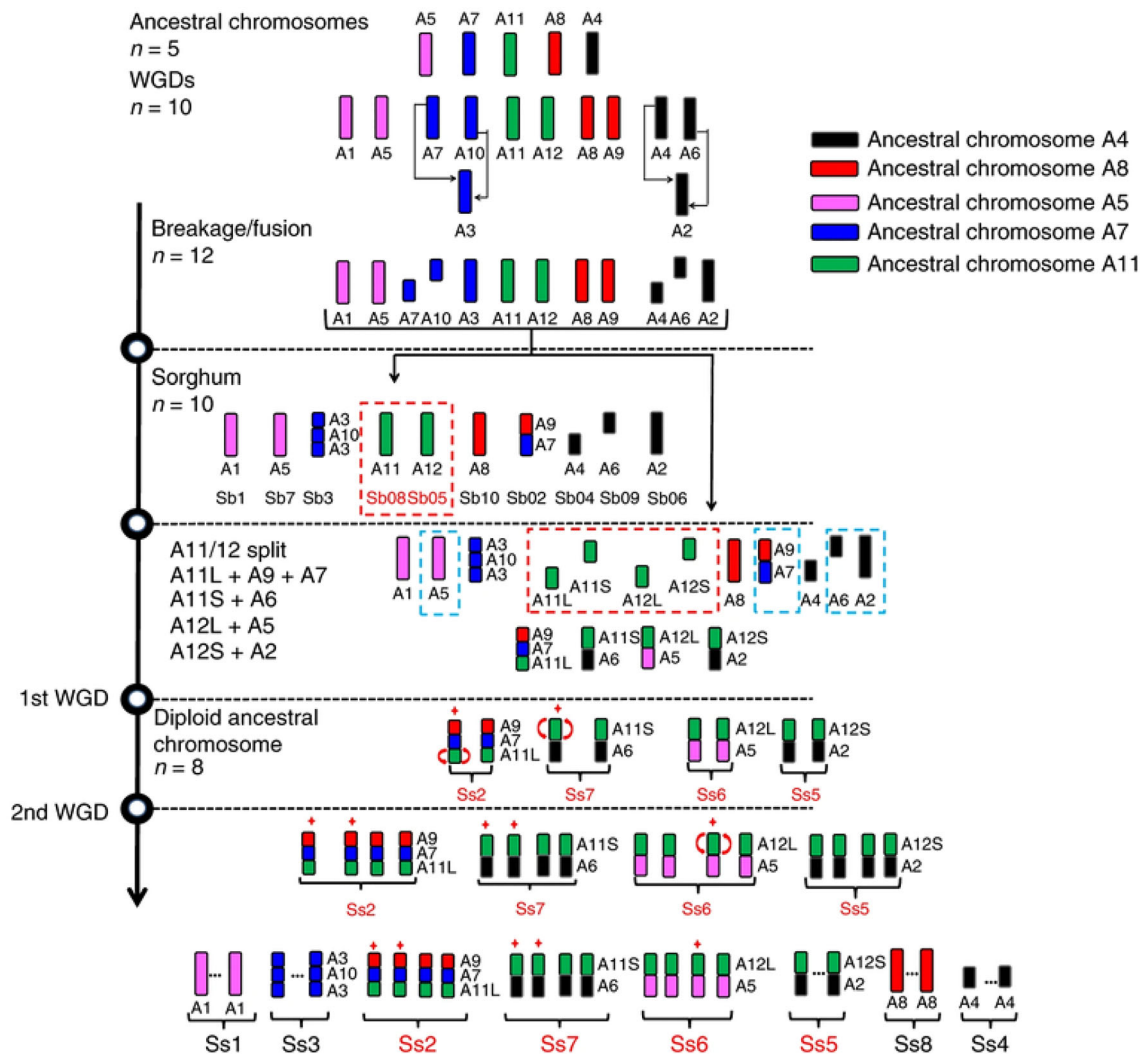


Fig. 2 Evolutionary history of *S. spontaneum* chromosomes (Zhang et al. 2018). The evolution of chromosome numbers in Poaceae, from $n = 10$ in sorghum to $n = 8$ in *S. spontaneum*. Chromosomes are represented by color codes to illustrate the evolution of segments from a common ancestor with 5 chromosomes. The genomes of ancestors are marked with AX (X is a number from 1 to 12). The

rearranged chromosomes of *S. spontaneum* are marked with dashed boxes. Red curved arrows are used to show the inverted events that occur in chromosome segments, and red plus signs are used to mark the chromosomes recombined with the inverted chromosome segments

is the polyploidy and recombination events and the parallel evolution that resulted in the varying basic chromosome numbers of sugarcane (Ha et al. 1999; Meng et al. 2020).

SNP-based genetic map of cultivar R570 also showed that the distinct basic chromosome number in *S. spontaneum* with $x = 8$, compared to *S. officinarum* with $x = 10$, resulted from a few large chromosomal rearrangements, i.e., two pairs of three chromosomes were rearranged into two chromosomes (Garsmeur et al. 2018). The AP85-441 genome assembled at chromosome level confirmed that a pair of ancient duplicated chromosomes that have undergone frequent recombination resulted in the reduction of *S. spontaneum* chromosomes from 10 to 8 (Zhang et al.

2018). Alignment to sorghum showed the chromosome fissions in ancestral homologs of sorghum chromosomes 5 and 8 (ancient duplicated chromosome pairs A5 and A11 in grasses) (Fig. 2). A12 and A11 are a paleo-duplicated pair of chromosomes which originated from p whole genome duplication in grass. Based on the comparative genomics, the ancestor of SbChro5 was originated from A12 and A11, respectively (Salse et al. 2008). In *S. spontaneum*, the common ancestor of SbChro5(A12) split into two large segments, C5S (A12S) and C5L (A12L), and translocated into the ancestral chromosomes of SbChr6 (A2) and SbChr7 (A5), respectively. Similarly, the ancestor of SbChr8 (A11) was divided into two segments, C8S (A11S)

and C8L (A11L), and translocated into the ancestors of SbChr09 (A6) and SbChr02 (A7 + A9), respectively. Sorghum stratum SSA formed 13.4 million years ago, well before sorghum and Saccharum diverged, so the homologous short fragments between SbChr08 and SsChr5, and between SbChr05 and SsChr7, are remains of homeologous genes in sorghum stratum SSA (Wang et al. 2011; Zhang et al. 2018).

Recently, a set of oligonucleotide (oligo) probes (a short sequence of nucleotides that are synthesized as a molecular probe to match a specific region of DNA or RNA to detect the sequence of those regions) based on the *S. spontaneum* genome ($x = 8$) was developed to examine a *S. spontaneum* accession Np-X ($2n = 4x = 40$), which belonged to the ancient Pan-Malaysia group (Meng et al. 2020). The results showed that it was a tetraploid with a basic chromosome number of $x = 10$ and had a global chromosome architecture conserved with sorghum and *S. officinarum*. This finding indicates a parallel evolution path of genomes and polyploid series in *S. spontaneum* with different basic chromosome numbers. Chromosome specific oligo probes based on the sequence assembly of R570 were used to analyze the genome architecture of some modern cultivars and their parental species (Piperidis and D'Hont 2020). The results validated the basic chromosome number of $x = 10$ for *S. officinarum*, and in *S. spontaneum* rearrangements occurred from a basic chromosome of $x = 10$, in two steps leading to $x = 9$ and then $x = 8$. The rearrangements that led to $x = 8$ followed by further polyploidization may have resulted in an increased adaptability of the species.

Polyploidization in Sugarcane

Whole genome duplication (WGD) has played an important role in plant evolution and diversification. Many crops are relatively recent autopolyploids or allopolyploids. WGD events have had a profound impact on the evolution of most grass (Poaceae) genomes. High polyploidization seems to provide increased adaptability to harsher environments and promotes the expansion of these species (Comai 2005). Therefore, a full understanding of the evolutionary consequences of autopolyploidy is critical to our understanding of crop domestication and agricultural improvement (Paterson et al. 2012; Renny-Byfield and Wendel 2014). Analysis of shared transposable elements (TE) insertions in single-copy genes suggested that two autopolyploidization events may have occurred in the lineage that gave rise to *S. officinarum*, after its divergence from *S. spontaneum* (Vilela et al. 2017). *S. officinarum* and *S. robustum* had diverged about 385,000 years ago, and the WGD events are thought to have occurred after the speciation event due to the shared interchromosomal rearrangements (Zhang et al. 2019). Two fissions in ancestral

homologs of sorghum chromosomes SbChr05 and SbChr08 occurred before the two rounds of WGDs in *Saccharum*, but after the divergence of *Saccharum* and *Miscanthus* (Fig. 3). Comparative genomics shows that the two rounds of WGD are autopolyploidization, and the time between them is very short (Garsmeur et al. 2018; Zhang et al. 2018).

Population Genomics of Saccharum

Cultivating sugarcane varieties with excellent agronomic traits is an efficient way to meet the growing demand for sugar and bioenergy. The improvement of sugarcane varieties has always been extremely challenging, mainly due to its complex genome with high polyploidy levels. Therefore, it is critically important to assemble a large germplasm panel, survey important agronomic traits, and analyze the loci that contribute to agronomic traits. The recent advancements in molecular markers, sequencing tools and development of robust statistical models have the potential to deliver affordable genomics-assisted breeding in sugarcane. Genome-wide association studies is a powerful tool for the investigation of complex traits, and identification of hundreds of thousands of SNPs is affordable through genome-wide sequencing of a selected panel of diverse genotypes (Tibbs Cortes et al. 2021). Genome-wide association study in sugarcane germplasms showed that certain markers were significantly associated with agronomic traits (Racedo et al. 2016; Yang et al. 2017a, 2019; Barreto et al. 2019; Fickett et al. 2019; Zan et al. 2020). However, the complexity of genome and the genetic mechanism regulating yield and quality traits increase the difficulty of molecular marker-assisted breeding in sugarcane. Statistical model and genomic tools that are widely used in population genetics are suitable for diploids. With the release of genome sequence of *S. spontaneum* and sugarcane hybrids, the possibilities of detecting novel variations linked to desirable traits have increased. It is of great significance to develop suitable statistical model and software algorithms for sugarcane to identify effective marker–trait associations (Banerjee et al. 2020).

Genomic Basis of Important Biological Characteristics

Sugarcane is a C4 plant with high photosynthetic efficiency to produce and store sucrose at unusually high concentrations. With the increasing demand for sugar and energy, significant efforts have been made to increase cane and sugar yield. Sugar accumulation is a complex process regulated by a variety of sugar metabolizing enzymes and

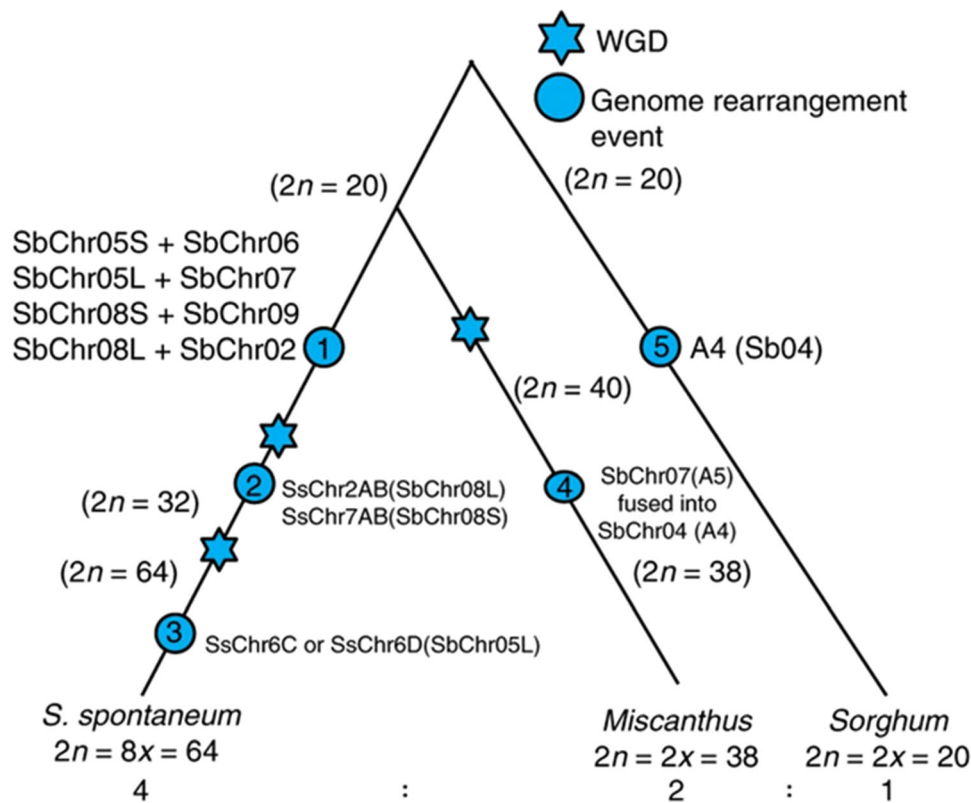


Fig. 3 Recent polyploidization events and genome rearrangements (Zhang et al. 2018). The genome duplications and rearrangements in *S. spontaneum*, *Miscanthus*, and sorghum are shown (4:2:1 indicates ratio of polyploidy level after WGD events, regardless of chromosome fissions). 1, The basic chromosome number reduction from 10 to 8 in *S. spontaneum* discussed in Evolution of Basic chromosomes Sect. 2. 2, Two inversions occurred after the first round of WGD are shown by pairs of inversions in SsChr2AB and SsChr7AB. 3, Three

segments in SsChr6ABD are in inverted position. 4, The ancestral chromosome SbChr07 (A5) fused with SbChr04 (A4) after an allopolyploidization event in *Miscanthus*. 5, Four chromosomal segments in SsChr4ABCD are in inverted position. This is a peculiar inversion of sorghum in SbChr04 (A4), because the orientation of this chromosomal fragment is the same in rice, *Miscanthus*, and *S. spontaneum*

various physiological processes, including sucrose synthesis, decomposition and transportation. The key factors regulating sugar accumulation are the activity of enzymes related to sugar synthesis and transport capacity, efficiency of phloem transport of sucrose as well as sucrose transport in the stem parenchyma cells. To ensure sugar yield, it is also necessary to increase disease and abiotic stress resistance of sugarcane. Here, we briefly summarize recent research on sugar accumulation and stress resistance of sugarcane.

Sugar Accumulation

Sucrose phosphate synthase (SPS)

SPS genes play vital roles in the production of sucrose in various plants. SPS is at the interface between sink and source, which is positively correlated with sugar content and may regulate the supply of sucrose to the sink. In *Saccharum*, SPS sequences were identified from the BAC

libraries of *S. officinarum* and *S. spontaneum*, respectively (Ma et al. 2020). All the SPS genes showed a trend of increased expression along the developmental gradient of leaves. *SPSA* and *SPSB* genes presented high expression and differential expression patterns between the two founding *Saccharum* species, suggesting that these two *SPSs* are important in the regulation of sucrose accumulation (Verma et al. 2011; Ma et al. 2020). Recent studies showed that SPS overexpression can increase the SPS activity and sucrose content of transgenic sugarcane leaves, and some transgenic lines increased the plant height and stem number (Anur et al. 2020). These studies strongly indicate that increasing SPS activity is an effective strategy to increase sugarcane yield.

Invertases (INVs)

Invertases (INVs) are involved in the process of cell differentiation, plant development and responses to environmental stresses, and they are key enzymes regulating

sucrose metabolism in plants (Ansari et al. 2013; Cardozo and Sentelhas 2013). Studies showed that in sugarcane culm tissues with low sucrose content and high hexose content, neutral invertase transcripts and protein levels are higher than those with stored sucrose (Bosch et al. 2004). Sequences belonging to 14 non-redundant members of the INV gene family were cloned, and the expression levels of INV genes and the variation of enzyme activities were analyzed in sugarcane seedlings subjected to drought (10% (W/V) PEG6000, 16-h light at 28 °C/8-h dark at 24 °C), low temperature (16 h light at 15 °C/8 h dark at 10 °C), glucose, fructose, and sucrose (3% (W/W) sucrose, 3% (W/W) glucose, 3% (W/W) fructose, respectively) stress. (Wang et al. 2017). Results showed that the activities of all INVs were significantly inhibited in response to all those five abiotic stresses, and the neutral/alkaline INVs activity decreased less in these abiotic stresses than the acid INVs and cell-wall INVs. These studies provide a basis and framework for understanding the physiological effects of INVs on sucrose accumulation and abiotic stress tolerance in sugarcane.

Sucrose Synthase (SUS) Gene Family

SUS is widely recognized as a key enzyme in sugar metabolism, primarily acting in sink tissues (Haigler et al. 2001; Baroja-Fernández et al. 2012). Haplotypes of five SUS genes in *S. officinarum*, *S. robustum*, and *S. spontaneum* were identified and phylogenetic analysis showed that SUSs in *Saccharum* evolved before the divergence of the genera in the tribe Andropogoneae at least 12 million years ago (Zhang et al. 2013). Expression profiles of SUS genes in the top, middle and bottom internodes of *S. spontaneum* and *S. officinarum* culm indicated that all SUS transcripts were differentially expressed between the top and bottom tissues, with high expression in the top tissues, lower expression in the bottom and moderate expression in the middle (Thirugnanasambandam et al. 2019). RNA-seq analysis of two *Saccharum* species revealed that SUS genes were at higher expression levels in *S. officinarum* than in *S. spontaneum*, and *SUS1* is mainly expressed in the internodes and basal areas of leaves; *SUS2* presented different expression patterns during the circadian rhythm in *S. spontaneum* and *S. officinarum* (Shi et al. 2019). These studies indicate that SUS genes may contribute to sugar content variation in these species and the functions of SUS genes in sugarcane, however, need to be studied further.

Fructokinase Gene Family

Fructokinases (FRKs) have high substrate specificity and affinity, which are the main fructose phosphorylating enzymes (Renz and Stitt 1993). As fructose accounts for

half of the hexose generated by sucrose cleavage in sink tissues, FRKs are considered to be essential for the metabolic pathways in sink tissues. *FRK2* is inhibited by fructose concentrations exceeding 0.1 mM, while *FRK1* activity is not negatively affected even at 1.0 mM fructose (Hoepfner and Botha 2003). By combining comparative genomics approaches, seven fructokinase genes were identified in *S. spontaneum* and the genomic constraints within the FRKs alleles between *S. spontaneum* and *Sorghum* were analyzed under drought stress and exogenous application of plant hormones (Chen et al. 2017). It is suggested that *FRK1* is conserved among allelic haplotypes and plays a major role in the phosphorylation of fructose; *FRK3* and *FRK5* are induced in response to drought stress; *FRK7* may be functionally redundant in *Saccharum* (Chen et al. 2017). The FRK gene family in sugarcane is not studied extensively.

Sugar Transporters

Sugar transporter (ST) is considered to be the most important gene family for sugar accumulation. The Sugar Will Eventually be Exported Transporters (SWEET) family plays an indispensable role in sugar efflux, phloem loading, and reproductive tissue development. *SWEETs* express differently in different tissues and regions, which may result in differences in sugar content between *S. spontaneum* and *S. officinarum* (Hu et al. 2018). Sucrose transporters (SUTs) are considered to be the key genes in regulating sucrose storage and are critical for phloem loading in source tissue and the sucrose uptake in sink tissue. Transcriptome data from seedlings under drought stress and mature plants of three *Saccharum* species (*S. officinarum*, *S. spontaneum* and *S. robustum*) showed that *SUT1* and *SUT4* expressed abundantly at different conditions (Zhang et al. 2016b). Phylogenetic analyses and expression profile indicate that *SUT1* and *SUT4* play an important role in sucrose transport, and these SUTs have sub-functional in sucrose accumulation and plant development (Zhang et al. 2016b). Tonoplast sugar transporters (TSTs) are highly associated with vacuolar sucrose accumulation, and it is speculated that TSTs are involved in the process of sequestering sucrose into the vacuoles of sugarcane stems (Casu et al. 2015; Bihmidine et al. 2016; Zhang et al. 2018). Recently, the evolutionary expansion and functional differentiation of ST genes in *S. spontaneum* were analyzed and some of them have been verified through heterologous expressions in defective yeast strain (Zhang et al. 2021c).

Starch Synthase (SS) Gene Families

Starch is the main storage carbohydrate in many plants, and starch synthases (SSs) have a central role in starch biosynthesis (Dry et al. 1992; Li et al. 2000). The structure and phylogeny of nine starch synthase genes identified in *S. spontaneum* were analyzed (Ma et al. 2019). At pre-mature and mature stages, the expression levels of SSs in the stem were significantly higher in *S. spontaneum* than in *S. officinarum*, and negatively correlated with the sucrose content between the two sugar species, which indicated that SSs were involved in the difference in carbohydrate metabolism between the two *Saccharum* species (Ma et al. 2019). These results are of great significance for further research on SSs and carbohydrate metabolism in sugarcane.

Photosynthesis in Saccharum

Malic Enzyme-Type C₄ Pathway

Some plants initially fixed CO₂ as C₄ instead of C₃ compounds such as sugarcane, maize and sorghum. Almost all C₄ lineages arose about 25 million years ago when atmospheric CO₂ levels had fallen, most of them occurred in hot, dry, and/or saline areas with high photorespiration potential (Sage et al. 2018). Increased expression of core C₄ genes played a major role in the evolution of C₄ photosynthesis (Sage 2004). In *S. spontaneum*, 24 genes of 7 key enzymes related to the nicotinamide adenine dinucleotide phosphate-malic enzyme (NADP-ME) C₄ pathway were identified, and neofunctionalization of *SsNADP-ME1* in sugarcane occurred after the divergence of maize and the Andropogoneae tribe (Zhang et al. 2018).

Magnesium Transporter (MGT) Gene Family

In plants, Mg²⁺ is essential for the function of many cellular enzymes and is the central atom of chlorophyll, which is considered an important nutrient for plant growth and photosynthesis (Shaul 2002). The magnesium transporter (MGT) family is composed of many membrane proteins, which play a vital role in maintaining magnesium homeostasis (Li et al. 2001). *MGTs* followed a circadian rhythm in the leaf tissues of *S. spontaneum*. The expression levels of *MGT3*, *MGT7*, and *MGT10* in *S. officinarum* were higher than those in *S. spontaneum*, (Wang et al. 2019). This study provides a foundation for further analysis of *MGTs* and their roles in sugarcane photosynthesis.

Transcription Factors in Saccharum

Transcription factors are DNA-binding proteins that play a key role in gene transcription and regulate plant physiological and biochemical processes. We have identified 1,933 and 2,022 transcription factors in the genomes of *S. spontaneum* and *S. officinarum*, respectively. Based on the Plant Transcription Factor Database of the University of Potsdam (<http://plntfdb.bio.uni-potsdam.de/v3.0/>), these transcription factors could be classified into 59 gene families, accounting for 5.43% (*S. spontaneum*) and 5.38% (*S. officinarum*) of their predicted proteins. We also compared sugarcane transcription factors with those in sorghum, Arabidopsis, rice and maize, in order to study their function (Unpublished data).

WRKY Gene Family

WRKY is one of the largest transcription factor families in plants and plays significant roles in plant stress responses. The expression patterns of WRKY genes varied in response to different stresses in plants (Yang et al. 2017b, 2020b; Chen et al. 2020). The expression of ScWRKY3 was stable in the smut-resistant *Saccharum* hybrid cultivar Yacheng05-179, while it was down-regulated in the smut-susceptible cultivar ROC22 following smut pathogenesis (Wang et al. 2018). The expression level of ScWRKY5 was significantly increased in two smut-resistant varieties, while it was decreased in three smut-susceptible varieties after inoculation with *Sporisorium scitamineum* for 1 day (Wang et al. 2020). In *S. spontaneum*, 294 sequences for 154 WRKY genes (SsWRKY) were identified, of which 13 (8.4%) had four alleles, 29 (18.8%) with three alleles and 41 (26.6%) with two alleles. Among them, 73.8% and 16.0% originated from segmental duplications and tandem duplications, respectively (Li et al. 2020). *SsWRKYs* showed distinct temporal and spatial expression patterns; 52 genes were expressed in all tissues, of which 21 may be involved in photosynthesis (Li et al. 2020). Further exploration on WRKY genes-mediated regulatory mechanisms associated with biotic and abiotic stress tolerance in sugarcane will be useful.

NBS-LRR

Breeding for disease resistance in sugarcane is difficult. Plant disease resistance genes share common structural features among distantly related species of plants. The most common features are nucleotide binding site and the leucine-rich repeat sequence (NBS-LRR) domains (Ye and Ting 2008; Luo et al. 2012). Isolation and identification of disease resistance genes can be time-consuming and expensive, but it is more straightforward to isolate

Resistance Gene Analogs (RGAs). NBS sequences were amplified from the genomic DNA and cDNA of smut-resistant sugarcane variety NCo376 (Que et al. 2009). In *S. spontaneum*, 80% of NBS-encoding genes are located in the four rearranged chromosomes (SsChr02, SsChr05, SsChr06 and SsChr07), and 51% of them are located in the rearranged region (Zhang et al. 2018). Most of the RGAs in modern sugarcane cultivars that respond to smut may originate in chromosome 5 of the ancestral *S. spontaneum* genotype. Smut resistant and susceptible genotypes of sugarcane have distinct expression patterns of RGAs (Rody et al. 2019). These results may help to develop a more practical procedure to improve resistance in sugarcane.

Future Prospects

In recent years, the advancement and application of high-throughput sequencing technology, especially the combination of third-generation ultra-long read sequencing and Hi-C technology, offer infinite possibilities to unveil the mysteries of sugarcane genome. Given the complex genetic background of hybrid sugarcane, it is essential to sequence the founding species *S. officinarum* and *S. spontaneum* as reference for assisting sugarcane hybrid genome assembly.

The breeding of modern sugarcane was restricted by the complexity of its genome (Jannoo et al. 1999a; Dal-Bianco et al. 2012). With the decrease in genome sequencing cost and the availability of a reference for *S. spontaneum* and the imminent publication of the *S. officinarum* sequence, it could be possible to detect genotype for the whole genetic populations. Since the pan-genome analysis for *Saccharum* based on large-scale genetic resource panels become practicable in the near future, one could expect discovery of novel genetic resources for sugarcane breeding. Given the fact that the modern sugarcane cultivars are hybrids which were mainly derived from crosses between *S. spontaneum* and *S. officinarum*, the cultivars are assumed to have open (or extensive) pan-genomes (Open pan-genome indicates number of gene of the pan-genome increases with the number of additionally sequenced populations). As *S. spontaneum* displays a high level of the genetic diversity, *S. officinarum* has a narrow genetic background. Therefore, it is necessary to conduct pan-genomic studies on both *S. spontaneum* and *S. officinarum*. For *S. spontaneum*, the research community may select the *S. spontaneum* accessions that represent maximum genetic diversities for this species (Aitken et al. 2018; Silva et al. 2018; Zhang et al. 2018; Yang et al. 2019; Meng et al. 2020) focusing on the genomic information related to stress tolerance, biomass production, ratooning, disease resistance and tillering. Compared to *S. spontaneum*, a few accessions are needed for the pan-genomic study on *S.*

officinarum mainly because of its narrow genetic background, with the target traits such as sugar accumulation, biomass production and photosynthesis. On these bases, the core genome and the “dispensable” genes would be further identified for these two *Saccharum* species, which will provide substantial knowledge for solving the genetic bottleneck of sugarcane breeding. Obviously, the pan-genomic study for modern sugarcane hybrids is very important and challenging, and preparation of a reference genome for *Saccharum* hybrids is still under progress due to its complex genome with high polyploidy.

So far, very few molecular makers have been used for selection in sugarcane breeding. The construction of genetic populations for elite agronomic traits such as sugar content, biomass, and disease resistance will become the key step for identifying the associated molecular markers, genes and quantitative trait loci. With the development of new molecular techniques and the associated statistical analysis methods, molecular breeding may become practicable in sugarcane breeding despite its complex genome.

Funding This work was supported by the Science and Technology Planting Project of Guangdong Province (2019B020238001), the National key research and development program [2018YFD1000104]; the National High-tech R&D Program [2013AA100604]; the Science and Technology Major Project of Guangxi [AA17202025]; the Fujian Provincial Department of Education [No. JA12082]; the Natural Science Foundation of Fujian Province, China (Grant Number 2019J0102).

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

Ethics Approval Not applicable.

References

- Aitken, K., J. Li, G. Piperidis, C. Qing, F. Yuanhong, and P. Jackson. 2018. Worldwide genetic diversity of the wild species *Saccharum spontaneum* and level of diversity captured within sugarcane breeding programs. *Crop Science* 58(1): 218–229.
- Aitken, K.S., J.C. Li, P. Jackson, G. Piperidis, and C.L. McIntyre. 2006. AFLP analysis of genetic diversity within *Saccharum officinarum* and comparison with sugarcane cultivars. *Australian Journal of Agricultural Research* 57(11): 1167–1184.
- Aitken, K.S., M.D. McNeil, P.J. Berkman, S. Hermann, A. Kilian, P.C. Bundock, and J. Li. 2014a. Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. *BMC Plant Biology* 14: 190.
- Aitken, K.S., M.D. McNeil, S. Hermann, P.C. Bundock, A. Kilian, K. Heller-Uszynska, R.J. Henry, and J. Li. 2014b. A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput Diversity Array Technology (DArT) markers. *BMC Genomics* 15(1): 152.

- Alkan, C., S. Sajjadian, and E.E. Eichler. 2011. Limitations of next-generation genome sequence assembly. *Nature Methods* 8(1): 61–65.
- Amalraj, V.A., and N. Balasundaram. 2006. On the taxonomy of the members of *Saccharum* complex. *Genetic Resources & Crop Evolution* 53(1): 35–41.
- Ansari, M.I., A. Yadav, and R. Lal. 2013. An-overview on invertase in sugarcane. *Bioinformation* 9: 464–465.
- Anur, R.M., N. Mufithah, W.D. Sawitri, H. Sakakibara, and B. Sugiharto. 2020. Overexpression of sucrose phosphate synthase enhanced sucrose content and biomass production in transgenic sugarcane. *Plants (Basel)* 9(2): 200.
- Banerjee, N., M.S. Khan, M. Swapna, R.K. Singh, and S. Kumar. 2020. Progress and prospects of association mapping in sugarcane (*Saccharum* species hybrid), a complex polyploid crop. *Sugar Tech* 22(6): 939–953.
- Baroja-Fernández, E., F. José Muñoz, J. Li, A. Bahaji, G. Almagro, M. Montero, E. Etxeberria, M. Hidalgo, M.T. Sesma, and J. Pozueta-Romero. 2012. Sucrose synthase activity in the *sus1/sus2/sus3/sus4* Arabidopsis mutant is sufficient to support normal cellulose and starch production. *Proceedings of the National Academy of Science* 109: 321.
- Barreto, F.Z., J. Rosa, T.W.A. Balsalobre, M.M. Pastina, R.R. Silva, H.P. Hoffmann, A.P. de Souza, A.A.F. Garcia, and M.S. Carneiro. 2019. A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). *PLoS ONE* 14(7): 9843.
- Bihmidine, S., B.T. Julius, I. Dweikat, and D.M. Braun. 2016. Tonoplast sugar transporters (SbTSTs) putatively control sucrose accumulation in sweet sorghum stems. *Plant Signaling & Behavior* 11(1): e1117721.
- Bosch, S., C.P.L. Grof, and F.C. Botha. 2004. Expression of neutral invertase in sugarcane. *Plant Science* 166(5): 1125–1133.
- Brown, J.S., R.J. Schnell, E.J. Power, S.L. Douglas, and D.N. Kuhn. 2007. Analysis of clonal germplasm from five *Saccharum* species: *S. barberi*, *S. robustum*, *S. officinarum*, *S. sinense* and *S. spontaneum*. A study of inter- and intra species relationships using microsatellite markers. *Genetic Resources & Crop Evolution* 54(3): 627–648.
- Cardozo, N.P., and P.C. Sentelhas. 2013. Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola* 70(6): 449–456.
- Casu, R.E., A.L. Rae, J.M. Nielsen, J.M. Perroux, G.D. Bonnett, and J.M. Manners. 2015. Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. *Plant Molecular Biology* 89(6): 607–628.
- Chan, A.P., J. Crabtree, Q. Zhao, H. Lorenzi, J. Orvis, D. Puiu, A. Melake-Berhan, K.M. Jones, J. Redman, G. Chen, E.B. Cahoon, M. Gedil, M. Stanke, B.J. Haas, J.R. Wortman, C.M. Fraser-Liggett, J. Ravel, and P.D. Rabinowicz. 2010. Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology* 28(9): 951–956.
- Chen, C., X. Chen, J. Han, W. Lu, and Z. Ren. 2020. Genome-wide analysis of the WRKY gene family in the cucumber genome and transcriptome-wide identification of WRKY transcription factors that respond to biotic and abiotic stresses. *BMC Plant Biology* 20(1): 443.
- Chen, Y., Q. Zhang, W. Hu, X. Zhang, L. Wang, X. Hua, Q. Yu, R. Ming, and J. Zhang. 2017. Evolution and expression of the fructokinase gene family in *Saccharum*. *BMC Genomics* 18(1): 1–15.
- Comai, L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6(11): 836–846.
- D’Hont, A., L. Grivet, P. Feldmann, S. Rao, N. Berding, and J.C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular Genetics and Genomics* 250(4): 405–413.
- D’Hont, A., D. Ison, K. Alix, C. Roux, and J.C. Glaszmann. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41(2): 221–225.
- da Silva, J., R.J. Honeycutt, W. Burnquist, S.M. Al-Janabi, M.E. Sorrells, S.D. Tanksley, and B.W.S. Sobral. 1995. *Saccharum spontaneum* L. ‘SES 208’ genetic linkage map combining RFLP- and PCR-based markers. *Molecular Breeding* 1(2): 165–179.
- Dal-Bianco, M., M.S. Carneiro, C.T. Hotta, R.G. Chapola, H.P. Hoffmann, A.A.F. Garcia, and G.M. Souza. 2012. Sugarcane improvement: How far can we go? *Current Opinion in Biotechnology* 23(2): 265–270.
- Daniels, D.C. 1993. Sugarcane in prehistory. *Archaeology in Oceania* 28(1): 1–7.
- Della Coletta, R., Y. Qiu, S. Ou, M.B. Hufford, and C.N. Hirsch. 2021. How the pan-genome is changing crop genomics and improvement. *Genome Biology* 22(1): 3.
- Deschamps, S., Y. Zhang, V. Llaca, L. Ye, A. Sanyal, M. King, G. May, and H. Lin. 2018. A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *Nature Communications* 9(1): 1–10.
- Dominguez Del Angel, V., E. Hjerde, L. Sterck, S. Capella-Gutierrez, C. Notredame, O. Vinnere Pettersson, J. Amselem, L. Bouri, S. Bocs, C. Klopp, J.-F. Gibrat, A. Vlasova, B.L. Leskosek, L. Soler, M. Binzer-Panchal, and H. Lantz. 2018. Ten steps to get started in genome assembly and annotation. *Research* 7: 148.
- Dry, I., A. Smith, A. Edwards, M. Bhattacharyya, P. Dunn, and C. Martin. 1992. Characterization of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. *The Plant Journal* 2(2): 193–202.
- Fickett, N., A. Gutierrez, M. Verma, M. Pontif, A. Hale, C. Kimbeng, and N. Baisakh. 2019. Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111(6): 1794–1801.
- Garsmeur, O., 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, and A. D’Hont. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications* 9(1): 2638.
- Greilhuber, J., J. Dolezel, M.A. Lysák, and M.D. Bennett. 2005. The origin, evolution and proposed stabilization of the terms “genome size” and “C-value” to describe nuclear DNA contents. *Annals of Botany* 95(1): 255–260.
- Ha, S., P.H. Moore, D. Heinz, S. Kato, N. Ohmido, and K. Fukui. 1999. Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescence in situ hybridization and imaging methods. *Plant Molecular Biology* 39(6): 1165–1173.
- Haigler, C.H., M. Ivanova-Datcheva, P.S. Hogan, V.V. Salnikov, S. Hwang, K. Martin, and D.P. Delmer. 2001. Carbon partitioning to cellulose synthesis. *Plant Molecular Biology* 47(1–2): 29–51.
- Hamilton, J.P., and C. Robin Buell. 2012. Advances in plant genome sequencing. *The Plant Journal* 70(1): 177–190.
- Hatch, M., and C. Slack. 1998. C 4 photosynthesis: discovery, resolution, recognition, and significance. In *Discoveries in plant biology: (Volume 1)*, 175–196. World Scientific.
- Hatch, M., C. Slack, and H. Johnson. 1967. Further studies on a new pathway of photosynthetic carbon dioxide fixation in sugar-cane

- and its occurrence in other plant species. *Biochemical Journal* 102(2): 417.
- Hert, D.G., C.P. Fredlake, and A.E. Barron. 2008. Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29(23): 4618–4626.
- Hodkinson, T.R., M.W. Chase, D.M. Lledó, N. Salamin, and S.A. Renvoize. 2002. Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnL intron and trnL-F intergenic spacers. *Journal of Plant Research* 115(5): 381–392.
- Hoeftner, S., and F. Botha. 2003. Expression of fructokinase isoforms in the sugarcane culm. *Plant Physiology and Biochemistry* 41: 741–747.
- Hu, W., X. Hua, Q. Zhang, J. Wang, Q. Shen, X. Zhang, K. Wang, Q. Yu, Y.R. Lin, and R. Ming. 2018. New insights into the evolution and functional divergence of the SWEET family in *Saccharum* based on comparative genomics. *BMC Plant Biology* 18(1): 1–20.
- International Rice Genome Sequencing, P. 2005. The map-based sequence of the rice genome. *Nature* 436(7052): 793–800.
- Irvine, J.E. 1999. *Saccharum* species as horticultural classes. *Theoretical & Applied Genetics* 98(2): 186–194.
- Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Huguency, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pè, G. Valle, M. Morgante, M. Caboche, A.-F. Adam-Blondon, J. Weissenbach, F. Quétier, P. Wincker, and C. The French-Italian Public Consortium for Grapevine Genome. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463–467.
- Jannoo, N., L. Grivet, A. Dookun, A. D'Hont, and J.C. Glaszmann. 1999a. Linkage disequilibrium among modern sugarcane cultivars. *Theoretical and Applied Genetics* 99(6): 1053–1060.
- Jannoo, N., L. Grivet, M. Seguin, F. Paulet, R. Domaigne, P.S. Rao, A. Dookun, A. D'Hont, and J.C. Glaszmann. 1999b. Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical & Applied Genetics* 99(1–2): 171–184.
- Jiao, Y., P. Peluso, J. Shi, T. Liang, M.C. Stitzer, B. Wang, M.S. Campbell, J.C. Stein, X. Wei, C.S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K.L. Schneider, T.K. Wolfgruber, M.R. May, N.M. Springer, E. Antoniou, W.R. McCombie, G.G. Presting, M. McMullen, J. Ross-Ibarra, R.K. Dawe, A. Hastie, D.R. Rank, and D. Ware. 2017. Improved maize reference genome with single-molecule technologies. *Nature* 546(7659): 524–527.
- Li, L., A.F. Tutone, R.S. Drummond, R.C. Gardner, and S. Luan. 2001. A novel family of magnesium transport genes in *Arabidopsis*. *The Plant Cell* 13(12): 2761–2775.
- Li, Z., X. Hua, W. Zhong, Y. Yuan, Y. Wang, Z. Wang, R. Ming, and J. Zhang. 2020. Genome-wide identification and expression profile analysis of WRKY family genes in the autopolyploid *Saccharum spontaneum*. *Plant and Cell Physiology* 61(3): 616–630.
- Li, Z., G. Mouille, B. Kosar-Hashemi, S. Rahman, B. Clarke, K.R. Gale, R. Appels, and M.K. Morell. 2000. The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family. *Plant Physiology* 123(2): 613–624.
- Liu, J., A.S. Seetharam, K. Chougule, S. Ou, K.W. Swentowsky, J.I. Gent, V. Llaca, M.R. Woodhouse, N. Manchanda, G.G. Presting, D.A. Kudrna, M. Alabady, C.N. Hirsch, K.A. Fengler, D. Ware, T.P. Michael, M.B. Hufford, and R.K. Dawe. 2020. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biology* 21(1): 121.
- Luo, S., Y. Zhang, Q. Hu, J. Chen, K. Li, C. Lu, H. Liu, W. Wang, and H. Kuang. 2012. Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiology* 159(1): 197–210.
- Ma, P., Y. Yuan, Q. Shen, Q. Jiang, and J. Zhang. 2019. Evolution and expression analysis of starch synthase gene families in *Saccharum spontaneum*. *Tropical Plant Biology* 12: 158–173.
- Ma, P., X. Zhang, L. Chen, Q. Zhao, Q. Zhang, X. Hua, Z. Wang, H. Tang, Q. Yu, M. Zhang, R. Ming, and J. Zhang. 2020. Comparative analysis of sucrose phosphate synthase (SPS) gene family between *Saccharum officinarum* and *Saccharum spontaneum*. *BMC Plant Biology* 20(1): 422.
- Meng, Z., J. Han, Y. Lin, Y. Zhao, Q. Lin, X. Ma, J. Wang, M. Zhang, L. Zhang, Q. Yang, and K. Wang. 2020. Characterization of a *Saccharum spontaneum* with a basic chromosome number of $x = 10$ provides new insights on genome evolution in genus *Saccharum*. *Theoretical and Applied Genetics* 133(1): 187–199.
- Metzker, M.L. 2010. Sequencing technologies—The next generation. *Nature Reviews Genetics* 11(1): 31–46.
- Mitros, T., A.M. Session, B.T. James, G.A. Wu, M.B. Belaffif, L.V. Clark, S. Shu, H. Dong, A. Barling, J.R. Holmes, J.E. Mattick, J.V. Bredeson, S. Liu, K. Farrar, K. Glowacka, S. Jeżowski, K. Barry, W.B. Chae, J.A. Juvik, J. Gifford, A. Oladeinde, T. Yamada, J. Grimwood, N.H. Putnam, J. De Vega, S. Barth, M. Klaas, T. Hodkinson, L. Li, X. Jin, J. Peng, C.Y. Yu, K. Heo, J.H. Yoo, B.K. Ghimire, I.S. Donnison, J. Schmutz, M.E. Hudson, E.J. Sacks, S.P. Moose, K. Swaminathan, and D.S. Rokhsar. 2020. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nature Communications* 11(1): 5442.
- Mukherjee, S.K. 1957. Origin and distribution of *Saccharum*. *Botanical Gazette* 119(1): 55–61.
- Nagaki, K., and M. Murata. 2005. Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. *Chromosome Research an International Journal on the Molecular Supramolecular & Evolutionary Aspects of Chromosome Biology* 13(2): 195.
- Nagaki, K., H. Tsujimoto, and T. Sasakuma. 1998. A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. *Chromosome Research an International Journal on the Molecular Supramolecular & Evolutionary Aspects of Chromosome Biology* 6(4): 295.
- Okura, V.K., R.S. de Souza, S.F. de Siqueira Tada, and P. Arruda. 2016. BAC-pool sequencing and assembly of 19 Mb of the complex sugarcane genome. *Frontiers in Plant Science* 7: 342.
- Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Pliakov, 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. O'tillar, 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. Mehboobur, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, and D.S. Rokhsar. 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229): 551–556.
- Paterson, A. H., X. Wang, J. Li, and H. Tang. 2012. Ancient and recent polyploidy in monocots. In *Polyploidy and genome evolution*, 93–108. Springer.

- Pellicer, J., O. Hidalgo, S. Dodsworth, and I.J. Leitch. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* 9(2): 88.
- Piperidis, N., and A. D'Hont. 2020. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *The Plant Journal* 103(6): 2039–2051.
- Price, S. 1963. Cytogenetics of modern sugar canes. *Economic Botany* 17(2): 97–106.
- Que, Y.-X., L.-P. Xu, J.-W. Lin, and R.-K. Chen. 2009. Isolation and characterization of NBS-LRR resistance gene analogs from sugarcane. *Acta Agronomica Sinica* 35(4): 631–639.
- Racedo, J., L. Gutiérrez, M.F. Perera, S. Ostengo, E.M. Pardo, M.I. Cuenya, B. Welin, and A.P. Castagnaro. 2016. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biology* 16(1): 142.
- Renny-Byfield, S., and J.F. Wendel. 2014. Doubling down on genomes: Polyploidy and crop plants. *American Journal of Botany* 101(10): 1711–1725.
- Renz, A., and M. Stitt. 1993. Substrate specificity and product inhibition of different forms of fructokinases and hexokinases in developing potato tubers. *Planta* 190(2): 166–175.
- Riaño-Pachón, D.M., and L. Mattiello. 2017. Draft genome sequencing of the sugarcane hybrid SP80–3280. *F1000Res* 6: 861.
- Rody, H.V.S., R.G.H. Bombardelli, S. Creste, L.E.A. Camargo, M.-A. Van Sluys, and C.B. Monteiro-Vitorello. 2019. Genome survey of resistance gene analogs in sugarcane: Genomic features and differential expression of the innate immune system from a smut-resistant genotype. *BMC Genomics* 20(1): 809–809.
- Sage, R., R. Monson, J. Ehleringer, S. Adachi, and R.W. Pearcy. 2018. Some like it hot: The physiological ecology of C4 plant evolution. *Oecologia* 187: 941–966.
- Sage, R.F. 2004. The evolution of C4 photosynthesis. *New Phytologist* 161(2): 341–370.
- Salse, J., S. Bolot, M. Throude, V. Jouffe, B. Piegu, U.M. Quraishi, T. Calcagno, R. Cooke, M. Delseny, and C. Feuillet. 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20(1): 11–24.
- Sanger, F., G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, J.C. Fiddes, C.A. Hutchison, P.M. Slocombe, and M. Smith. 1977. Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* 265(5596): 687–695.
- Schatz, M.C., A.L. Delcher, and S.L. Salzberg. 2010. Assembly of large genomes using second-generation sequencing. *Genome Research* 20(9): 1165–1173.
- Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, and T.A. Graves. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956): 1112–1115.
- Shaul, O. 2002. Magnesium transport and function in plants: The tip of the iceberg. *BioMetals* 15(3): 307–321.
- Shi, Y., H. Xu, Q. Shen, J. Lin, and J. Zhang. 2019. Comparative analysis of SUS gene family between *Saccharum officinarum* and *Saccharum spontaneum*. *Tropical Plant Biology* 12: 174–185.
- Silva, J.A.G.D., P.M.D.A. Costa, T.G. Marconi, E.J.D.S. Barreto, N. Solís-Gracia, J.-W. Park, and N.C. Glynn. 2018. Agronomic and molecular characterization of wild germplasm *Saccharum spontaneum* for sugarcane and energycane breeding purposes. *Scientia Agricola* 75(4): 329–338.
- Soreng, R.J., P.M. Peterson, K. Romaschenko, G. Davidse, F.O. Zuloaga, E.J. Judziewicz, and T.S. Filgueiras. 2015. A world-wide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* 53(002): 117–137.
- Souza, G.M., H. Berges, S. Bocs, R. Casu, A. D'Hont, J.E. Ferreira, R. Henry, R. Ming, B. Potier, and M.A.V. Sluys. 2011. The sugarcane genome challenge: Strategies for sequencing a highly complex genome. *Tropical Plant Biology* 4(3–4): 145–156.
- Souza, G.M., M.-A. Van Sluys, C.G. Lembke, H. Lee, G.R.A. Margarido, C.T. Hotta, J.W. Gaiarsa, A.L. Diniz, M.D.M. Oliveira, S.D.S. Ferreira, M.Y. Nishiyama, F. Ten-Caten, G.T. Ragagnin, P.D.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, and D. Heckerman. 2019. 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(12): giz129.
- Swaminathan, K., W.B. Chae, T. Mitros, K. Varala, L. Xie, A. Barling, K. Glowacka, M. Hall, S. Jezowski, and R. Ming. 2012. A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics* 13(1): 1–17.
- Talbert, P.B., R. Masuelli, A.P. Tyagi, L. Comai, and S. Henikoff. 2002. Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *The Plant Cell* 14(5): 1053–1066.
- Thirugnanasambandam, P.P., N.V. Hoang, and R.J. Henry. 2018. The challenge of analyzing the sugarcane genome. *Frontiers in Plant Science* 9: 616.
- Thirugnanasambandam, P.P., P.J. Mason, N.V. Hoang, A. Furtado, F.C. Botha, and R.J. Henry. 2019. Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. *BMC Plant Biology* 19(1): 160.
- Tibbs Cortes, L., Z. Zhang, and J. Yu. 2021. Status and prospects of genome-wide association studies in plants. *The Plant Genome* 14(1): e20077.
- Tu, Y., J. Sun, X. Ge, and Z. Li. 2009. Chromosome elimination, addition and introgression in intertribal partial hybrids between *Brassica rapa* and *Isatis indigotica*. *Annals of Botany* 103: 1039–1048.
- van Dijk, E.L., Y. Jaszczyszyn, D. Naquin, and C. Thermes. 2018. The third revolution in sequencing technology. *Trends in Genetics* 34(9): 666–681.
- VanBuren, R., C. Man Wai, X. Wang, J. Pardo, A.E. Yocca, H. Wang, S.R. Chaluvadi, G. Han, D. Bryant, P.P. Edger, J. Messing, M.E. Sorrells, T.C. Mockler, J.L. Bennetzen, and T.P. Michael. 2020. Exceptional subgenome stability and functional divergence in the allotetraploid *Ethiopian cereal teff*. *Nature Communications* 11(1): 884.
- Velasco, R., A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S.K. Bhatnagar, M. Troggio, D. Pruss, S. Salvi, M. Pindo, P. Baldi, S. Castelletti, M. Cavaiuolo, G. Coppola, F. Costa, V. Cova, A. Dal Ri, V. Goremykin, M. Komjanc, S. Longhi, P. Magnago, G. Malacarne, M. Malnoy, D. Micheletti, M. Moretto, M. Perazzolli, A. Si-Ammour, S. Vezzulli, E. Zini, G. Eldredge, L.M. Fitzgerald, N. Gutin, J. Lanchbury, T. Macalma, J.T. Mitchell, J. Reid, B. Wardell, C. Kodira, Z. Chen, B. Desany, F. Niazi, M. Palmer, T. Koepke, D. Jiwan, S. Schaeffer, V. Krishnan, C. Wu, V.T. Chu, S.T. King, J. Vick, Q. Tao, A. Mraz, A. Stormo, K. Stormo, R. Bogden, D. Ederle, A. Stella, A. Vecchiatti, M.M. Kater, S. Masiero, P. Lasserre, Y. Lespinasse, A.C. Allan, V. Bus, D. Chagné, R.N. Crowhurst, A.P. Gleave, E. Lavezzo, J.A. Fawcett, S. Proost, P. Rouzé, L. Sterck, S. Toppo, B. Lazzari, R.P. Hellens, C.-E. Durel, A. Gutin, R.E. Bumgarner, S.E. Gardiner, M. Skolnick, M. Egholm, Y. Van de Peer, F. Salamini, and R. Viola. 2010. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature Genetics* 42(10): 833–839.
- Verma, A.K., S. Upadhyay, P.C. Verma, S. Solomon, and S.B. Singh. 2011. Functional analysis of sucrose phosphate synthase (SPS)

- and sucrose synthase (SS) in sugarcane (*Saccharum*) cultivars. *Plant Biology* 13(2): 325–332.
- Vilela, M.M., L.E. Del Bem, M.A. Van Sluys, N. de Setta, J.P. Kitajima, G.M. Cruz, D.A. Sforça, A.P. de Souza, P.C. Ferreira, C. Gratiol, C.B. Cardoso-Silva, R. Vicentini, and M. Vincentz. 2017. Analysis of three sugarcane homo/homeologous regions suggests independent polyploidization events of *Saccharum officinarum* and *Saccharum spontaneum*. *Genome Biology and Evolution* 9(2): 266–278.
- Waclawovsky, A.J., P.M. Sato, C.G. Lembke, P.H. Moore, and G.M. Souza. 2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal* 8(3): 263–276.
- Wang, D., L. Wang, W. Su, Y. Ren, C. You, C. Zhang, Y. Que, and Y. Su. 2020. A class III WRKY transcription factor in sugarcane was involved in biotic and abiotic stress responses. *Scientific Reports* 10(1): 20964.
- Wang, J., B. Roe, S. Macmil, Q. Yu, J.E. Murray, H. Tang, C. Chen, F. Najjar, G. Wiley, J. Bowers, M.-A. Van Sluys, D.S. Rokhsar, M.E. Hudson, S.P. Moose, A.H. Paterson, and R. Ming. 2010. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11(1): 261.
- Wang, L., F. Liu, X. Zhang, W. Wang, T. Sun, Y. Chen, M. Dai, S. Yu, L. Xu, Y. Su, and Y. Que. 2018. Expression characteristics and functional analysis of the ScWRKY3 gene from sugarcane. *International Journal Molecular Science* 19(12): 4059.
- Wang, L., Y. Zheng, S. Ding, Q. Zhang, and Y. Chen. 2017. Molecular cloning, structure, phylogeny and expression analysis of the invertase gene family in sugarcane. *BMC Plant Biology* 17(1): 1–20.
- Wang, X., H. Tang, and A.H. Paterson. 2011. Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. *The Plant Cell* 23(1): 27–37.
- Wang, Y., X. Hua, J. Xu, Z. Chen, T. Fan, Z. Zeng, H. Wang, A. L. Hour, Q. Yu, and R. Ming. 2019. Comparative genomics revealed the gene evolution and functional divergence of magnesium transporter families in *Saccharum*. *BMC Genomics* 20(1).
- Yang, X., Z. Luo, J. Todd, S. Sood, and J. Wang. 2020a. Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *Plant Genome* 13(1): e20006.
- Yang, X., J. Song, J. Todd, Z. Peng, D. Paudel, Z. Luo, X. Ma, Q. You, E. Hanson, Z. Zhao, Y. Zhao, J. Zhang, R. Ming, and J. Wang. 2019. Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (*Saccharum* spp.), a ‘sweet’ crop with ‘bitter’ genomes. *Plant Biotechnology Journal* 17(2): 488–498.
- Yang, X., J. Song, Q. You, D.R. Paudel, J. Zhang, and J. Wang. 2017a. Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics* 18(1): 594.
- Yang, Y., J. Liu, X. Zhou, S. Liu, and Y. Zhuang. 2020b. Identification of WRKY gene family and characterization of cold stress-responsive WRKY genes in eggplant. *PeerJ* 8: e8777.
- Yang, Y., Y. Zhou, Y. Chi, B. Fan, and Z. Chen. 2017b. Characterization of soybean WRKY gene family and identification of soybean WRKY genes that promote resistance to soybean cyst nematode. *Scientific Reports* 7(1): 17804.
- Ye, Z., and J.P. Ting. 2008. NLR, the nucleotide-binding domain leucine-rich repeat containing gene family. *Current Opinion in Immunology* 20(1): 3–9.
- Zan, F., Y. Zhang, Z. Wu, J. Zhao, C. Wu, Y. Zhao, X. Chen, L. Zhao, W. Qin, and L. Yao. 2020. Genetic analysis of agronomic traits in elite sugarcane (*Saccharum* spp.) germplasm. *PLoS ONE* 15(6): e0233752.
- Zhang, J., C. Nagai, Q. Yu, Y.B. Pan, T. Ayala-Silva, R.J. Schnell, J.C. Comstock, A.K. Arumuganathan, and R. Ming. 2012. Genome size variation in three *Saccharum* species. *Euphytica* 185(3): 511–519.
- Zhang, J., J. Arro, Y. Chen, and R. Ming. 2013. Haplotype analysis of sucrose synthase gene family in three *Saccharum* species. *BMC Genomics* 14.
- Zhang, J., M. Zhou, J. Walsh, L. Zhu, Y. Chen, and R. Ming. 2014. Sugarcane genetics and genomics. *Sugarcane: Physiology, biochemistry, and functional biology*. Edited by Moore PH, Botha FC. Wiley-Blackwell physiology, biochemistry, and functional biology. Edited by Moore PH, Botha FC. Wiley-Blackwell Publishing:623–643.
- Zhang, J., L.L. Chen, F. Xing, D.A. Kudrna, W. Yao, D. Copetti, T. Mu, W. Li, J.M. Song, W. Xie, S. Lee, J. Talag, L. Shao, Y. An, C.L. Zhang, Y. Ouyang, S. Sun, W.B. Jiao, F. Lv, B. Du, M. Luo, C.E. Maldonado, J.L. Goicoechea, L. Xiong, C. Wu, Y. Xing, D.X. Zhou, S. Yu, Y. Zhao, G. Wang, Y. Yu, Y. Luo, Z.W. Zhou, B.E. Hurtado, A. Danowitz, R.A. Wing, and Q. Zhang. 2016a. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proceedings of the National Academy of Sciences of the United States of America* 113(35): E5163–5171.
- Zhang, Q., W. Hu, F. Zhu, L. Wang, Q. Yu, R. Ming, and J. Zhang. 2016b. Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in *Saccharum*. *BMC Genomics* 17(1): 88.
- Zhang, W., S. Zuo, Z. Li, Z. Meng, J. Han, J. Song, Y.-B. Pan, and K. Wang. 2017. Isolation and characterization of centromeric repetitive DNA sequences in *Saccharum spontaneum*. *Scientific Reports* 7(1): 41659.
- Zhang, J., X. Zhang, H. Tang, Q. Zhang, X. Hua, X. Ma, F. Zhu, T. Jones, X. Zhu, and J. Bowers. 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* 50(11): 1565–1573.
- Zhang, J., Q. Zhang, L. Li, H. Tang, Q. Zhang, Y. Chen, J. Arrow, X. Zhang, A. Wang, and C. Miao. 2019. Recent polyploidization events in three *Saccharum* founding species. *Plant Biotechnology Journal* 17(1): 264–274.
- Zhang, G., C. Ge, P. Xu, S. Wang, S. Cheng, Y. Han, Y. Wang, Y. Zhuang, X. Hou, T. Yu, X. Xu, S. Deng, Q. Li, Y. Yang, X. Yin, W. Wang, W. Liu, C. Zheng, X. Sun, Z. Wang, R. Ming, S. Dong, J. Ma, X. Zhang, and C. Chen. 2021a. The reference genome of *Miscanthus floridulus* illuminates the evolution of Saccharinae. *Nature Plants* 7(5): 608–618.
- Zhang, J., Z. Qing, Q. Yiyi, P. Haoran, W. Gang, H. Xiuting, W. Yongjun, L. Lianyu, L. Zhen, L. Yihan, M. Panpan, D. Meijie, W. Yibin, W. Hengbo, Z. Xingtang, Y. Wei, W. Yuntong, L. Xinlong, W. Maojun, W. Jianping, D. Zuhu, Y. Qinghui, C. Baoshan, Z. Muqing, T. Haibao, and M. Ray. 2021b. *Research Square*.
- Zhang, Q., X. Hua, H. Liu, Y. Yuan, Y. Shi, Z. Wang, M. Zhang, R. Ming, and J. Zhang. 2021c. Evolutionary expansion and functional divergence of sugar transporters in *Saccharum* (*S. spontaneum* and *S. officinarum*). *The Plant Journal* 105(4): 884–906.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.