Identifying Candidate Genes for Sugar Accumulation in Sugarcane Cultivars: From a Syntenic Genomic Region to a Gene Coexpression Network

Mônica Letícia Turibio Martins[1,2]; Danilo Augusto Sforça[1], Luís Paulo dos Santos[1,2]; Ricardo José Gonzaga Pimenta[1,2]; Melina Cristina Mancini[1,2]; Alexandre Hild Aono[1,2]; Cláudio Benício Cardoso da Silva[1,2,3]; Sonia Vautrin[5]; Arnaud Bellec[5]; Renato Vicentini[1,2]; Helene Bérgès[5]; Carla Cristina da Silva[1,2,4]; Anete Pereira de Souza[1,2]

1 Center for Molecular Biology and Genetic Engineering (CBMEG) – State University of Campinas (UNICAMP), Campinas, SP, Brazil

2 Institute of Biology (IB) – State University of Campinas (UNICAMP), Campinas, SP, Brazil

3 National Laboratory of Biorenewables—LNBR/CNPEM, Campinas, SP, Brazil

4 Agronomy Department, Federal University of Viçosa, Viçosa, MG, Brazil

5 Centre National de Ressources Génomiques Végétales (CNRGV/INRA) Toulouse, France

Abstract

Elucidating the intricacies of the sugarcane genome is essential for breeding superior cultivars. This economically important crop originates from hybridizations of highly polyploid *Saccharum* species. However, the large size (10 Gb), high polyploidy, and aneuploidy of the sugarcane genome pose significant challenges to complete genome sequencing, assembly, and annotation. One successful strategy for identifying candidate genes linked to agronomic traits, particularly those associated with sugar accumulation, leverages synteny and potential collinearity with related species. In this study, we explored synteny between sorghum and sugarcane. Genes from a sorghum Brix QTL were used to screen bacterial artificial chromosome (BAC) libraries from two Brazilian sugarcane varieties (IACSP93-3046 and SP80-3280). The entire region was successfully recovered, confirming synteny and collinearity between the species. Manual annotation

identified 51 genes in the hybrid varieties that were subsequently confirmed to be present in *Saccharum spontaneum*. To identify candidate genes for sugar accumulation, this study employed a multifaceted approach, including retrieving the genomic region of interest, performing gene-by-gene analysis, analyzing RNA-seq data of internodes from *Saccharum officinarum* and *S. spontaneum* accessions, constructing a coexpression network to examine the expression patterns of genes within the studied region and their neighbors, and finally identifying differentially expressed genes (DEGs). This comprehensive approach led to the discovery of three candidate genes potentially involved in sugar accumulation: an ethylene-responsive transcription factor (ERF), an ABA 8'-hydroxylase, and a prolyl oligopeptidase (POP). These findings could be valuable for identifying additional candidate genes for other important agricultural traits and directly targeting candidate genes for further work in molecular breeding.

Keywords:

Brix, Polyploidy, Candidate genes**,** Sugar accumulation, Gene expression

Introduction

In the 1880s, sugarcane (*Saccharum* spp.) farmers crossed *Saccharum spontaneum* (2n=5x=40 to 16x=128, x=8), which is resistant to biotic and abiotic stress, with *Saccharum officinarum* (2n=8x=80, x=10), which is considered a noble sugarcane due its high amount of sugar. To maintain a high sugar content in hybrids, successive backcrosses with *S. officinarum* were performed (Bremer, 1961, D'Hont et al., 1996; Garsmeur et al., 2018; Babu et al., 2022 and Healey et al., 2024). The crosses between both species generated modern sugarcane cultivars: plants with large genomes (10 Gb) that are highly polyploid and aneuploid with at least 50% repetitive regions (Cuadrado, 2004; Piperidis et al., 2010; Garcia et al., 2013; Thirugnanasambandam et al., 2018, Healey et al., 2024). The hybrid genome is a mixture of chromosomes originating from *S. officinarum* (70-80% of all chromosomes of the hybrids) and *S. spontaneum* (10-20%) and recombinant chromosomes (5-10%) (D'Hont et al., 1996; Cuadrado, 2004; Piperidis et al., 2010; Garsmeur et al., 2018 and Healey et al., 2024). The variable ploidy intrinsic

to each genotype creates a unique genomic structure with chromosome numbers varying between 40 and 128 (D'Hont et al., 1996; Zhang et al., 2012; Garsmeur et al., 2018; Thirugnanasambandam et al., 2018), which renders the study of the sugarcane genome a challenge (Thirugnanasambandam et al., 2018, Babu et al., 2022).

Grasses with a reference genome, such as rice (International Rice Genome Sequencing Project and Sasaki, 2005), maize (Schnable et al., 2009), wheat (IWGSC et al., 2018), and miscanthus *(Miscanthus sinensis)* (Kim et al., 2014; Tsuruta et al., 2017), even sugarcane with an allele-defined genome of *Saccharum spontaneum* (Zhang et al., 2018) or a monoploid sequence reference for sugarcane (Garsmeur et al., 2018) and sorghum (McCormick et al., 2018), are commonly used as references for studies on the sugarcane genome (Thirugnanasambandam et al, 2018; Wang et al., 2021; Babu et al., 2022, Mancini et al., 2018; Garsmeur et al., 2018; Zhang et al., 2018, Zhang et al., 2022 and Healey et al., 2024). Recently, a highly representative genome of the R570 hybrid variety was presented to the scientific community (Healey et al., 2024). The *Miscanthus* genome has sorghum as an important reference for its assembly and annotation (Mitros et al., 2020), and sorghum is an ancestor of the *Saccharum* and *Miscanthus* groups. Sorghum has an assembled and annotated diploid genome that is one-tenth of the sugarcane genome size and diverged approximately 8 million years ago; nevertheless, its genome has maintained strong synteny and collinearity (Ming et al., 1998; Paterson et al., 2009; Wang et al., 2010; Garsmeur et al., 2018, Thirugnanasambandam et al., 2018; Babu et al., 2022). The R570 has a high inbreeding coefficient, with approximately half being identical by descent. Therefore, it is expected that its genome will have a small portion missing and another collapsed, totaling 8.72 Gb, approaching the estimated size of 10 Gb (Healey et al., 2024).

Therefore, choosing a sorghum quantitative trait locus (QTL) for a trait of interest and recovering its orthologous region in sugarcane can be an efficient strategy to retrieve a potentially target sugarcane genomic region (Mancini et al., 2018). Recent studies with the R570 variety confirmed that a significant portion of the alleles originated from *S. officinarum*, so the sugar-accumulating origin is identical and thus largely inaccessible to QTL mapping efforts (Healey et al., 2024). However, the genomic sequence of a cultivar does not fully reflect the genetic information about the species (Montenegro et al., 2017), where a cultivar may not be representative of the entire genomic content of the species (Garsmeur et al., 2018; Thirugnanasambandam et al., 2018 and Healey et al., 2024). The

development of new commercial sugarcane varieties with higher yields is the main goal of most breeding programs. To increase yield, they seek genotypes that can tolerate biotic and abiotic stress but also have increased sugar accumulation (CURSI et al., 2021 and Healey et al., 2024). The exploration of candidate genes can be aided by the use of other omic technologies, such as RNA sequencing (RNA-seq), which offers a differential assessment not only between specific tissues but also between notably different varieties and related species (Stark et al., 2019). The transcriptome allows for the identification of differentially expressed genes (DEGs) and can be leveraged to construct coexpression networks, aiming to identify coexpressed genes and expand the evidence leading to candidate genes.

Many genes involved in the synthesis and transport of sucrose have been identified in sugarcane (Zhu et al. 2000; Carson and Botha 2002; Grivet and Arruda 2002; Casu et al. 2003; Vasantha et al., 2022). Although sucrose is synthesized in the cytosol of mesophyll cells in most plants, sugarcane requires the involvement of two cell types: the bundle sheath and mesophyll. Sucrose synthesis occurs predominantly in the mesophyll, utilizing glucose phosphates, which are then translocated through the conducting strands of sheath to the vascular compartments of internodal tissues, where they finally accumulate (Vasantha et al., 2022). Moreover, during the plant maturation phase, the sucrose concentration in culms increases, while the proportion of glucose and fructose decreases (Chandra et al., 2012). However, many processes related to sugar accumulation in sugarcane internodes are not fully understood, and possible pathways and related genes have yet to be identified.

The accumulation of proline in plant cells is associated with various physiological processes, such as cellular homeostasis, aiding in water absorption, and adaptation to abiotic stresses, enhancing the plant's adaptive response (Rejeb et al., 2014; Kazemi-Shahandashti & Maali-Amiri, 2018; Sharma et al., 2014; Kazemi-Shahandashti & Maali-Amiri, 2018; Sharma et al., 2014; Kazemi-Shahandashti & Maali-Amiri, 2018; Sharma et al., 2019; Ghosh et al., 2021). The interaction of proline accumulation with sucrose during salt stress has been reported in sugarcane (Ghosh et al., 2019). The enzyme prolyl oligopeptidase (POP—serine protease family clan SC, family S9) is a cytoplasmic enzyme that hydrolyzes oligopeptides up to 30 residues and occurs at the C-terminal side of proline residues (Gutierrez et al., 2008; Baharin et al., 2022). In plants, POP has been

associated with responses to biotic and abiotic stresses (Gutierrez et al., 2008; Singh et al., 2011; Tan et al., 2013).

In this context, a sorghum QTL for Brix (Shiringani et al., 2010) in sorghum was chosen as a target because its orthologous region was recovered in two Brazilian cultivars (SP80-3280 and IACSP93-3046), in R570 (Garsmeur et al., 2018) and in the *S. spontaneum* genome (Zhang et al. 2018). These genomic regions were compared to understand the level of genomic structural variation and genetic differences among sugarcane and sorghum. The genes found in this region were used to search for candidate genes for sugar accumulation through gene annotation evaluation, differential expression analysis in sugarcane stem transcriptomes (Aono et al., 2021) and a coexpression network. The combination of such strategies provides a more comprehensive and robust perspective in the search for candidate genes related to sugar accumulation in sugarcane. In exploring the region's genes, a set of evidence combining genetic/genomic factors revealed three candidate genes related to sugar accumulation characteristics.

Materials and methods

Sorghum region of interest

A partial QTL that was mapped in sorghum for Brix, which is genetically located on chromosome SBI-02 between the EST-SSR markers Xtxp56 and Stgnhsbm36 (Shiringani et al., 2010), was selected. The marker sequences were used to define the physical chromosomal position using the v3.1 version of the S*orghum bicolor* genome (Paterson et al., 2009) available in the Phytozome 13.0 database (https://phytozome-next.jgi.doe.gov/, Goodstein et al., 2012). The QTL has a phenotypic variation with 21.9% explained by genotype (R^2) % and a logarithm of odds (LOD) value of 10.08 (Shiringani et al., 2010). It spans from 61,568 kb to 61,952 kb on the sorghum chromosome SBI-02, totaling an approximate length of 385 kb. The target region was defined between 61,500 kb and 62,000 kb.

Plant material

Two Brazilian sugarcane cultivars were analyzed in the present study. SP80-3280 is known for its high production of sucrose and good tillering. It is resistant to smut, mosaic, and rust and tolerant to scald (Embrapa—Brazilian Agricultural Research Company, 2022). The SP80-3280 variety has been widely used in studies to understand

sugarcane genomics and genetics. This variety has a collection of sugarcane expressed sequence tags (SUCEST, Vettore, 2003), transcriptomes (Cardoso-Silva et al., 2014; Nishiyama et al., 2014; Mattiello et al., 2015), mapped QTLs (Aitken et al., 2006; Costa et al., 2016), a draft genome (Riaño-Pachón and Mattiello et al., 2017), and bacterial artificial chromosome (BAC) libraries (Figueira et al., 2012; Sforça et al., 2019). The use of sorghum synteny and collinearity has also been the focus of an approach for restoring genomic regions of agronomic interest (Mancini et al., 2018). The economic importance of the IACSP93-3046 cultivar is due to its high sucrose content, good tillering, resistance to rust and suitability for mechanized harvesting (Mancini et al., 2012). This cultivar also has a transcriptome (Cardoso-Silva et al., 2014) and a BAC library (Sforça et al., 2019).

Recovering the sorghum ortholog region in sugarcane

Primer design: The coding sequences (CDSs) of the genes within the target sorghum genomic region were recovered, as well as five genes before the delimited region and two genes after the delimited region, totaling 58 genes; the BLASTn algorithm (AltschuP et al., 1990) was used to align the CDS against sugarcane leaf transcripts (Cardoso-Silva et al., 2014) with a cutoff of E < 1e-10. Sorghum gene sequences that did not have similar transcripts in the sugarcane leaf transcriptome were compared to those in the SUCEST database (Vettore, 2003) and the NCBI database (AltschuP et al., 1990). Only genes that aligned with sugarcane leaf transcripts, were in the SUCEST database or NCBI database, had putative exons of 200 base pairs (bp) or larger and were not duplicated in the sorghum genome were used for primer development.

Identification of BAC clones, sequencing and assembly: To recover the sequences of interest in the equivalent target region of sorghum in sugarcane varieties, BAC libraries from the varieties SP80-3280 and IACSP93-3046 (Sforça et al., 2019) were used. Positive clone selection and preparation of BAC DNA for sequencing and pooling followed the steps described by Mancini et al. (2018). Sequencing was performed on the PacBio® Sequel platform (Pacific Biosciences) at the Arizona Genomics Institute (AGI—Tucson, USA). Vector and *Escherichia coli* genomic sequences were removed with the BBtools package (https://sourceforge.net/projects/bbmap/). Assembly was performed with the Canu v2.1 program (Koren et al., 2017) with default parameters, except for corOutCoverage = 200. The refinement of the final contig consensus sequence was

performed by aligning the raw reads against the assembled contigs with the pbalign program, and error correction was performed with the Arrow program. Both programs are present in the SMRTLink v7.0 package (Pacific Biosciences).

Annotation of contig sequences: The annotation of BACs for repetitive elements was performed using the LTR FINDER retrotransposon predictor (Xu and Wang, 2007) and the giriREPBASE database (Kohany et al., 2006). Gene annotation was performed with the NCBI (Altschul et al., 1990) and Phytozome v12.0 (Goodstein et al., 2012) databases. The Artemis program of the Sanger Institute (Rutherford et al., 2000) was used to visualize genes and repetitive elements. The sorghum CDSs and the manually annotated sugarcane variety CDSs were used to perform similarity searches using BLASTn tools (Altschul et al., 1990) against the following databases: NCBI (Sayers et al., 2022), UniProt (The UniProt Consortium, 2023) and Pfam (Protein Families) (Mistry et al., 2021). Genes were considered similar if they exhibited a sequence identity of 80% or greater. Contigs that did not have genes, had only one gene or were smaller than 25 kb in size were discarded.

Manual curation of orthologous regions in S. spontaneum: Manual homology curation of the orthologous regions in *S. spontaneum* was performed. The CDS of each QTL sorghum gene was aligned against the four alleles of the Sspon02 chromosomal set using BLASTn tools (Altschul et al., 1990). This allowed for enhanced accuracy of the automated annotation performed by Zhang et al., 2018, including the identification of pseudogenes, thereby providing a more precise definition of the genomic architecture in this specific region in *S. spontaneum.*

The information obtained was used for a detailed literature review of each gene. This review describes the proteins and their functions, the biological pathways in which they are supposedly implicated, and their potential role in sugar accumulation in plants, particularly in grasses and sugarcane.

Comparative genomic analyses: Comparative analyses were performed between the genes present in the target region in both varieties. In addition, the genes of the orthologous region in the *S. bicolor, S. spontaneum* (Zhang et al., 2018) and the sugarcane hybrid variety R570 (Garsmeur et al., 2018) genomes (Phytozome 12) were also used for comparative analysis. The analyses were performed to determine the synteny, collinearity and genomic structure of the region.

The homologous region in *S. spontaneum* was located using the BLASTn tool (Altschul et al. 1990) against the four homologous chromosome sequences of the *S. spontaneum* homologous chromosome 02 group (Sspon02), thus called Sspon2A, Sspon2B, Sspon2C and Sspon2D (Zhang et al., 2018). Genes were manually curated only in the sorghum orthologous region using the Artemis program from the Sanger Institute (Rutherford et al., 2000) for visualization. Manual curation was performed with NCBI databases (Sayers et al., 2022) and Phytozome 12.0 (Goodstein et al., 2012) databases with visualization through the Artemis program of the Sanger Institute (Rutherford et al., 2000).

Differential gene expression analysis

The expression of genes within the QTL region was analyzed in internode tissues using sugarcane RNA-Seq data. Gene expression data of the top (3) and bottom (8) internodes of the IACSP93-3046 and SP80-3280 varieties, as well as of the parental species *S. officinarum* (Badila de Java) and *S. spontaneum* (Krakatau), were obtained as described by Aono et al. (2021). Briefly, RNA-Seq reads were trimmed, and gene expression was quantified with Salmon (Patro et al., 2015) using the longest isoforms of *S. spontaneum* CDSs as a reference and automatic annotations by Zhang et al. (2018). A heatmap depicting the expression of all genes within the QTL was generated using the pheatmap R package (Kolde et al., 2012) in R software (R Core Team, 2011).

Differentially expressed genes (DEGs) were identified using the edgeR package version 3.38.4 (Robinson et al., 2010). The raw count data first underwent normalization using the counts per million (CPM) method. Genes with a CPM value ≥ 1 in all samples of at least one biological condition were retained. To identify DEGs, counts were subsequently normalized using the trimmed mean of M-values (TMM) method. Statistical comparisons were conducted between *S. spontaneum* samples and all other samples. DEGs were determined using a false discovery rate (FDR) threshold of $p \leq 0.05$ and a $\log_2$ fold change (FC) cutoff of ≥ 1.

Gene coexpression network analyses

To further investigate the biological processes associated with the genes within the QTL, a gene coexpression network was constructed with R software employing the highest reciprocal rank (HRR) methodology (Mutwill et al., 2010). Raw count data were normalized using the transcripts per million (TPM) method, and genes with a TPM > 0 in

all samples of at least one biological condition were retained. Pairwise Pearson R correlation coefficients were calculated for pairs of filtered genes. To ensure robust associations, a minimum absolute correlation coefficient threshold of 0.8 was used to consider two genes to be connected.

Results

QTL gene identification, BAC clone selection, sequencing, assembly and annotation

In the QTL for Brix in sorghum, 51 genes were identified, and seven genes in the expanded region were also identified; of these genes, 21 aligned with sugarcane leaf transcripts and presented exons with sizes equal to or greater than 200 bp. Primer pairs were developed for these 21 genes, and one pair failed to produce amplicons. The 20 primer pairs developed were used for screening clones of interest in the BAC libraries of the SP80-3280 and IACSP93-3046 varieties. Of the remaining 38 genes, 14 were found to be duplicated in the sorghum genome, and 24 did not meet the other selection criteria. For each gene, a number was assigned, except for two tandemly duplicated genes, which were given a single number (25), as shown in Supplementary Table 3, for a total of 57 genes.

In the screening of the BAC library of IACSP93-3046, 37 clones were positive for at least two genes, and 30 clones were sequenced. Among these contigs, 28 were assembled and manually annotated, representing 26 BACs. In the screening of the BAC library of SP80-3280, 56 clones were positive for at least two genes, and 31 clones were sequenced. Of these, genes from the region were found in 16 assembled contigs, and these were manually annotated, representing 16 BACs. The size of all contigs varied between 3,960 bp (pool 25) and 192,924 bp (pool 17), and the total length of the contigs was 5,850,46 bp (Supplementary Table 1).

From the 71 contigs that were generated, 43 carried the target region's genes. Each contig was related to a BAC, and some BACs were represented by two contigs (Supplementary Table 2).

The sorghum orthologous region was recovered in the variety IACSP93-3046 (Figure 2), which has 50 annotated genes. Seven sorghum genes were not found in the recovered sequence (Supplementary Figure 5). Between genes 35 and 36, there was a gap. In one of the haplotypes, one annotated gene did not belong to this region in sorghum,

although it is located in another region of sorghum chromosome SBI-2. In the SP80-3280 variety (Figure 1), the region was recovered almost in its entirety, with 44 annotated genes. Of the seven genes that were not found in the IACSP93-3046 contig sequences, six were not found in this variety, and one was annotated as a pseudogene. It is possible to observe two gaps, one between genes 24 and 26 and the other between genes 43 and 47.

There were 45 pseudogenes among the homo(e)logous genes in the variety IACSP93-3046 and nine probable pseudogenes in the variety SP80-3280. In the variety IACSP93-3046, genes with insertions of transposons in intronic regions (6-13.3%), insertions/deletions of one or more nucleotides (36–80%) and partial gene sequences (3-6.7%) were considered pseudogenes. Among nine homo(e)logous genes considered probable pseudogenes in SP80-3280, four (44.5%) exhibited an insertion/deletion of one or more nucleotides, three (33.5%) exhibited a transposon insertion in intronic regions, and in two (22.2%) of these genes, the pseudogene was a fragment of the gene.

Main differences in sorghum-sugarcane synteny and collinearity in the target region

Chromosome Sspon2A (Supplementary Figure 1)—The orthologous region on chromosome Sspon2A is 794,054 bp long and is the closest in size to chromosome SbI-02 of sorghum. It is located between bases 35,019,101 and 35,813,155. Among the 57 genes present in sorghum, 50 orthologs were found in Sspon2A. The seven missing orthologous genes (03, 14, 24, 40, 46, 55 and 56) were not detected throughout the chromosome and not only in the delimited region; they were not detected in the IACSP93-3046, SP80-3280 and R570 varieties. In the region delimited in Sspon2A, the gene Sspon.02G0013290 was found, and it is orthologous to a sorghum gene from chromosome Sb10 (Sobic.010G093001). Three additional genes were not detected in the IACSP93-3046 and R570 varieties, 31 and 43; these genes were detected on chromosome Sspon2A but as pseudogenes. Gene 51 was also detected as a pseudogene in the SP80-3280 variety but was the IACSP93-3046 and R570 varieties. Eight inversions were observed, two of which were common to the varieties IACSP93-3046 and SP80-3280, and they involved from two to eight genes. Duplications, some *in tandem*, were also observed. Therefore, there is synteny, as almost all the genes are present, but there are many breaks in collinearity.

Chromosome Sspon2B (Supplementary Figure 2)—On this chromosome, between the first and last genes of the studied region, there are 1,134,230 bp, more than double the region in sorghum, located between bases 32,312,667 and 33,446,897. In this chromosome, it was possible to observe many collinearity breaks, with inversions and an insertion within a cluster of 12 genes. In this orthologous region, it was possible to observe rearrangements and reorganizations, but most of the genes were present, guaranteeing synteny. Of the 57 genes present in sorghum, six were absent from the entire chromosome: 07, 26, 27 and 28. Genes 14 and 31 were also missing and were also not found in the IACSP93-3046, SP80-3280 and R570 varieties. An insertion with an eight-gene cluster, similar to a region immediately posterior to the one studied, in sorghum is present in this allele.

Chromosome Sspon2C (Supplementary Figure 3)—The chromosome Sspon2C region is the region that most resembles the sorghum chromosome SbI-02. Considering synteny, although between the first and the last gene, it is almost twice the size of the region, reaching 969,275 bp, and is located between bases 37,413,201 and 38,382,476. As in the other alleles, Sspon2C also has collinearity breaks with inversions and insertions, and there are gene sequences from the region that are displaced and inserted in other stretches. Among the 57 sorghum genes in the region, there are two that are absent on this chromosome, and these genes are also absent in the hybrid varieties IACSP93-3046, SP80-3280 and R570: genes 14 and 31. Although the region is quite large, compared to sorghum, there are no insertions with genes similar to those of other chromosomes in *S. spontaneum*.

Chromosome Sspon2D (Supplementary Figure 4)—This chromosome also maintains synteny with sorghum. Of the 57 genes in the region, 51 remained. Among the six missing genes, four were not detected in the IACSP93-3046, SP80-3280 or R570 varieties. The region was divided into two subregions. The first subregion is between bases 28,425,371 and 29,097,118 (671,747 bp), and the second is between bases 49,915,069 and 50,120,906 (205,837 bp). These two regions are approximately 21 Mb in length.

SP80-3280 variety (Figure 1)- Of the sequenced 31 clones, 16 were recognized as part of the target region using BLASTn. The orthologous region was partially recovered using region-belonging BACs. Among the recovered genes, synteny and collinearity may be presumed. Although there are breaks in collinearity with the variety IACSP93-3046, some of those observed are the same in both varieties, such as an inversion between genes

11 and 13 and another between genes 34 and 35. Out of 15 sequenced and annotated BACs, 11 were positive for the Sobic.002G223900 gene (gene 08), and of these, eight were also positive for one of the last 20 genes in the target region (genes 38 to 57). Some genes were not observed in the annotations of the varieties IACSP93-3046 and R570; this also occurred with the variety SP80-3280, except for one gene (51) that was observed as a possible pseudogene. Two pronounced gaps were detected: the absence of BACs containing genes 24 to 26 and the absence of BACs containing genes 43 to 47. The last gene flanking the region, 57, was also not recovered.

IACSP93-3046 (Figure 2)- This region was recovered with 29 annotated BACs. The synteny between sorghum and sugarcane in this specific region was confirmed, but some collinearity breaks were detected. Two inversions were observed, one between genes 11 and 13 and the other between genes 34 and 35. These inversions are observed in all haplotypes where these genes could be present. Gene 25, which is duplicated in sorghum, appeared in a single copy in the annotated haplotypes; on the other hand, gene 26 was duplicated in one of the three haplotypes observed. A sequence of three genes (53, 54 and 55) was found to be duplicated exactly in this sequence, resulting in a collinearity break; however, this finding appears in only one of the seven haplotypes that could have these genes. In one annotated BAC, an insertion of gene 57 between genes 52 and 53 was observed. Another interesting insertion was found between genes 16 and 17; it was a gene similar to Sobic.002G135950 from sorghum, and chromosome SbI-02 at positions 20,517,004-20,519,005, and Sobic.002G195033 from sorghum was located at sites Sb02 58,315,718-58,317,644; in other words, this gene was in another region but on the same chromosome.

R570 (Supplementary Figure 5)—Upon comparing the findings of Brazilian varieties with those of R570, some commonalities were observed. The inversion between genes 11 and 13 is present in all three hybrid cultivars, indicating that this observation is a characteristic of the *Saccharum* genus, as is also observed in *S. spontaneum*. Tandem duplications, such as that of gene 25, were noted, mirroring observations in sorghum. Interestingly, IACSP93-3046 lacks this duplication, and due to a gap in the sequencing of this region, this duplication could not be detected in SP80-3280. In R570, genes 48 and 49 are duplicated in tandem, a feature not observed in sorghum. Similar findings were not observed in *S. spontaneum* or in the varieties SP80-3280 and IACSP93-3046.

Expression analysis and search for candidate genes related to sugar accumulation: Investigation of selected genes

A summary of genes 01 to 57, their orthologs in *S. spontaneum* and *S. bicolor*, as well as their proteins, is provided in Supplementary Table 03. Based on this analysis, 10 candidate genes for sugar accumulation were selected: 02, 06, 09, 10, 15, 19, 20, 23, 25, and 43. These genes are possibly involved directly, indirectly, or in fundamental upstream steps involved in some phase of the process of sugar accumulation, which begins with carbon fixation from the atmosphere (photosynthesis), sucrose biosynthesis and transport to the stems, and subsequent accumulation (Supplementary Table 03).

DEG analyses

The expression of the genes within the QTL was evaluated using RNA-Seq data from internodes 3 (younger) and 8 (more mature) of IACSP93-3046 and SP80-3280, as well as data from accessions of the two species considered the main ancestors of modern cultivars, namely, *S. spontaneum* and *S. officinarum*. Seven of the 51 sorghum genes under analysis had no orthologs in the *S. spontaneum* genome, which was used for the gene quantification procedures; therefore, they are not represented in the expression data.

A heatmap (Figure 3) depicting the expression of the remaining 44 genes normalized by TPM is shown in Figure 3. This allowed us to observe internode gene expression patterns in two commercial sugarcane varieties, IACSP93-3046 and SP80-3280, and the parental species *S. officinarum* and *S. spontaneum*. For ten genes (10, 11, 12, 13, 15, 22, 25, 34, 45, 46, and 56), no expression was detected in any biological replicate, or there was minimal expression in up to three biological replicates. These genes may play crucial roles in other plant organs, such as leaves or roots, or could also be relevant in other stages of plant maturation. However, due to a lack of evidence of expression in the organ/tissue and maturation stages under analysis, these genes were not considered candidates for involvement in sugar production.

After filtering, 22,859 of the 35,471 genes present in the *S. spontaneum* CDSs were stably expressed under at least one biological condition and were thus retained for DEG analyses. By comparing varieties with high (*S. officinarum*, IACSP93-3046, and SP80-3280) and low (*S. spontaneum*) sugar contents, 6,264 DEGs were identified

(Supplementary Table 4). Seven genes within the QTL region were DEGs; their log_2(FC) values, FDR-corrected p values and annotations are available in Table 1.

Gene coexpression network analyses

Based on the expression data, an HRR coexpression network was constructed to explore new evidence that could contribute to the search for candidate genes. During filtering procedures, 7,565 genes were excluded, and the remaining 27,906 genes were used as input to construct the network. The final network had 6,809 connected nodes (genes) and an average of 17 neighbors per node. Among these genes, 3,397 genes were identified as DEGs, and six genes were identified within the QTL. A first neighbor search was employed to identify genes related to potential sugar accumulation candidates and to assess whether these genes could support their role in this process. The first neighbors of the genes within the QTL represented in the network can be seen in Table 2. Three of these genes—01, 23, and 26—were also identified as DEGs (Table 1).

Gene 01 (prolyl oligopeptidase—POP) exhibited relatively low expression in the stems of *S. spontaneum* and relatively high expression in samples from sugar-accumulating plants. Gene 23 (abscisic acid 8'-hydroxylase 3—ABA8'OH) has virtually no expression in the internodes of the sugar-accumulating plants sampled and is expressed at low levels in *S. spontaneum.* Gene 26 (ethylene responsive factor 109—ERF109) also has almost no expression in sugarcane plants, while it is expressed in *S. spontaneum*; however, in this case, there is significantly greater expression in the more mature internodes of *S. spontaneum* (I8) than in less mature internodes. This evidence led to the selection of genes 01, 23, and 26 as the primary candidate genes in the QTL for sugar accumulation.

Discussion

Main differences in genomic architecture

Synteny and collinearity have been used to compare and recover genomic regions of interest in sugarcane using sorghum (Ming et al., 1998; Paterson et al., 2009; Wang et

al., 2010; Figueira et al., 2012; Mancini et al., 2018; Garsmeur et al., 2018; Thirugnanasambandam et al., 2018; Zhang et al., 2018; Sforça et al., 2019; Aono et al., 2021; Federico et al., 2022 and Healey et al., 2024) and miscanthus (Mitros et al., 2020 and Zhang et al., 2021) genomes as references, revealing high gene retention (Mancini et al., 2018; Garsmeur et al., 2018; Sforça et al., 2019 and Feng et al., 2021). The comparison of the same region between sugarcane varieties and their ancestral species can provide insight into the genomic complexity of sugarcane. The region evaluated in this work showed substantial differences among the genotypes studied, such as gene duplications, loss of gene exons, pseudogenization, gene inversions, gene deletions and insertions.

For example, gene 25 (similar to alpha-amylase—AMY—Supplementary Table 3) is duplicated in tandem in sorghum and in R570, but in four IACSP93-3046 haplotypes, it is in a single copy (there is a gap in the region SP80-3280). The sequences of genes 53, 54, and 55 (Supplementary Table 3) were duplicated in tandem in BAC Shy141H03 of IACSP93-3046 (Figure 2), but they were not duplicated in sorghum, neither in the recovered SP80-3280 haplotypes nor in any of the alleles of the *S. spontaneum* genome. In addition to gene duplications, gene inversions were detected in the orthologous region between sorghum and all the *Saccharum* accessions evaluated (Figure 1), which suggests that both inversions occurred after sorghum–sugarcane divergence. Overall, when inversions do not significantly disrupt the gene balance of an organism, the direct consequences tend to be minimal. Documented cases exist where inversion results in pseudogenization or even deletion of one of the genes (Jurka et al., 2001; Zhao et al., 2016; Redd et al., 2023).

In BAC Shy411A07 of the IACSP93-3046 variety, gene 11 (Supplementary Table 3) was absent, yet the remaining genes (12 and 13—Supplementary Table 3) indicated that an inversion occurred. Fragments of Harbinger-type (HARB) repetitive elements were found near pseudogenes 12 and 13 (Supplementary Table 3). HARB transposons are classified as class II transposable elements (TEs) that carry out the cleavage and transfer of single DNA strands mediated by transposases (Zhao et al., 2016; Redd et al., 2023). The presence of these HARB transposons suggests a possible relationship between these elements and these inversions, which were present in all examined varieties, especially with the probable pseudogenization of genes 12 and 13 (Supplementary Table 3). The process of cleavage followed by fusion may have led to the deletion of bases, resulting in the truncation of genes and, consequently, the loss of their functions. In the

SP80-3280 variety, gene 13 (Supplementary Table 3) exhibited a single exon spanning 2,682 base pairs. On the other hand, gene 12 (Supplementary Table 3) maintains two introns, even in its pseudogenized state, and in this case, it is situated between two TEs, similar to the HARB type (BAC Shy260G24). In BAC Shy492F12, gene 12 (Supplementary Table 3) is also close to a HARB-type TE flanking the last exon. In this case, gene 12 (Supplementary Table 3) exhibited characteristics indicative of a functional gene. The gene had different CDS base pair compositions among the haplotypes but was always between 1347 bp and 1488 bp. Additionally, gene 11 (Supplementary Table 3) retained a single intron, with a length ranging between 1,200 and 1,209 bp. However, neither variation was detected in sorghum, suggesting that it might be a unique characteristic of the *Saccharum* genus. On chromosome Sspon2B of *S. spontaneum*, gene 11 (Supplementary Table 3) has a single exon.

In a haplotype of the variety IACSP93-3046, represented by BAC Shy112C03, a gene whose ortholog in sorghum is not found in the QTL studied was detected. Notably, this gene is similar to the two sorghum genes Sobic.002G135950 and Sobic.002G195033. These sorghum genes have 91% sequence identity, and both have a zinc finger domain. The probable orthologous gene in the IACSP93-3046 variety is inserted in a retrotransposon similar to Copia22-ZM_I/LTR. Interestingly, this gene exhibits all the characteristic features of being a functional gene, even though it is inserted in a TE. One possible explanation for this insertion is that the gene was cotransported with the retrotransposon. As Class II TEs, they can replicate a copy of themselves, which is subsequently inserted into different genomic regions. As such, there is a substantial likelihood that this haplotype is a copy of the gene. The presence of a TE within an expressed gene (CENP-C) in sugarcane has been previously described (Sforça et al., 2019), demonstrating that the proximity or overlap of TEs and genes does not hinder the function of the gene, at least in sugarcane. This gene, specific to the IACSP93-3046 haplotype (BAC Shy112C03), has a zinc finger domain; in plants, proteins featuring this domain are transcription factors (TFs) related to the control of cell division in totipotent tissues (petunias), histone-DNA binding (wheat), leaf budding (Chinese cabbage), soil salinity tolerance (Arabidopsis) and carbon metabolism (potato) (Takatsuji, 1999).

This gene has also been detected in the orthologous region of the *S. spontaneum* chromosome Sspon2B but as a gene fragment. In IACSP93-3046, the gene is located between genes 16 and 17 (Supplementary Table 3), in reverse orientation, and in

Sspon2B, it is located between genes 39 and 41 (Supplementary Table 3), with gene 40 (Supplementary Table 3) being inserted into another fragment of the orthologous region, in strand orientation. It is possible that this gene could also have been transposed with a TE, as possibly occurred with the hybrid. Importantly, *S. spontaneum* is a wild species that has been evolving under the pressure of natural selection, without the same level of human interference that fully domesticated plants undergo, as is the case with modern sugarcane cultivars—commercial hybrids. Despite such variability, we can observe the presence of potential genomic structure characteristics of *S. spontaneum* in commercial hybrids, such as inversions 11-13 (Supplementary Table 3) and 34-35 (Supplementary Table 3), which are present in at least three of the four alleles and are also present in all recovered haplotypes of the SP80-3280 and IACSP94-3046 varieties, where these inversions could be observed, as well as in the R570 variety (Garsmeur et al., 2018).

The orthologous regions in the hybrid varieties appear to be more similar to those in sorghum than to those in *S. spontaneum*. Some fundamental characteristics are shared, such as synteny. However, differences such as inversions, duplications, insertions of orthologous genes from the same genomic region and even from sequences that are similar to genes from sorghum chromosomes other than SBI-02, possible pseudogenization and translocation were detected. However, this finding is not surprising considering that the chromosomes originating from *S. spontaneum* found in hybrids constitute only 10% to 20% of the chromosomes of modern hybrids (D'Hont et al., 1996; Cuadrado, 2004; Piperidis et al., 2010; Garsmeur et al., 2018). Furthermore, wild species such as *S. spontaneum* have been subjected to natural selection pressure, resulting in a high level of expected heterozygosity for wild plants. While wild species have evolved naturally, commercial varieties have been selected and improved over the past 120 years to meet human needs (Singh et al., 2020), which has led to substantial genomic differences between them.

Investigation of genes involved in sugar accumulation

Of the 51 studied genes, 17 were annotated as genes associated with tolerance or response to stress (genes 01, 04, 11, 12, 13, 20, 22, 23, 24, 26, 32, 33, 34, 36, 37, 41, and 57—Supplementary Table 3). The period when sugarcane accumulates sucrose in its stems coincides with the dry and high luminosity period in a significant portion of the crops. During this period, leaves gradually fall, while sugar accumulates in culms (Garcia et al., 2019). Sucrose accumulation occurs in response to stressful conditions (Souza et

al., 2018). Therefore, there may be a connection between sucrose accumulation in sugarcane and genes related to abiotic stress.

Considering the characteristics of the crop, the significant number of genes related to the abscisic acid (ABA) response were also observed (genes 20, 23, 32, 33 and 37—Supplementary Table 3). In addition to being a crucial hormone for the photosynthetic process by regulating stomatal closure and opening, ABA is highly responsive to stress, particularly water stress (Chen et al., 2019). Leaf water potential and stomatal conductance are crucial factors for sugarcane to be able to produce carbohydrates that are converted into sucrose, transported to stalks, and subsequently accumulate (Smit et al., 2006 and Aluko et al., 2021).

In addition to these genes, five genes (genes 17, 18, 45, 46, and 47—Supplementary Table 3) belonging to the lipolytic enzyme GDXG family were detected. This enzyme family is characterized by having two consensus sequences containing a histidine residue and a serine residue as putative active site residues (van der Vlugt-Bergmans et al., 2001). In some genes encoding these enzymes, the presence of the alpha/beta hydrolase (ABH) domain may occur, which is known to play a role in catalyzing the cleavage of carbon double bonds and decarboxylation. Additionally, six genes (8, 49, 50, 52, 53, and 54—Supplementary Table 3) containing the ABH domain but belonging to the carboxylesterase (CXE) family were identified, five of which are sequential. The specificities of genes within the same family characterized by shared domains may vary significantly, necessitating further exploration of the biological roles of each gene. Notably, a group of genes sharing closely related domains and families on the same chromosome, even sequentially, may suggest a potential origin through duplication events that diverged during evolution into distinct genes while maintaining some similarity (Zhang et al., 2003).

Candidate genes for sugar accumulation

The ERF109 (gene 26—Supplementary Table 3), ABA 8' OH (gene 23—Supplementary Table 3), and POP (gene 01—Supplementary Table 3) genes are candidates for sucrose accumulation in sugarcane, considering that sugarcane needs soil with low humidity, approximately 15%, for greater sugar accumulation (FAO, 2024).

Gene 26 (Supplementary Table 3) is similar to ERF109, an ethylene-responsive TF. The expression of ERF109 is related to anthocyanin accumulation in apples, as

ERF109 directly binds to the promoters of anthocyanin synthesis genes (Ma et al., 2021). Jasmonic acid (JA) accumulation in plant wounds also activates the expression of ERF109. ERF109 induces the biosynthesis of the auxin protein ASA1, which aids in the process of secondary root formation mediated by JA-dependent ERF109 signaling (Guarneri et al., 2023). In sugarcane, more than 16,000 genes have been identified as potential targets regulated by ERF109, indicating that ERF109 has a broad influence on gene expression. Functions are diverse and include metabolic activities such as Rubisco activity, triggering hormone biosynthesis such as that of cytokinins, and gibberellin-mediated responses (Yu et al., 2024). ERF109 was not expressed in the internodes of any of the analyzed sugar-accumulating plants or in the younger internodes of *S. spontaneum*, but there was significant expression in its older internodes. In transgenic lemon, the overexpression of ERF109 causes global reprogramming of plant expression. ERF109 acts as a stress-responsive TF, but theorizing that this gene is a candidate gene for sugar accumulation requires further investigation.

Proline-dependent genes

Genes 01 and 23 (Supplementary Table 3) are related to proline accumulation. Proline is an amino acid with a unique configuration, restricting its free rotation at the α-carbon because the nitrogen and α-carbon are combined in a pyrrolidine ring. This structure contributes to the rigidity of proteins containing proline residues and requires specific enzymes, including POP, for cleavage (Dong et al., 2017). Enzymes that specifically cleave proline are known to be involved in proline accumulation in the cytosol of plant cells. This accumulation is essential for the plant's adaptive response to adverse situations (Ghiffari et al., 2022). Plants accumulate proline to maintain cellular homeostasis, aid in water absorption, and better adapt to abiotic stresses such as drought, salinity, and heavy metals. These adverse conditions lead to excess production of reactive oxygen species (ROS) and consequences such as lipid peroxidation, increased osmolyte levels, and activation of antioxidant systems (Rejeb et al., 2014; Kazemi-Shahandashti & Maali-Amiri, 2018; Sharma et al., 2019; Ghosh et al., 2021). As such, proline accumulation enhances adaptive responses in plants. Plants may increase proline biosynthesis in response to the above conditions or reuse presynthesized proline from proteins and peptides that are not essential (Ghifari et al., 2022).

The exogenous application of proline in maize has been reported to increase sugar, oil, moisture, and protein levels in seeds under drought conditions (Ali et al., 2013; Gosh et al., 2021). In sugarcane, the efficiency of photosynthesis and stomatal conductance is especially related to sucrose accumulation in stalks (Singels et al., 2021). In mature plants, the relationship between leaves as a sugar source and other organs, including internodes, is critical for the regulation of photosynthesis rates and sucrose accumulation in stalks (Souza et al., 2018). Proline interacts with other metabolites, including soluble sugars. The phenomenon of proline accumulation interacting with sucrose, for example, to adjust osmotic balance during salt stress, has been reported (Ghosh et al., 2019). However, it is unclear whether this interaction could be driven by other physiological changes in different plants with different stimuli. Although the connection between stress and plants has been clarified, there is still much to be elucidated. In sugarcane, the accumulation of sucrose and starch in leaves coincides with a reduction in photosynthetic rates, which occurs during low water availability (Garcia et al., 2019).

Gene 23 (Supplementary Table 3), which shares similarities with ABA 8'-hydroxylase enzymes from the cytochrome-P450 family, converts ABA into 8'-hydroxy ABA and then into phaseic acid (Kronchko et al., 1998), regulating ABA metabolism and influencing plant responses to environmental stress and development, including germination, root growth, and fruit maturation (Wang et al., 2023). Inhibition of this enzyme affects the balance of processes involving ABA (Wang et al., 2023), such as stomatal closure in response to water, salt, and thermal stresses. Studies in grapes have shown that inhibiting ABA 8'-hydroxylase results in reduced leaf water potential and stomatal conductance (Tomiyama et al., 2020), accompanied by proline accumulation in leaves and the growth of adventitious roots (Tomiyama et al., 2020).

In sugar-accumulating plants, ABA 8'-hydroxylase is expressed at low levels, suggesting a potential role for ABA regulation in sugar storage tissues (Figure 3— gene 13240). The dry climate during the sugar accumulation period in sugarcane, as observed during sample collection, indicates potential moderate water stress. Under water stress conditions, the sugarcane genotypes with the most efficient sugar accumulation tend to maintain greater stomatal conductance (Sajid et al., 2023). The lack of expression of the ABA catabolism gene suggested that the need for stomatal conductance regulation in these plants may be linked to maintaining open stomata. Additionally, in grapes, inhibition of the ABA 8'-hydroxylase gene improved tolerance to dehydration and

promoted adventitious root formation, demonstrating an effective strategy for coping with water stress. Gene 23 was not expressed in the internodes (gene 13240—Figure 3), suggesting that the plant may use this strategy to improve its tolerance to potential water deficits.

Gene 01 (Supplementary Table 3) shares similarities with POP, which belongs to the serine protease family (clan SC, family S9) that includes various peptidases. POP is a cytoplasmic enzyme that hydrolyzes peptide bonds at the C-terminal side of proline residues (Gutierrez et al., 2008; Baharin et al., 2022). The enzyme's three-dimensional structure allows for the postproline cleavage of peptides containing up to 30 amino acid residues (Gutierrez et al., 2008; Baharin et al., 2022). Post- or preproline cleavage enzymes can belong to different peptidase families, including aminopeptidases, endopeptidases, or oligopeptidases (PAP/PEP/POP). The most common domain in family S9 is a substrate-limiting β-propeller domain preventing unwanted digestion, while the α/β hydrolase domain catalyzes the reaction at the carboxy terminus of proline residues (Baharin et al., 2022). POP is a ubiquitous protein with a well-established structure and mechanism of action. However, its biological role in plants has not been fully elucidated. Increased expression in plants is known to be associated with tolerance to various types of abiotic stresses (Gutierrez et al., 2008; Singh et al., 2011; Tan et al., 2013). In flax, this phenomenon seems to be related to a fundamental mechanism for embryo growth in seeds (Gutierrez et al., 2008). In coffee, POP overexpression is linked to a significant increase in the number of branches in transgenic plants (Singh et al., 2011). Other peptidases that hydrolyze with proline specificity are related to plant development, such as pollen development (Ghifari et al., 2022), flowering, increased ABA activity, protection of photosynthetic activity during salt stress, elimination of reactive oxygen species, and overall osmotic potential adjustment (Ghosh et al., 2021).

Gene 01 is a DEG (Table 1) that is more highly expressed in the internodes of sugar-accumulating sampled plants and significantly less expressed in *S. spontaneum*, a sugar nonaccumulating sugarcane species known for its resistance to various stress types. Although POP is related to stress resistance and tolerance, there is more evidence suggesting that this gene is involved in this process. It has been observed that proline cleavage enzymes occur when a plant needs to accumulate proline, a phenomenon that usually occurs when the plant requires osmotic regulation due to stresses such as water

and saline stress. It is also known that proline can bind to soluble sugars such as sucrose when there is a need to regulate the homeostasis of plant cells. Gene 01 has three first neighbors, one of which possesses the ENTH domain—a lipid-binding region crucial for clathrin-coated vesicle formation, endocytosis at the trans-Golgi network (TGN), and vacuolar transport. This gene could play a role in the transport of proline and sucrose, given its correlation with POP.

While sugarcane cannot accumulate sucrose under severe stress, previous studies have shown that mild water deficiency enhances photosynthetic rates and the accumulation of starch and sucrose in leaves (Garcia et al., 2018). The mechanism of proline accumulation is related to plants facing challenging situations, and the ability of proline to bind to sucrose adds another layer of complexity. Therefore, considering the paramount importance of comprehending sucrose accumulation processes in sugarcane and its connection with water deficit events during the period of peak sugar accumulation in the stems, a thorough examination of the role of POP is crucial. This includes exploring its potential association with proline accumulation and understanding how this accumulation might impact sugar storage.

Identifying genes that control or influence agronomic traits is one of the objectives of molecular breeding. Sugarcane, however, lags behind sorghum in terms of available genetic and genomic information. This study proposes a novel approach for transferring genetic knowledge from sorghum (donor) to sugarcane (recipient). Building upon existing methods (Shiringani et al., 2010; Garsmeur et al., 2011; Mancini et al., 2018; Sforça et al., 2019; Aono et al., 2021), we integrated genomic and coexpression network analysis to validate the relevance of sorghum-derived information in sugarcane. Furthermore, we analyzed the same genomic region in two Brazilian cultivars, revealing their differential genomic architecture and potential impact on sugar accumulation, using expression information to validate the results.

References

Aitken, K. S., Jackson, P. A., and McIntyre, C. L. (2006). Quantitative trait loci identified for sugar related traits in a sugarcane (Saccharum spp.) cultivar × Saccharum officinarum population. Theor Appl Genet 112, 1306–1317. doi:10.1007/s00122-006-0233-2.

Ali, Q.; Anwar, F.; Ashraf, M.; Saari, N.; Perveen, R. Ameliorating Effects of Exogenously Applied Proline on Seed Composition, Seed Oil Quality and Oil

Antioxidant Activity of Maize (Zea mays L.) under Drought Stress. Int. J. Mol. Sci. 2013, 14, 818-835. https://doi.org/10.3390/ijms14010818

AltschuP, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. 8.

Aluko OO, Li C, Wang Q, Liu H. Sucrose Utilization for Improved Crop Yields: A Review Article. International Journal of Molecular Sciences. 2021; 22(9):4704. https://doi.org/10.3390/ijms22094704

Alvarez María E., Arnould Savouré, László Szabados, Proline metabolism as regulatory hub, Trends in Plant Science, Volume 27, Issue 1, 2022, Pages 39-55, ISSN 1360-1385, https://doi.org/10.1016/j.tplants.2021.07.009.

Aono AH, Pimenta RJG, Garcia ALB, Correr FH, Hosaka GK, Carrasco MM, Cardoso-Silva CB, Mancini MC, Sforça DA, Dos Santos LB, Nagai JS, Pinto LR, Landell MGA, Carneiro MS, Balsalobre TW, Quiles MG, Pereira WA, Margarido GRA, de Souza AP. The Wild Sugarcane and Sorghum Kinomes: Insights Into Expansion, Diversification, and Expression Patterns. Front Plant Sci. 2021 Jul 7;12:668623. doi: 10.3389/fpls.2021.668623. PMID: 34305969; PMCID: PMC8294386.

Babu, K. S. D., Janakiraman, V., Palaniswamy, H., Kasirajan, L., Gomathi, R., & Ramkumar, T. R. (2022). A short review on sugarcane: its domestication, molecular manipulations, and future perspectives. *Genetic Resources and Crop Evolution*, *69*, 2623-2643.

Badet, T., Voisin, D., Mbengue, M., Barascud, M., Sucher, J., Sadon, P., Balagué, C., Roby, D., & Raffaele, S. (2017). Parallel evolution of the POQR prolyl oligo peptidase gene conferring plant quantitative disease resistance. *PLoS Genetics*, *13*(12), e1007143. https://doi.org/10.1371/journal.pgen.1007143.

Baharin, A., Ting, T. Y., & Goh, H. H. (2022). Post-Proline Cleaving Enzymes (PPCEs): Classification, Structure, Molecular Properties, and Applications. *Plants (Basel)*, *11*(10), 1330. https://doi.org/10.3390/plants11101330

Bozzi, A. T., & Gaudet, R. (2021). Molecular Mechanism of Nramp-Family Transition Metal Transport. *Journal of Molecular Biology*, *433*(16), 166991. https://doi.org/10.1016/j.jmb.2021.166991

Bremer, G. (1961). Problems in breeding and cytology of sugar cane. *Euphytica*, *10*, 59-78.

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., et al. (2014). De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. *PLoS ONE*, *9*, e88462. https://doi.org/10.1371/journal.pone.0088462

Carson, D.L., and F.C. Botha. 2002. Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Reports* 20: 1075–1081.

Casu, R.E., C.P.L. Grof, A.L. Rae, C.L. McIntyre, C.M. Dimmock, and J.M. Manners. 2003. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Molecular Biology* 52: 371–386.

Chandra, A., Jain, R., & Solomon, S. (2012). Complexities of invertases controlling sucrose accumulation and retention in sugarcane. *Current Science*, *102*(6), 857–866. http://www.jstor.org/stable/24084500

Chen K, Li GJ, Bressan RA, Song CP, Zhu JK, Zhao Y. Abscisic acid dynamics, signaling, and functions in plants. J Integr Plant Biol. 2020 Jan;62(1):25-54. doi: 10.1111/jipb.12899. PMID: 31850654.

Costa, E. A., Anoni, C. O., Mancini, M. C., Santos, F. R. C., Marconi, T. G., Gazaffi, R., et al. (2016). QTL mapping including codominant SNP markers with ploidy

level information in a sugarcane progeny. Euphytica 211, 1–16. doi:10.1007/s10681-016-1746-7.

Cuadrado, A. (2004). Genome remodelling in three modern S. officinarumxS. spontaneum sugarcane cultivars. Journal of Experimental Botany 55, 847–854. doi:10.1093/jxb/erh093.

CURSI, D. E. et al. History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. Sugar Tech, p. 1–22, 2021.

De Souza AP, Grandis A, Arenque-Musa BC, Buckeridge MS. Diurnal variation in gas exchange and nonstructural carbohydrates throughout sugarcane development. Funct Plant Biol. 2018 Jul;45(8):865-876. doi: 10.1071/FP17268. PMID: 32291068.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J. C., Rao, S., and Berding, N. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Molec. Gen. Genet. 250, 405–413. doi:10.1007/BF02174028.

Diniz AL, da Silva DIR, Lembke CG, Costa MDL, ten-Caten F, Li F, Vilela RD, Menossi M, Ware D, Endres L, Souza GM. Amino acid and carbohydrate metabolismo are coordinated to maintain balance during drought in sugarcane (2020) Int. J. Mol. Sci. doi.org/10.3390/ijms21239124

Dong Z, Shuangshuang Yang, Zhengtian Zhang, Cunduo Tang, Yunchao Kan, Lunguang Yao; Prolyl aminopeptidases: Reclassification, properties, production and industrial applications, Process Biochemistry, Volume 118, 2022, Pages 121-132, ISSN 1359-5113, https://doi.org/10.1016/j.procbio.2022.04.025.

Federico, M. L., Diniz, A. L., Souza, G. M., Snowdon, R., & Erazzú, L. (2022). Translational genomics from sorghum (Sorghum bicolor) to sugarcane (Saccharum spp.) for bioenergy breeding. *Plant Breeding*, *141*(3), 389-398.

Figueira, T. R., Okura, V., Rodrigues da Silva, F., Jose da Silva, M., Kudrna, D., Ammiraju, J. S., et al. (2012). A BAC library of the SP80-3280 sugarcane variety (saccharum sp.) and its inferred microsynteny with the sorghum genome. BMC Res Notes 5, 185. doi:10.1186/1756-0500-5-185.

Gallaher TJ, Peterson PM, Soreng RJ, Zuloaga FO, Li D, Clark LG, Tyrrel CD, Welker CAD, Kellog EA, Teisher JK (2022). Grasses through space and time: An overview of the biogeographical and macroevolutionary history of Poaceae. Journal of Systematics and Evolution. Vol60,Issue3, Pages 522-569, doi.org/10.1111/jse.12857

Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., et al. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep 3, 3399. doi:10.1038/srep03399.

Garcia F.H.S., Mendonça A. M. das C., Rodrigues M., Matias F. I., Filho M. P. da S., Santos H. R. B., Taffner J.and Barbosa J. P. R. A. D; Water deficit tolerance in sugarcane is dependent on the accumulation of sugar in the leaf. Annals of Applied, Biology, Vol 176, Issue 1, January 2020, https://doi.org/10.1111/aab.12559

Garsmeur O., Charron C., Bocs S., Jouffe V., Samain S., Couloux A., et al.. (2011). High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. New Phytol. 189, 629–642. 10.1111/j.1469-8137.2010.03497.x

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun 9, 2638. doi:10.1038/s41467-018-05051-5.

Ghifari, A. S., Huang, S., Murcha, M. W. (2019). The peptidases involved in plant mitochondrial protein import. J. Exp. Bot. 70, 6005–6018. doi: 10.1093/jxb/erz365

Ghifari AS, Teixeira PF, Kmiec B, Singh N, Glaser E, Murcha MW. The dual-targeted prolyl aminopeptidase PAP1 is involved in proline accumulation in response to

stress and during pollen development. J Exp Bot. 2022 Jan 5;73(1):78-93. doi: 10.1093/jxb/erab397. PMID: 34460901.

Ghosh UK, Islam MN, Siddiqui MN, Cao X, Khan MAR. Proline, a multifaceted signalling molecule in plant responses to abiotic stress: understanding the physiological mechanisms. Plant Biol (Stuttg). 2022 Mar;24(2):227-239. doi: 10.1111/plb.13363. Epub 2021 Nov 18. PMID: 34796604.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40, D1178–D1186. doi:10.1093/nar/gkr944.

Grivet, L., and P. Arruda. 2002. Sugarcane genomics: Depicting the complex genome of an important tropical crop. *Current Opinion in Plant Biology* 5: 122–127.

Guarneri N, Willig JJ, Sterken MG, Zhou W, Hasan MS, Sharon L, Grundler FMW, Willemsen V, Goverse A, Smant G, Lozano-Torres JL. Root architecture plasticity in response to endoparasitic cyst nematodes is mediated by damage signaling. New Phytol. 2023 Feb;237(3):807-822. doi: 10.1111/nph.18570. Epub 2022 Dec 1. PMID: 36285401; PMCID: PMC10108316.

Gutierrez L, Castelain M, Verdeil JL, Conejero G, Van Wuytswinkel O. A possible role of prolyl oligopeptidase during Linum usitatissimum (flax) seed development. Plant Biol (Stuttg). 2008 May;10(3):398-402. doi: 10.1111/j.1438-8677.2008.00038.x. PMID: 18426487.

Healey, A. L., Garsmeur, O., Lovell, J. T., Shengquiang, S., Sreedasyam, A., Jenkins, J., ... D'Hont, A. (2024). The complex polyploid genome architecture of sugarcane. *Nature*. https://doi.org/10.1038/s41586-024-0535-6

International Rice Genome Sequencing Project, and Sasaki, T. (2005). The map-based sequence of the rice genome. Nature 436, 793–800. doi:10.1038/nature03895.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. 4.

Kazemi-Shahandashti K, Reza Maali-Amiri; Global insights of protein responses to cold stress in plants: Signaling, defence, and degradation, Journal of Plant Physiology, Volume 226, 2018, Pages 123-135, ISSN 0176-1617, https://doi.org/10.1016/j.jplph.2018.03.022.

Kim, C., Wang, X., Lee, T.-H., Jakob, K., Lee, G.-J., and Paterson, A. H. (2014). Comparative Analysis of Miscanthus and Saccharum Reveals a Shared Whole-Genome Duplication but Different Evolutionary Fates. Plant Cell 26, 2420–2429. doi:10.1105/tpc.114.125583.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915. doi:10.1038/s41587-019-0201-4.

Kavi Kishor PB, Suravajhala P, Rathnagiri P, Sreenivasulu N. Intriguing Role of Proline in Redox Potential Conferring High Temperature Stress Tolerance. Front Plant Sci. 2022 Jun 10;13:867531. doi: 10.3389/fpls.2022.867531. PMID: 35795343; PMCID: PMC9252438.

Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 474. doi:10.1186/1471-2105-7-474.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version, 1(2), 726.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Res. 27, 722–736. doi:10.1101/gr.215087.116.

Krochko JE, Abrams GD, Loewen MK, Abrams SR, Cutler AJ. (+)-Abscisic acid 8'-hydroxylase is a cytochrome P450 monooxygenase. Plant Physiol. 1998 Nov;118(3):849-60. doi: 10.1104/pp.118.3.849. PMID: 9808729; PMCID: PMC34795.

Ma H, Yang T, Li Y, Zhang J, Wu T, Song T, Yao Y, Tian J. The long noncoding RNA MdLNC499 bridges MdWRKY1 and MdERF109 function to regulate early-stage light-induced anthocyanin accumulation in apple fruit. Plant Cell. 2021 Oct 11;33(10):3309-3330. doi: 10.1093/plcell/koab188. PMID: 34270784; PMCID: PMC8505877.

Mancini, M. C., Cardoso-Silva, C. B., Sforça, D. A., and Pereira de Souza, A. (2018). "Targeted Sequencing by Gene Synteny," a New Strategy for Polyploid Species: Sequencing and Physical Structure of a Complex Sugarcane Region. Front. Plant Sci. 9, 397. doi:10.3389/fpls.2018.00397.

Mancini, M. C., Leite, D. C., Perecin, D., Bidóia, M. A. P., Xavier, M. A., Landell, M. G. A., et al. (2012). Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. Sugar Tech 14, 119–125. doi:10.1007/s12355-012-0141-5.

Marin, F. R.; Variedades- EMBRAPA 2022. Available in https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/cana/pre-producao/caracteristicas/variedades. Access in march/2024.

Mattiello, L., Riaño-Pachón, D. M., Martins, M. C. M., da Cruz, L. P., Bassi, D., Marchiori, P. E. R., et al. (2015). Physiological and transcriptional analyses of developmental stages along sugarcane leaf. BMC Plant Biol 15, 300. doi:10.1186/s12870-015-0694-z.

Meyer, R., Purugganan, M. Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet 14, 840–852 (2013). https://doi.org/10.1038/nrg3605

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, Mattison A, Morishige DT, Grimwood J, Schmutz J, Mullet JE, The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization., The Plant journal : for cell and molecular biology. 2017 Nov 21

Miao, J., Feng, Q., Li, Y. *et al.* Chromosome-scale assembly and analysis of biomass crop *Miscanthus lutarioriparius* genome. *Nat Commun* 12, 2458 (2021). https://doi.org/10.1038/s41467-021-22738-4

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH. Detailed alignment of saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics. 1998 Dec;150(4):1663-82. doi: 10.1093/genetics/150.4.1663. PMID: 9832541; PMCID: PMC1460436.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mitros, T., Session, A.M., James, B.T. et al. Genome biology of the paleotetraploid perennial biomass crop Miscanthus. Nat Commun 11, 5442 (2020). https://doi.org/10.1038/s41467-020-18923-6

Moloi SJ, Ngara R. The roles of plant proteases and protease inhibitors in drought response: a review. Front Plant Sci. 2023 Apr 18;14:1165845. doi: 10.3389/fpls.2023.1165845. PMID: 37143877; PMCID: PMC10151539.

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C.-K. K., et al. (2017). The pangenome of hexaploid bread wheat. Plant J 90, 1007–1013. doi:10.1111/tpj.13515.

Mutwil M, Usadel B, Schütte M, Loraine A, Ebenhöh O, Persson S. 2009. Assembly of an interactive correlation network for the *Arabidopsis* genome using a novel heuristic clustering algorithm. Plant Physiology 152:29-43. doi: 10.1104/pp.109.145318.

Nishiyama, M. Y., Ferreira, S. S., Tang, P.-Z., Becker, S., Pörtner-Taliana, A., and Souza, G. M. (2014). Full-Length Enriched cDNA Libraries and ORFeome Analysis of Sugarcane Hybrid and Ancestor Genotypes. PLoS ONE 9, e107351. doi:10.1371/journal.pone.0107351.

Ovens K, Eames BF and McQuillan I (2021) Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution. Front. Genet. 12:695399. doi: 10.3389/fgene.2021.695399

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556. doi:10.1038/nature07723.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417-419. doi:10.1038/nmeth.4197

Peng R, Zhang B. Foxtail Millet: A New Model for C4 Plants. Trends Plant Sci. 2021 Mar;26(3):199-201. doi: 10.1016/j.tplants.2020.12.003. Epub 2020 Dec 21. PMID: 33358112.

Pereira-Santana, A., Alvarado-Robledo, E. J., Zamora-Briseño, J. A., Ayala-Sumuano, J. T., Gonzalez-Mendoza, V. M., Espadas-Gil, F., et al. (2017). Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. PLoS ONE 12, e0189271. doi:10.1371/journal.pone.0189271.

Phytozome 13 - Miscanthus sinensis v7.1 DOE-JGI, http://Pi.jgi.doe.gov/

Piperidis, G., Piperidis, N., and D'Hont, A. (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Genet Genomics 284, 65–73. doi:10.1007/s00438-010-0546-3.

R Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Redd PS, Payero L, Gilbert DM, Page CA, King R, McAssey EV, Bodie D, Diaz S, Hancock N (2023). Transposase expression, element abundance, element size, and DNA repair determine the mobility and heritability of PIF/Pong/Harbinger transposable elements. Frontiers in Cell and Developmental Biology, 10.3389/fcell.2023.1184046.

Rejeb, K B, Chedly Abdelly, Arnould Savouré, How reactive oxygen species and proline face stress together, Plant Physiology and Biochemistry, Volume 80, 2014, Pages 278-284, ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2014.04.007.

Riaño-Pachón, D.M. and Mattiello, L. (2017) Draft genome sequencing of the sugarcane hybrid SP80-3280 [version 2; peer view: 2 approved]. F1000 Research, 6:861

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. bioinformatics, 26(1), 139-140.

Rowland, E., Kim, J., Wijk, K. J., Van Friso, G., Poliakov, A., Ponnala, L. (2022). The CLP and PREP protease systems coordinate maturation and degradation of the chloroplast proteome in Arabidopsis thaliana. New Phytol. 236, 1339–1357. doi: 10.1111/nph.18426

Ruan, J., Dean, A.K. & Zhang, W. A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC Syst Biol 4, 8 (2010). https://doi.org/10.1186/1752-0509-4-8

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al. (2000). Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945. doi:10.1093/bioinformatics/16.10.944.

Sajid, M., Amjid, M., Munir, H., Ahmad, M., Zulfiqar, U., Ali, M. F., Farah, M. A., Ahmed, M. A. A., & Artyszak, A. (2023). Comparative Analysis of Growth and Physiological Responses of Sugarcane Elite Genotypes to Water Stress and Sandy Loam Soils. Plants, 12(15), 2759. https://doi.org/10.3390/plants12152759

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022 Jan 7;50(D1):D20-D26. doi: 10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326, 1112–1115. doi:10.1126/science.1178534.

Sforça, D. A., Vautrin, S., Cardoso-Silva, C. B., Mancini, M. C., Romero-da Cruz, M. V., Pereira, G. da S., et al. (2019). Gene Duplication in the Sugarcane Genome: A Case Study of Allele Interactions and Evolutionary Patterns in Two Genic Regions. Front. Plant Sci. 10, 553. doi:10.3389/fpls.2019.00553.

Sharma P, Priyanka Sharma, Priya Arora, Vinod Verma, Kanika Khanna, Poonam Saini, Renu Bhardwaj, Chapter 8 - Role and Regulation of ROS and Antioxidants as Signaling Molecules in Response to Abiotic Stresses, Editor(s): M. Iqbal R. Khan, Palakolanu Sudhakar Reddy, Antonio Ferrante, Nafees A. Khan, Plant Signaling Molecules, Woodhead Publishing, 2019, Pages 141-156, ISBN 9780128164518, https://doi.org/10.1016/B978-0-12-816451-8.00008-3.

Sheng, J., Zheng, X., Wang, J., Zeng, X., Zhou, F., Jin, S., et al. (2017). Transcriptomics and proteomics reveal genetic and biological basis of superior biomass crop Miscanthus. Sci Rep 7, 13777. doi:10.1038/s41598-017-14151-z.

Shiringani, A. L., Frisch, M., and Friedt, W. (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. Theor Appl Genet 121, 323–336. doi:10.1007/s00122-010-1312-y.

Singels A, Phillip Jackson, Geoff Inman-Bamber; Chapter 21 - Sugarcane, Editor(s): Victor O. Sadras, Daniel F. Calderini, Crop Physiology Case Histories for Major Crops, Academic Press, 2021, Pages 674-713, ISBN 9780128191941, https://doi.org/10.1016/B978-0-12-819194-1.00021-9.

Singh, A. (2020). Benefits of crop diversification in Fiji's sugarcane farming. *Asia & the Pacific Policy Studies*, 7(1), 65-80. https://doi.org/10.1002/app5.291

Singh, R., Irikura, B., Nagai, C. et al. Characterization of Prolyl Oligopeptidase Genes Differentially Expressed Between Two Cultivars of Coffea arabica L.. Tropical Plant Biol. 4, 203–216 (2011). https://doi.org/10.1007/s12042-011-9082-5

Smit M.A. , A. Singels,The response of sugarcane canopy development to water stress, Field Crops Research, Volume 98, Issues 2–3, 2006, Pages 91-97, ISSN 0378-4290, https://doi.org/10.1016/j.fcr.2005.12.009.

Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. Nat Rev Genet 20, 631–656 (2019). https://doi.org/10.1038/s41576-019-0150-2

Takatsuji, H – Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. Plant Molecular Biology, 1999

Tan CM, Chen RJ, Zhang JH, Gao XL, Li LH, Wang PR, Deng XJ, Xu ZJ. OsPOP5, a prolyl oligopeptidase family gene from rice confers abiotic stress tolerance in Escherichia coli. Int J Mol Sci. 2013 Oct 10;14(10):20204-19. doi: 10.3390/ijms141020204. PMID: 24152437; PMCID: PMC3821611.

Taraszkiewicz A, Sinkiewicz I, Sommer A, Staroszczyk H. The biological role of prolyl oligopeptidase and the procognitive potential of its peptidic inhibitors from food proteins. Crit Rev Food Sci Nutr. 2023 Feb 16:1-14. doi: 10.1080/10408398.2023.2170973. Epub ahead of print. PMID: 36798052.

The International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191. doi:10.1126/science.aar7191.

The UniProt Consortium , UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, Volume 51, Issue D1, 6 January 2023, Pages D523–D531, https://doi.org/10.1093/nar/gkac1052

Thirugnanasambandam, P. P., Hoang, N. V., and Henry, R. J. (2018). The Challenge of Analyzing the Sugarcane Genome. Front. Plant Sci. 9, 616. doi:10.3389/fpls.2018.00616.

Thirugnanasambandam, P. P., Mason, P. J., Hoang, N. V., Furtado, A., Botha, F. C., and Henry, R. J. (2019a). Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. BMC Plant Biol 19, 160. doi:10.1186/s12870-019-1733-y.

Thirugnanasambandam, P. P., Mason, P. J., Hoang, N. V., Furtado, A., Botha, F. C., and Henry, R. J. (2019b). Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. BMC Plant Biol 19, 160. doi:10.1186/s12870-019-1733-y.

Tomiyama, H., Sato, M., Opio, P., et al. (2020). Inhibition of Abscisic Acid 8′-Hydroxylase Affects Dehydration Tolerance and Root Formation in Cuttings of Grapes (*Vitis labrusca* L. × *Vitis vinifera* L. cv. Kyoho) Under Drought Stress Conditions. *Journal of Plant Growth Regulation*, 39, 1577–1586. https://doi.org/10.1007/s00344-020-10171-8

Tsuruta, S., Ebina, M., Kobayashi, M., and Takahashi, W. (2017). Complete Chloroplast Genomes of Erianthus arundinaceus and Miscanthus sinensis: Comparative Genomics and Evolution of the Saccharum Complex. PLoS ONE 12, e0169992. doi:10.1371/journal.pone.0169992.

van der Vlugt-Bergmans CJ, van der Werf MJ. Genetic and biochemical characterization of a novel monoterpene epsilon-lactone hydrolase from Rhodococcus erythropolis DCL14. Appl Environ Microbiol. 2001 Feb;67(2):733-41. doi: 10.1128/AEM.67.2.733-741.2001. PMID: 11157238; PMCID: PMC92642.

van Wijk, K. J. (2015). Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. Annu. Rev. Plant Biol. 66, 75–111. doi: 10.1146/annurev-arplant-043014-115547

Vasantha, S., Kumar, R.A., Tayade, A.S. *et al.* Physiology of Sucrose Productivity and Implications of Ripeners in Sugarcane. *Sugar Tech* 24, 715–731 (2022). doi.org/10.1007/s12355-021-01062-7

Vettore, A. L. (2003). Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. Genome Research 13, 2725–2735. doi:10.1101/gr.1532103.

Wang, J., Roe, B., Macmil, S., Yu, Q., Murray, J. E., Tang, H., et al. (2010). MReseiacrcrhoarctioclellinearity between autopolyploid sugarcane and diploid sorghum genomes. 17.

Wang, T., Fang, J. & Zhang, J. Advances in Sugarcane Genomics and Genetics. Sugar Tech 24, 354–368 (2022). https://doi.org/10.1007/s12355-021-01065-4

Wang, X., Jin, B., Yan, W., Wang, J., Xu, J., Cai, C., Qi, X., Xu, Q., Yang, X., Xu, X., & Chen, X. (2023). Cucumber abscisic acid 8′-hydroxylase Csyf2 regulates yellow flesh by modulating carotenoid biosynthesis. *Plant Physiology*, *193*(2), 1001–1015. https://doi.org/10.1093/plphys/kiad383

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research 35, W265–W268. doi:10.1093/nar/gkm286.

Yu G, Sun B, Zhu Z, Mehareb EM, Teng A, Han J, Zhang H, Liu J, Liu X, Raza G, Zhang B, Zhang Y, Wang K. Genome-wide DNase I-hypersensitive site assay reveals distinct genomic distributions and functional features of open chromatin in autopolyploid sugarcane. Plant J. 2024 Jan;117(2):573-589. doi: 10.1111/tpj.16513. Epub 2023 Oct 27. PMID: 37897092.

Zhang G, Ge C, Xu P, Wang S, Cheng S, Han Y, Wang Y, Zhuang Y, Hou X, Yu T, Xu X, Deng S, Li Q, Yang Y, Yin X, Wang W, Liu W, Zheng C, Sun X, Wang Z, Ming R, Dong S, Ma J, Zhang X, Chen C. The reference genome of Miscanthus floridulus illuminates the evolution of Saccharinae. Nat Plants. 2021 May;7(5):608-618. doi: 10.1038/s41477-021-00908-y. Epub 2021 May 6. Erratum in: Nat Plants. 2021 Jul;7(7):991. Erratum in: Nat Plants. 2021 Jul;7(7):990. PMID: 33958777; PMCID: PMC8238680.

Zhang J, Evolution by gene duplication: an update, Trends in Ecology & Evolution, Volume 18, Issue 6, 2003, Pages 292-298, ISSN 0169-5347, https://doi.org/10.1016/S0169-5347(03)00033-8.

Zhang, J., Nagai, C., Yu, Q., Pan, Y.-B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three Saccharum species. Euphytica 185, 511–519. doi:10.1007/s10681-012-0664-6.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nat Genet 50, 1565–1573. doi:10.1038/s41588-018-0237-2.

Zhao D, Fergunson AA, Jiang N. (2016) What makes up plante genomes: The vanishing line between transposable element and genes. Biochimica et Biophysica Acta 1859, 366–380.

Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat. 2010 Mar;20(2):281-300. doi: 10.1080/10543400903572753. PMID: 20309759.

Zhu, Y.J., H.H. Albert, and P.H. Moore. 2000. Differential expression of soluble acid invertase genes in the shoots of high-sucrose and low-sucrose species of *Saccharum* and their hybrids. *Australian Journal of Plant Physiology* 27: 193–199.

Zoltan S and Laszlo P, Structure, function and biological relevance of Prolyl Oligopeptidase; Current Protein and Peptide Science, Vol.9, Number 1, 2008, pp.96-107 (12)

Figure 1

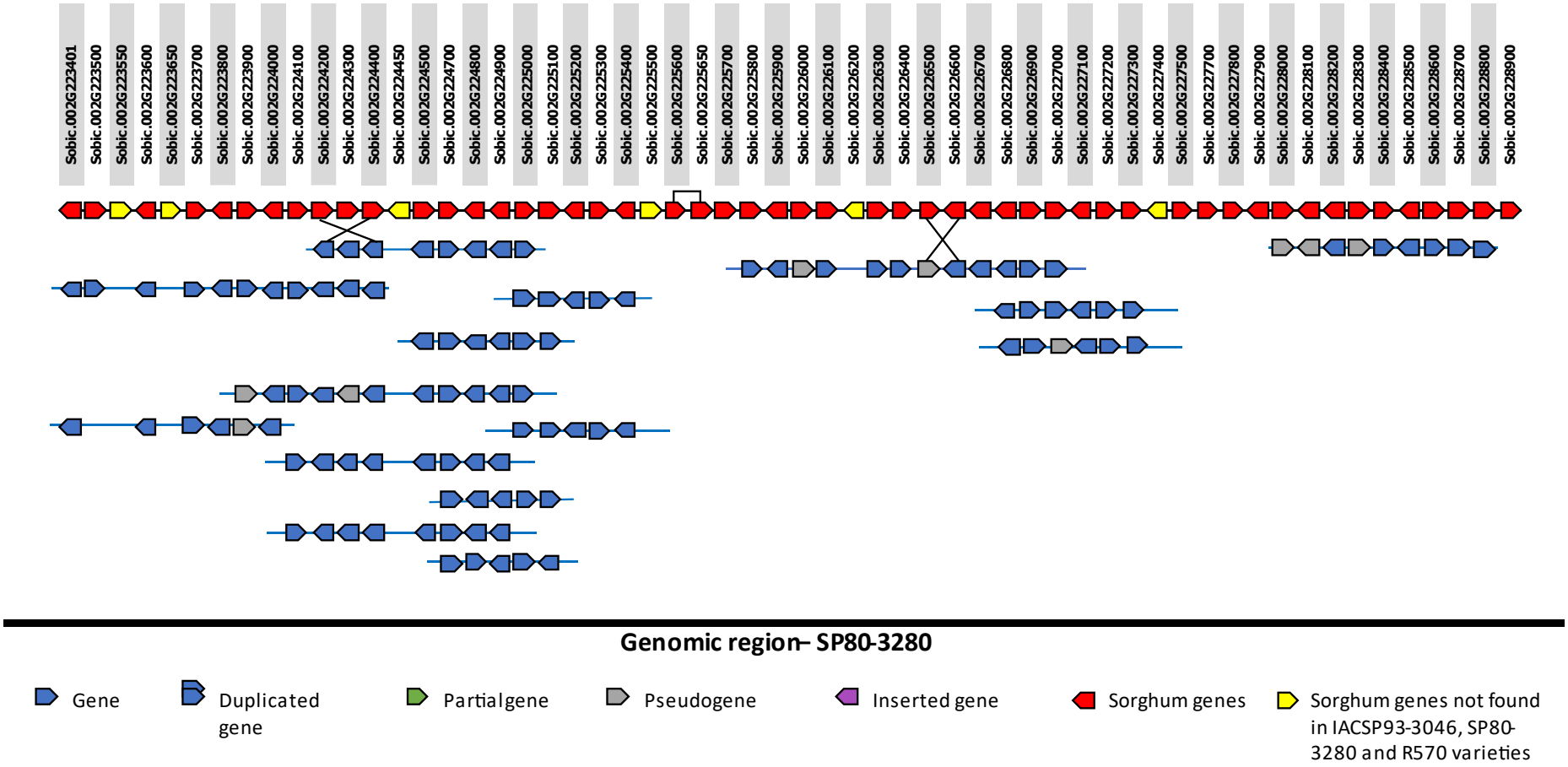

Sobic.002G223401
Sobic.002G223500
Sobic.002G223550
Sobic.002G223600
Sobic.002G223650
Sobic.002G223700
Sobic.002G223800
Sobic.002G223900
Sobic.002G224000
Sobic.002G224100
Sobic.002G224200
Sobic.002G224300
Sobic.002G224400
Sobic.002G224450
Sobic.002G224500
Sobic.002G224700
Sobic.002G224800
Sobic.002G224900
Sobic.002G225000
Sobic.002G225100
Sobic.002G225200
Sobic.002G225300
Sobic.002G225400
Sobic.002G225500
Sobic.002G225600
Sobic.002G225650
Sobic.002G225700
Sobic.002G225800
Sobic.002G225900
Sobic.002G226000
Sobic.002G226100
Sobic.002G226200
Sobic.002G226300
Sobic.002G226400
Sobic.002G226500
Sobic.002G226600
Sobic.002G226700
Sobic.002G226800
Sobic.002G226900
Sobic.002G227000
Sobic.002G227100
Sobic.002G227200
Sobic.002G227300
Sobic.002G227400
Sobic.002G227500
Sobic.002G227700
Sobic.002G227800
Sobic.002G227900
Sobic.002G228000
Sobic.002G228100
Sobic.002G228200
Sobic.002G228300
Sobic.002G228400
Sobic.002G228500
Sobic.002G228600
Sobic.002G228700
Sobic.002G228800
Sobic.002G228900
Genomic region– SP80-3280
Gene
Duplicated gene
Partialgene
Pseudogene
Inserted gene
Sorghum genes
Sorghum genes not found in IACSP93-3046, SP80-3280 and R570 varieties

Figure 2

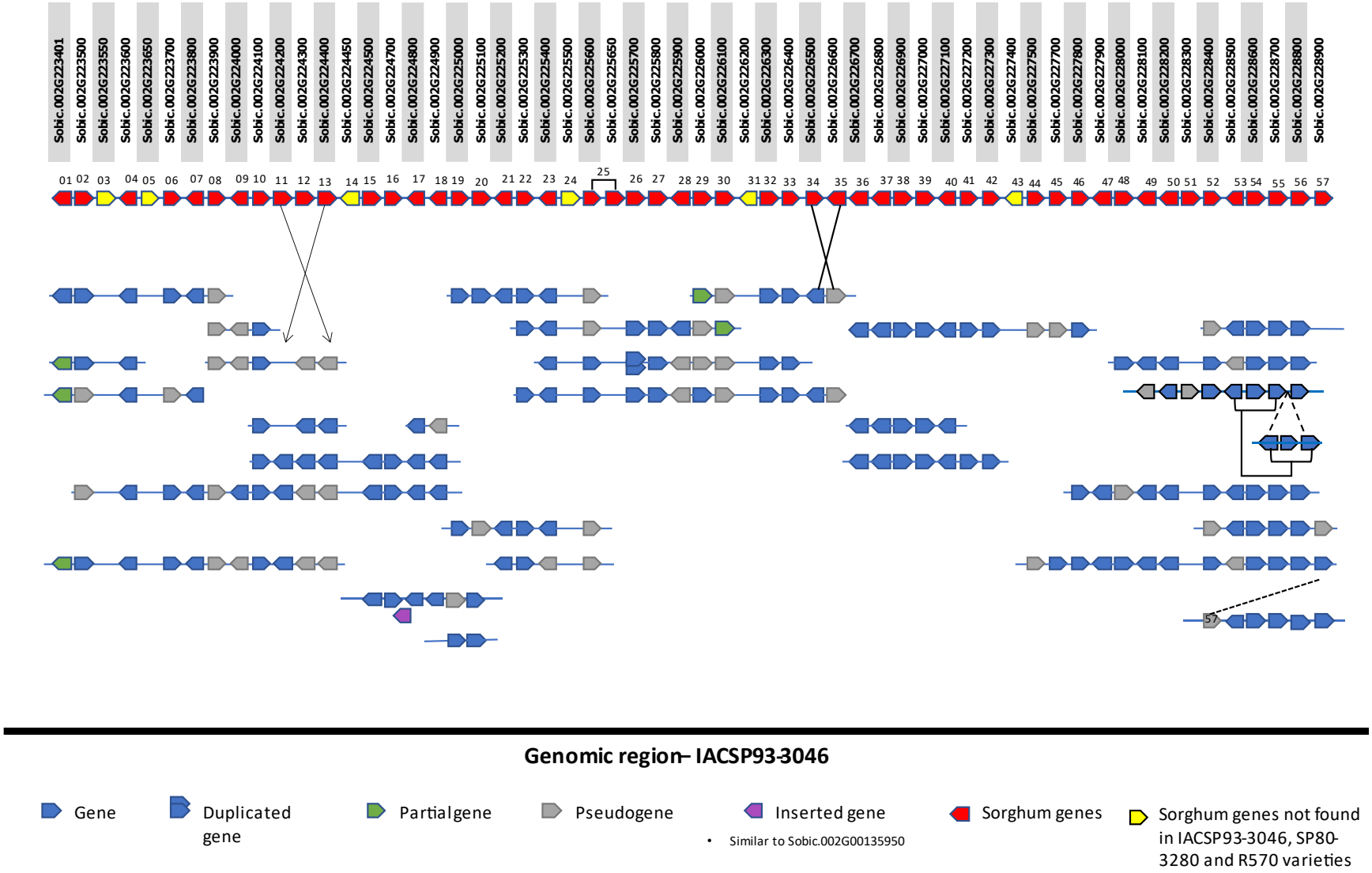

Sobic.002G223401
Sobic.002G223500
Sobic.002G223550
Sobic.002G223600
Sobic.002G223650
Sobic.002G223700
Sobic.002G223800
Sobic.002G223900
Sobic.002G224000
Sobic.002G224100
Sobic.002G224200
Sobic.002G224300
Sobic.002G224400
Sobic.002G224450
Sobic.002G224500
Sobic.002G224700
Sobic.002G224800
Sobic.002G224900
Sobic.002G225000
Sobic.002G225100
Sobic.002G225200
Sobic.002G225300
Sobic.002G225400
Sobic.002G225500
Sobic.002G225600
Sobic.002G225650
Sobic.002G225700
Sobic.002G225800
Sobic.002G225900
Sobic.002G226000
Sobic.002G226100
Sobic.002G226200
Sobic.002G226300
Sobic.002G226400
Sobic.002G226500
Sobic.002G226600
Sobic.002G226700
Sobic.002G226800
Sobic.002G226900
Sobic.002G227000
Sobic.002G227100
Sobic.002G227200
Sobic.002G227300
Sobic.002G227400
Sobic.002G227500
Sobic.002G227700
Sobic.002G227800
Sobic.002G227900
Sobic.002G228000
Sobic.002G228100
Sobic.002G228200
Sobic.002G228300
Sobic.002G228400
Sobic.002G228500
Sobic.002G228600
Sobic.002G228700
Sobic.002G228800
Sobic.002G228900
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
57
Genomic region– IACSP93-3046
Gene
Duplicated gene
Partialgene
Pseudogene
Inserted gene
• Similar to Sobic.002G00135950
Sorghum genes
Sorghum genes not found in IACSP93-3046, SP80-3280 and R570 varieties

Figure 3

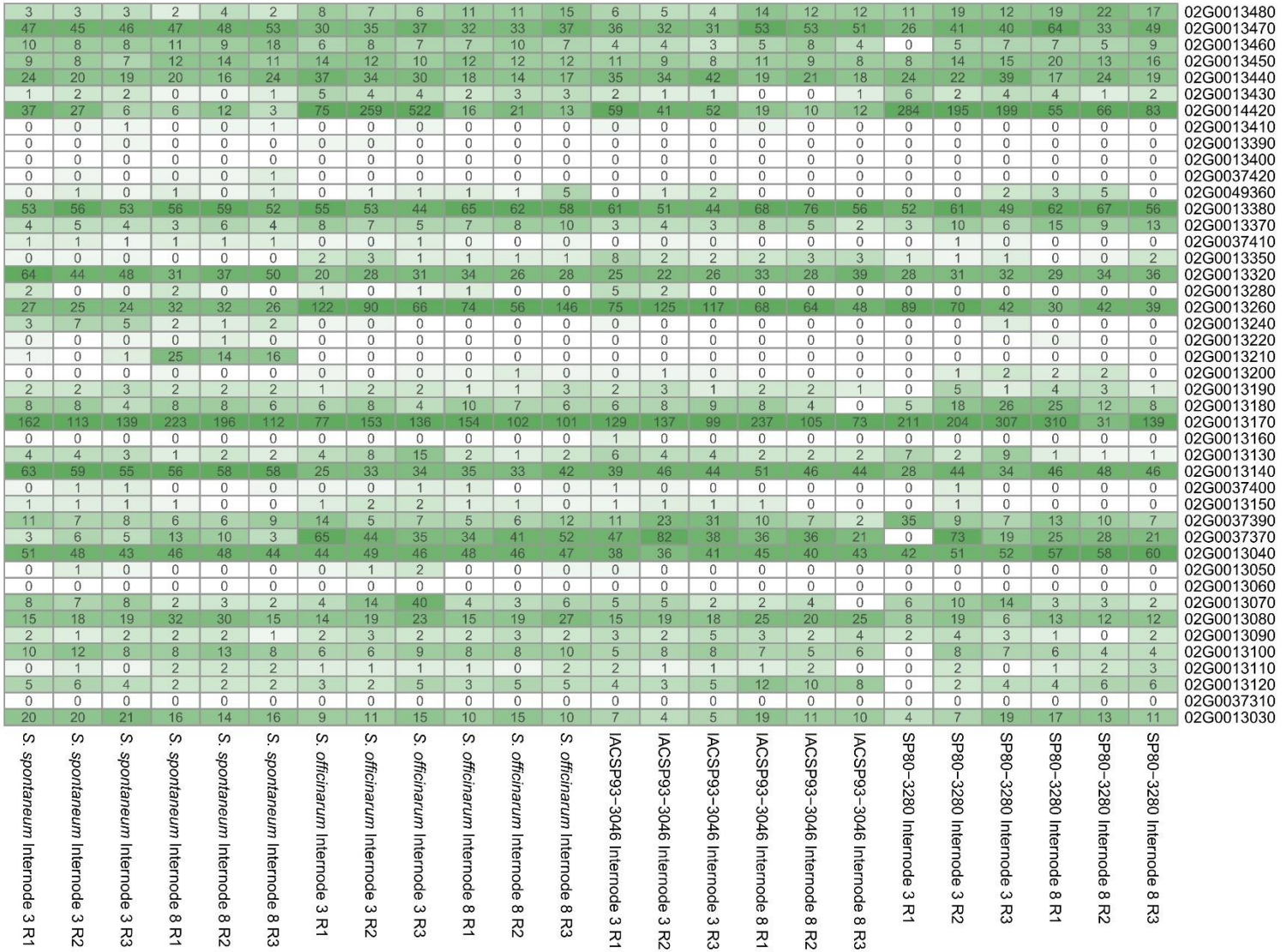

S. spontaneum Internode 3 R1
S. spontaneum Internode 3 R2
S. spontaneum Internode 3 R3
S. spontaneum Internode 8 R1
S. spontaneum Internode 8 R2
S. spontaneum Internode 8 R3
S. officinarum Internode 3 R1
S. officinarum Internode 3 R2
S. officinarum Internode 3 R3
S. officinarum Internode 8 R1
S. officinarum Internode 8 R2
S. officinarum Internode 8 R3
IACSP93−3046 Internode 3 R1
IACSP93−3046 Internode 3 R2
IACSP93−3046 Internode 3 R3
IACSP93−3046 Internode 8 R1
IACSP93−3046 Internode 8 R2
IACSP93−3046 Internode 8 R3
SP80−3280 Internode 3 R1
SP80−3280 Internode 3 R2
SP80−3280 Internode 3 R3
SP80−3280 Internode 8 R1
SP80−3280 Internode 8 R2
SP80−3280 Internode 8 R3
3 3 3 2 4 2 8 7 6 11 11 15 6 5 4 14 12 12 11 19 12 19 22 17 02G0013480
47 45 46 47 48 53 30 35 37 32 33 37 36 32 31 53 53 51 26 41 40 64 33 49 02G0013470
10 8 8 11 9 18 6 8 7 7 10 7 4 4 3 5 8 4 0 5 7 7 5 9 02G0013460
9 8 7 12 14 11 14 12 10 12 12 12 11 9 8 11 9 8 8 14 15 20 13 16 02G0013450
24 20 19 20 16 24 37 34 30 18 14 17 35 34 42 19 21 18 24 22 39 17 24 19 02G0013440
1 2 2 0 0 1 5 4 4 2 3 3 2 1 1 0 0 1 6 2 4 4 1 2 02G0013430
37 27 6 6 12 3 75 259 522 16 21 13 59 41 52 19 10 12 284 195 199 55 66 83 02G0014420
0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013410
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013390
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013400
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0037420
0 1 0 1 0 1 0 1 1 1 1 5 0 1 2 0 0 0 0 0 2 3 5 0 02G0049360
53 56 53 56 59 52 55 53 44 65 62 58 61 51 44 68 76 56 52 61 49 62 67 56 02G0013380
4 5 4 3 6 4 8 7 5 7 8 10 3 4 3 8 5 2 3 10 6 15 9 13 02G0013370
1 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 02G0037410
0 0 0 0 0 0 2 3 1 1 1 1 8 2 2 2 3 3 1 1 1 0 0 2 02G0013350
64 44 48 31 37 50 20 28 31 34 26 28 25 22 26 33 28 39 28 31 32 29 34 36 02G0013320
2 0 0 2 0 0 1 0 1 1 0 0 5 2 0 0 0 0 0 0 0 0 0 0 02G0013280
27 25 24 32 32 26 122 90 66 74 56 146 75 125 117 68 64 48 89 70 42 30 42 39 02G0013260
3 7 5 2 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 02G0013240
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013220
1 0 1 25 14 16 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013210
0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 2 2 2 0 02G0013200
2 2 3 2 2 2 1 2 2 1 1 3 2 3 1 2 2 1 0 5 1 4 3 1 02G0013190
8 8 4 8 8 6 6 8 4 10 7 6 6 8 9 8 4 0 5 18 26 25 12 8 02G0013180
162 113 139 223 196 112 77 153 136 154 102 101 129 137 99 237 105 73 211 204 307 310 31 139 02G0013170
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 02G0013160
4 4 3 1 2 2 4 8 15 2 1 2 6 4 4 2 2 2 7 2 9 1 1 1 02G0013130
63 59 55 56 58 58 25 33 34 35 33 42 39 46 44 51 46 44 28 44 34 46 48 46 02G0013140
0 1 1 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 02G0037400
1 1 1 1 0 0 1 2 2 1 1 0 1 1 1 1 0 0 0 1 0 0 0 0 02G0013150
11 7 8 6 6 9 14 5 7 5 6 12 11 23 31 10 7 2 35 9 7 13 10 7 02G0037390
3 6 5 13 10 3 65 44 35 34 41 52 47 82 38 36 36 21 0 73 19 25 28 21 02G0037370
51 48 43 46 48 44 44 49 46 48 46 47 38 36 41 45 40 43 42 51 52 57 58 60 02G0013040
0 1 0 0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013050
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0013060
8 7 8 2 3 2 4 14 40 4 3 6 5 5 2 2 4 0 6 10 14 3 3 2 02G0013070
15 18 19 32 30 15 14 19 23 15 19 27 15 19 18 25 20 25 8 19 6 13 12 12 02G0013080
2 1 2 2 2 1 2 3 2 2 3 2 3 2 5 3 2 4 2 4 3 1 0 2 02G0013090
10 12 8 8 13 8 6 6 9 8 8 10 5 8 8 7 5 6 0 8 7 6 4 4 02G0013100
0 1 0 2 2 2 1 1 1 1 0 2 2 1 1 1 2 0 0 2 0 1 2 3 02G0013110
5 6 4 2 2 2 3 2 5 3 5 5 4 3 5 12 10 8 0 2 4 4 6 6 02G0013120
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 02G0037310
20 20 21 16 14 16 9 11 15 10 15 10 7 4 5 19 11 10 4 7 19 17 13 11 02G0013030

Figure Legends

Figure 1: Genomic region of sorghum×Genomic region of SP80-3280. Each square represents a gene and shows the direction of the gene in the genome, where right-facing arrows indicate the forward direction, and left-facing arrows indicate the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. The genes in red are orthologous sorghum genes, while genes in yellow lack orthologs in the hybrid varieties. For each gene, a number has been designated, and above each number, the name allocated to it in Phytozome v.13 is indicated. Each solid line below the genomic region of sorghum represents a BAC. Genes shown with solid lines represent those successfully recovered within a BAC; together, the BACs reconstruct the genomic region of SP80-3280.

Figure 2: Sorghum×IACSP93-3046 genomic region. Each square represents a gene and shows the direction of the gene in the genome, where right-facing arrows indicate the forward direction, and left-facing arrows indicate the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. The genes in red are orthologous sorghum genes, while genes in yellow lack orthologs in the hybrid varieties. For each gene, a number has been designated, and above each one, the name allocated to it in Phytozome v.13 is indicated. Each solid line below the genomic region of sorghum represents a BAC. Genes shown with solid lines represent those successfully recovered within a BAC; together, the BACs reconstruct the genomic region of IACSP93-3046. The gene depicted in purple is an exclusive finding within this BAC. This gene is not one of the 57 sorghum QTL genes; however, it is similar to the sorghum gene Sobic.002G00135950 (Phytozome v.13). A pseudogene orthologous to this gene was observed on chromosome Sspon.2B (Supplementary Figure 2).

Figure 3: Heatmap representing the level of gene expression in the internodes of SP80-3280, IACSP93-3046, *S. officinarum* and *S. spontaneum.* A heatmap depicting gene expression levels across tissues, internodes 3 (top) and 8 (bottom) of sugarcane plants (varieties SP80-3280 and IACSP93-3046; *S. officinarum* and *S. spontaneum*) in triplicate. The darker the shade of green is, the higher the expression level.

Tables

Table 1: Differentially expressed genes (DEGs) located within the QTL under analysis.

Gene number	log2(FC)	P value	*S. spontaneum gene*/Protein and Description
01	2.049208548	1.5015E-06	**Sspon.02G0013480—**Similar to prolyl oligopeptidase (POP), an enzyme that cleaves oligopeptides up to 30 amino acid residues postproline (Gutierrez et al., 2008).
09	2.826212214	0.003517002	**Sspon.02G0014420—**The conversion of CO2 and water to bicarbonate and the release of a proton is catalyzed by carbonic anhydrase (CA), the first enzymatic step of photosynthesis in C4 plants. This reaction takes place in the mesophyll cells. Bicarbonate will initiate the first carboxylation of C4 (DiMario et al., 2022).
19	3.472221834	0.006550842	**Sspon.02G0013350—**They are membrane proteins, they are part of the so-called lipid rafts (Raft proteins—protein groups of membrane proteins that resemble a boat on the lipid group) (Rafaelle et al., 2012). Remorines may be associated with the regulation and translocation of photoassimilates. A specific type, GSD1, belonging to group 6 of remorines, has the function of regulating the conductance of photoassimilates through plasmodesmata in rice (Gui et al., 2015).
22	1.306514099	0.00043504	**Sspon.02G0013260—**B-type HSF are classically transcriptional repressor proteins (Ikeda et al., 2011). In grape HSFB1 has expression induced by heat stress. In rice, HSFB1 expression was related to cold. The gene has increased transcription when the plant is undergoing abiotic stresses, however, they act mainly as transcriptional repressors (Fragkostefanakis et al., 2018 and Chen et al., 2023).
23	-5.246110673	0.009820214	**Sspon.02G0013240—**ABA 8'-hydroxylase (CYP707A) is an enzyme involved in the catabolism of the hormone ABA (abscisic acid). ABA is inactive when it has a hydroxyl group (OH) at its 8' position, while its removal makes it active. ABA is involved in the closure of stomata, which are crucial structures in photosynthetic processes as they regulate gas exchange between the environment and the plant (Ng et al., 2014)."
26	-17.55155671	5.38665E-05	**Sspon.02G0013210—**Studies suggest that ERF109 is a positive regulator of cold tolerance (Wang et al., 2018). Other classic abiotic stresses also regulate ERF109 (Bahieldin et al., 2018). The gene, in joint action, also regulates defense against some pathogens (Zhao et al., 2022). The gene is present in the biosynthesis of tryptophan, in the metabolic pathway of auxin production (Zhang et al., 2019).

42	2.410548623	0.005106978	**Sspon.02G0037370—**Uncharacterized protein—predicted (Phytozome v13)

Table 2: Sugarcane genes and their first neighbors. The first neighbor genes are defined by their names in *S. spontaneum*, and their descriptions are based on the annotations of the EMBL database.

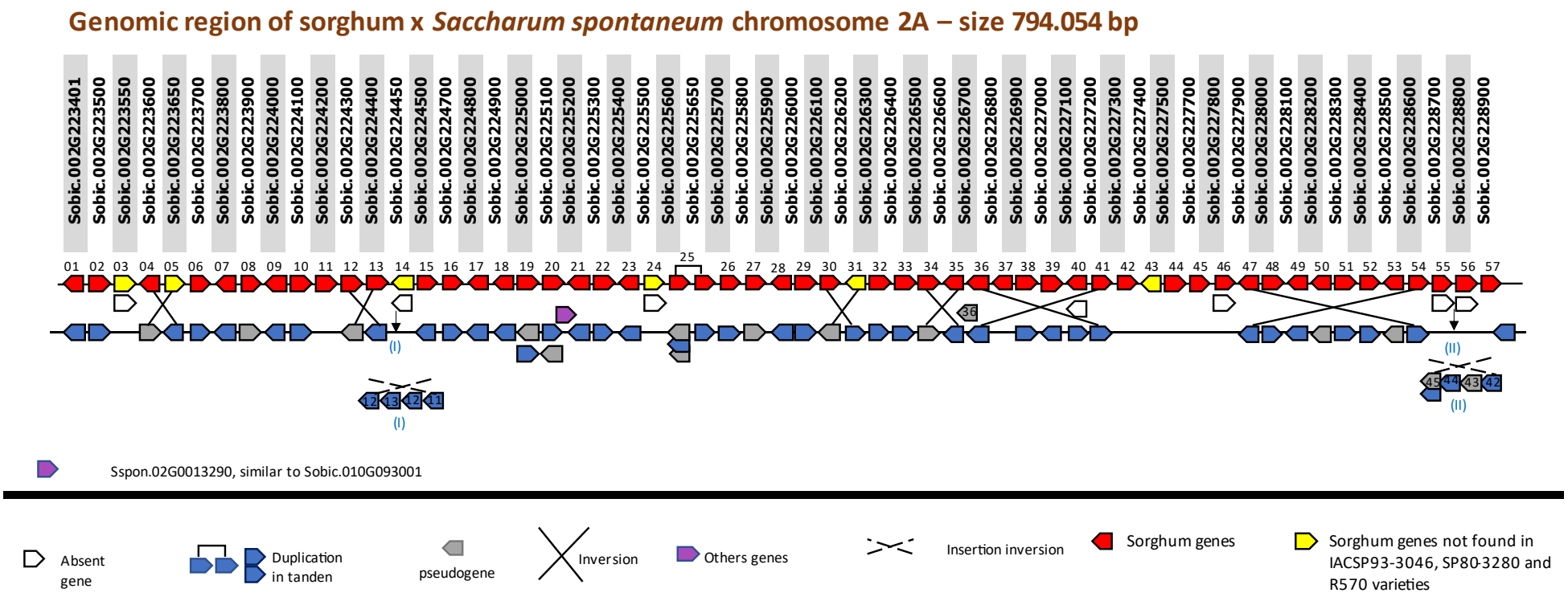
Gene Number	Protein (Phytozome v.13)	First Neighbors/Description
01	**POP** -Similar to prolyl oligopeptidase family	**Sspon.01G0011190—**The gene shares similarity with the IQ motif, known for its interaction with calmodulins in plant cells. This interaction influences the function of the target protein, particularly in cytoskeletal processes and cellular development, with a primary role in regulating signaling alongside CaM, CML, and CAMTA proteins (Teresinski et al., 2023).
		Sspon.01G0011270—No translating CDS (EMBL). The sequence has 73.7% identity with a protein containing the RRM domain from Miscanthus lutarioparius (Phytozome v.13).
		Sspon.01G0011300—This gene is equipped with a ENTH domain responsible for lipid binding, pivotal in the creation of clathrin-coated vesicles, trans-Golgi network (TGN) endocytosis, and vacuolar transport, contributing to the plant's immune response. Proteins within the ANTH/ENTH/VHS family display functional redundancy, likely shaped by natural selection, necessitating mutations in at least two genes to significantly affect plant function (Feng et al., 2022).
07	Similar to a membrane protein, possibly structural. (GO:0016020)	**Sspon.01G0010650—**similar to MYB46, a transcription factor, activates genes for cellulose, hemicellulose, and lignin synthesis, interacting with other transcription factors. In apples, MYB46 overexpression enhances salt tolerance, stress response, and promotes secondary cell wall biosynthesis, including lignin deposition by binding directly to relevant gene promoters. Activated during plant stress, MYB46 regulates genes involved in both

		biotic and abiotic stress responses (Chen et al., 2019).
18	Protein belonging to the **GDXG** family. Similar to alpha/beta hydrolase (**ABH**).	**Sspon.01G0019580—**No translating CDS, The intronic sequence of this gene is 100% identical to the Sobic.001G217300 gene in sorghum. However, in sorghum, this gene consists of three exons, whereas in *S. spontaneum,* it contains seven exons, with the additional four exons being the first four in the sequence (Phytozome v13).
		Sspon.01G0019780—Segments of the gene exhibit similarities to various segments on sorghum chromosome 1, encompassing both small gene fragments and intergenic regions (Phytozome v13).
23	**AA8' OH**- Similar to Abscisic acid 8'-hydroxylase	**Sspon.01G0005990—**This gene shares 94.1% similarity with *Miscanthus lutarioriparius*' s-acetyltransferase, identified as a palmitoyltransferase (PAT16). Acetyltransferases, influencing protein modification in plants, transfer acetyl groups from Acetyl-CoA. Palmitoyltransferases specifically add a 16-carbon palmitate to proteins, crucial for membrane protein function by anchoring to the cell membrane (Jiang et al., 2021).
26	**ERF109**- Similar to Ethylene Responsive factor 109	**Sspon.01G0005860—**NRAMP genes in plants are key players in selectively absorbing and transporting essential transition metals during heavy metal stress. Regulated by phytohormones, these genes maintain metal balance. A research with potatoes unveils molecular insights for potential development of low-metal-accumulating plant varieties. In essence, NRAMP genes are vital for plant resilience to heavy metal stress (Bozzi et al. 2021).

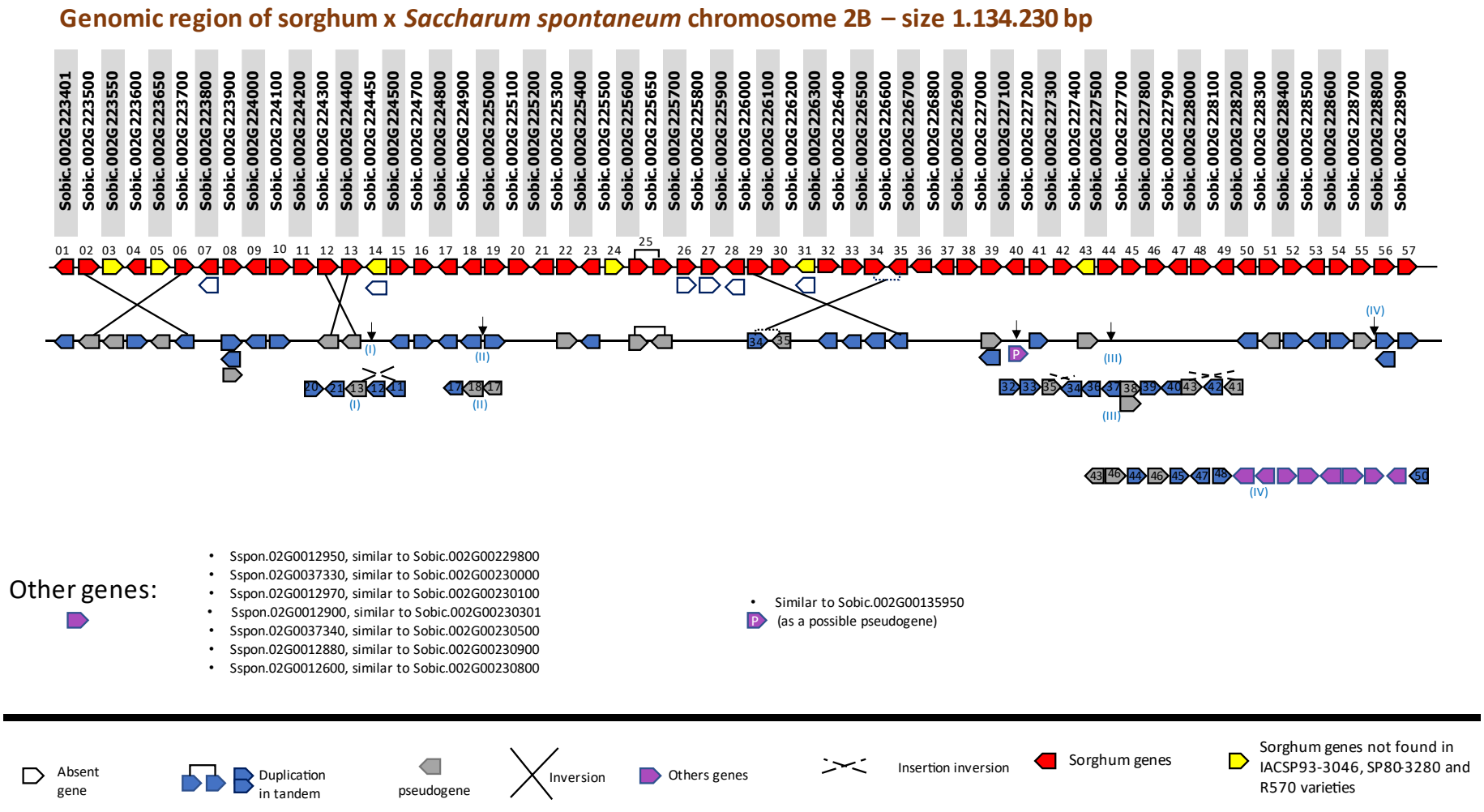
37	**PP2C**- Similar to protein phosphatase 2C	**Sspon.01G0005530—**In S. spontaneum, there are two isoforms of this gene, one encoding NPG1 (NO pollen germination 1) and a nontranslating CDS. The latter is annotated with two additional exons compared to its sorghum counterpart. This isoform features the tetratricopeptide repeat (TPR) domain. The TPR domain is explored in a study with tomatoes, influencing cell regulation, gene expression, and stress responses. The same study with tomatoes suggests the TPR gene has a potential link to energy metabolism and acts as a mediator in disease resistance (Zhou et al., 2021).

Supplementary Figures

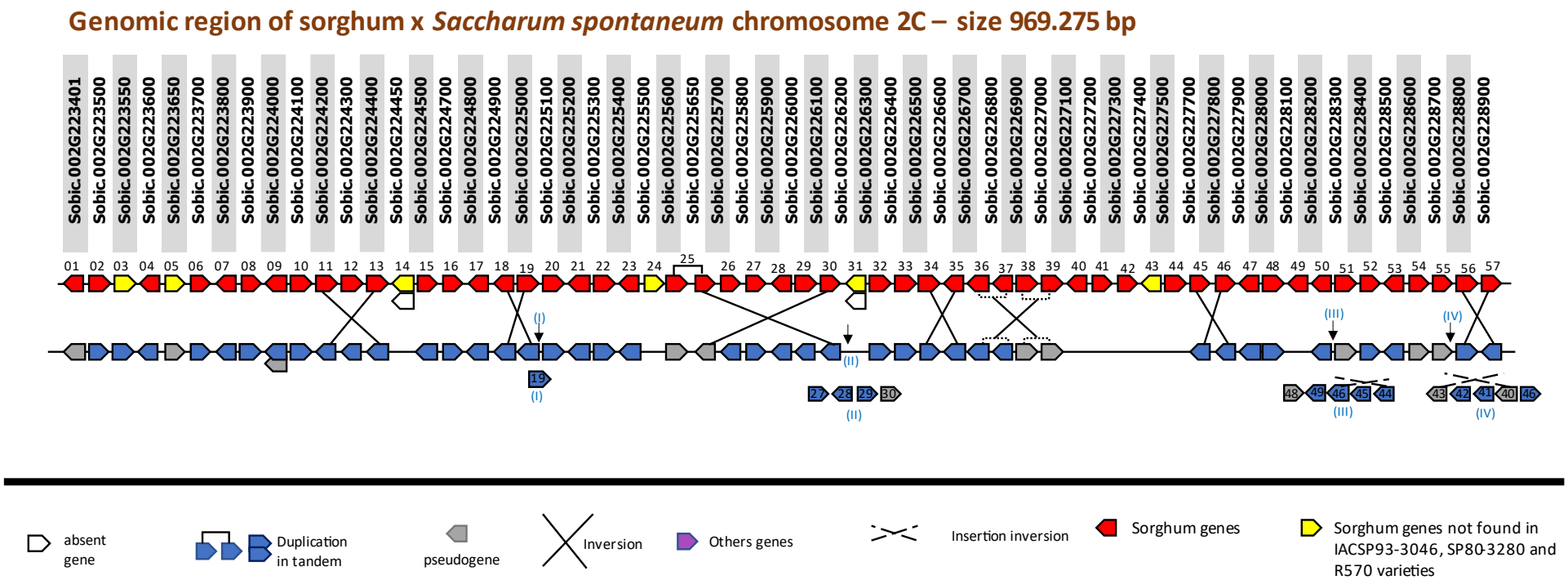
Supplementary Figure 1



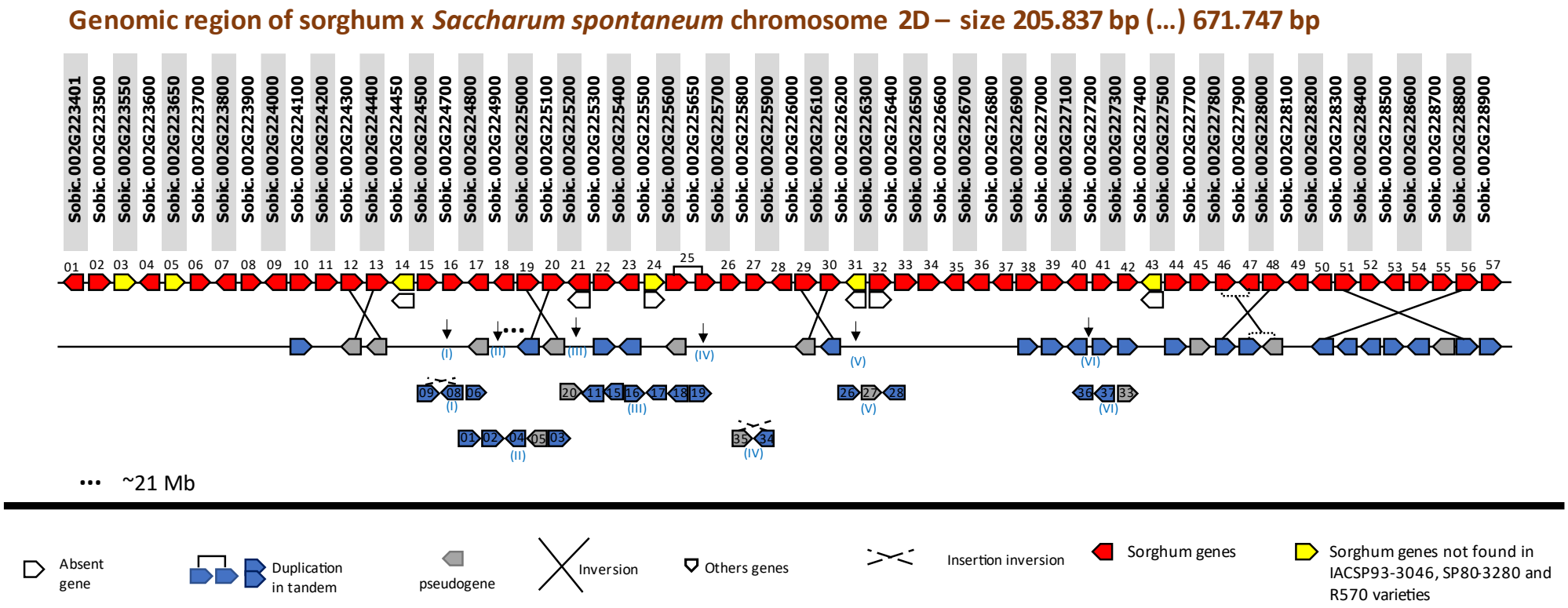
Supplementary Figure 2-



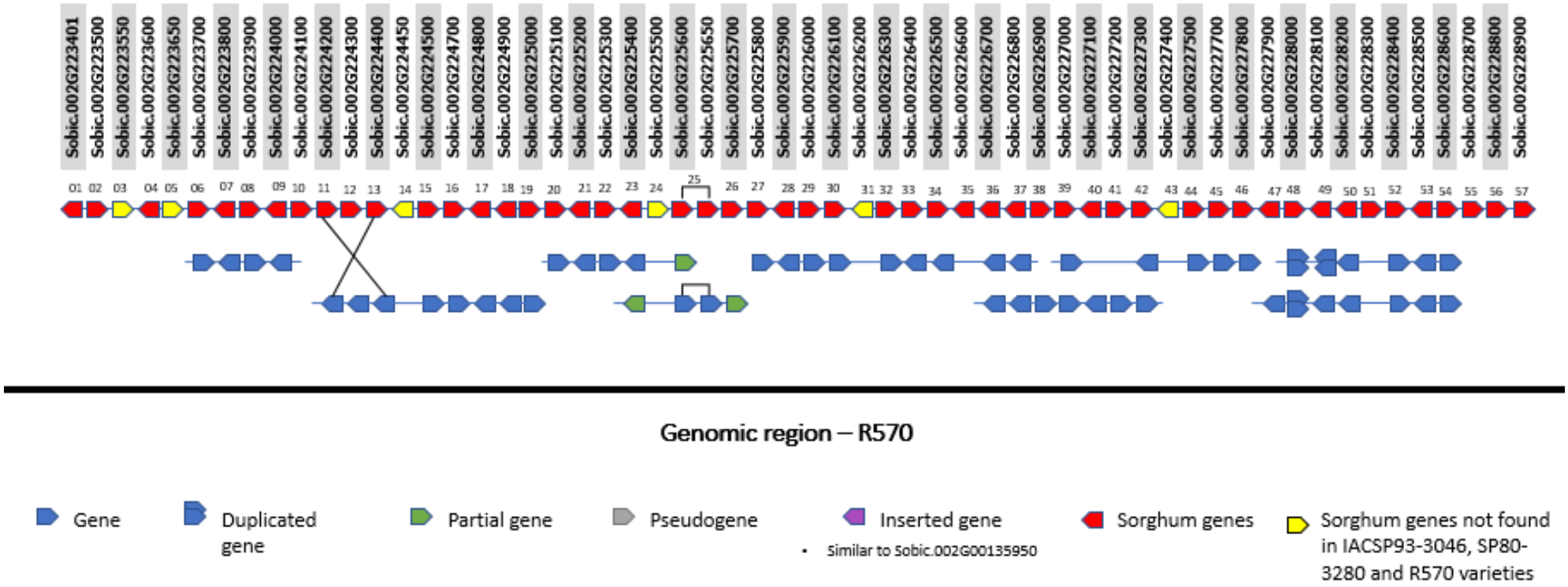
Supplementary Figure 3-



Supplementary Figure 4-



Supplementary Figure 5



Supplementary Figure 1- Sorghum×Sspon.2A genomic regions. *Genomic region of sorghum:* Each square represents a gene and shows the direction of the gene in the genome, where the right-facing arrow indicates the forward direction, and the left-facing arrow indicates the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. *Genomic region of chromosome Sspon.2A:* Below the representation of the genomic region of sorghum is the orthologous genomic region in *S. spontaneum* for the chromosome Sspon.2A. Genes are shown in blue squares on a solid line, pseudogenes in gray, and genes that do not have orthologs in the sorghum QTL in purple. Each gene was numbered from 1 to 57 (Supplementary Table 3). The insertions of one or more gene clusters are shown by the Roman numerals indicated by the black arrows. In white, genes are depicted as absent not only in the represented chromosomal region but also throughout the entire chromosome. Inversions are indicated for crossed lines. The dotted crossed lines show an inversion inside an insertion. Synteny can be observed despite breaks in collinearity. An insertion represented by the purple square can be observed for this insertion in the gene Sspon.02G0013290, a duplicate of a gene from the same chromosome that is orthologous to sorghum, Sobic.010G093001, but in this case, the gene is observed on chromosome Sb01.

Supplementary Figure 2- Sorghum×*S. spontaneum* Sspon.2B genomic regions. *Genomic region of sorghum:* Each square represents a gene and shows the direction of the gene in the genome, where the right-facing arrow indicates the forward direction, and the left-facing arrow indicates the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. *Genomic region of Chromosome Sspon.2B:* Below the representation of the genomic region of sorghum is the orthologous genomic region in *S. spontaneum* for the chromosome Sspon.2B. Genes are shown in blue squares on a solid line, pseudogenes in gray, and genes that do not have orthologs in the sorghum QTL in purple. Each gene was numbered from 1 to 57 (Supplementary Table 3). The insertions of one or more gene clusters are shown by the Roman numerals indicated by the black arrows. In white, genes are depicted as absent not only in the represented chromosomal region but also throughout the entire chromosome. Inversions are indicated for crossed lines. The dotted cross lines show an inversion inside an insertion. Synteny can be observed despite breaks in collinearity. The cluster of genes depicted in purple in insertion IV is orthologous to sorghum, which has the same chromosome and the same sequence (Phytozome v.13). The absence of similar findings in the other *S. spontaneum*

alleles, as well as in the IACSP93-3046, SP80-3280 and R570 haplotypes, suggests the possibility of a specific duplication in this particular allele.

Supplementary Figure 3- Sorghum×Sspon.2C. *Genomic region of sorghum:* Each square represents a gene and shows the direction of the gene in the genome, where the right-facing arrow indicates the forward direction, and the left-facing arrow indicates the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. *Genomic region of chromosome Sspon2C:* Below the representation of the genomic region of sorghum are the orthologous genomic regions in *S. spontaneum* for each allele of chromosome Sspon.2C. Genes are shown in blue squares on a solid line, and pseudogenes are shown in gray. Each gene was numbered from 1 to 57 (Supplementary Table 3). The insertions of one or more gene clusters are shown by the Roman numerals indicated by the black arrows. In white, genes are depicted as absent not only in the represented chromosomal region but also throughout the entire chromosome. Inversions are indicated for crossed lines. The dotted cross lines show an inversion inside an insertion. Synteny can be observed despite breaks in collinearity.

Supplementary Figure 4- Sorghum×Sspon.2D. *Genomic region of sorghum:* Each square represents a gene and shows the direction of the gene in the genome, where the right-facing arrow indicates the forward direction, and the left-facing arrow indicates the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. *Genomic region of chromosome Sspon.2D:* Below the representation of the genomic region of sorghum are the orthologous genomic regions in *S. spontaneum* for each allele of chromosome Sspon.2D. Genes are shown in blue squares on a solid line, and pseudogenes are shown in gray. Each gene was numbered from 1 to 57 (Supplementary Table 3). The insertions of one or more gene clusters are shown by the Roman numerals indicated by the black arrows. In white, genes are depicted as absent not only in the represented chromosomal region but also throughout the entire chromosome. Inversions are indicated for crossed lines. The dotted cross lines show an inversion inside an insertion. The three dots indicate a significant break in collinearity with sorghum, specifically on chromosome Sspon.2D, where the genomic region is separated by approximately 21 million base pairs. In addition to many other differences, many genes that seem to be absent in the allele are indeed present. Chromosomal absences are marked by white squares.

Supplementary Figure 5 - Sorghum×R570 genomic regions. *Genomic region of sorghum:* Each square represents a gene and shows the direction of the gene in the genome, where the right-facing arrow indicates the forward direction, and the left-facing arrow is the reverse direction. The solid lines with squares in red and yellow represent the 57 sorghum QTLs. *Genomic region of R570:* Below the representation of the sorghum genomic region, the orthologous genomic region in R570 is shown. Genes are depicted in blue. Possible pseudogenes are not represented (Garsmeur et al., 2018).

Supplementary Tables

Supplementary Table 1: Summary of PacBio® Sequel sequencing and sequence assembly plus annotations.

	Pool Name	BAC	Contigs	Longest contig size (pb)	Shortest contig size (pb)	Total bases (pb)
SP80-3280	2	2	2	138,727	102,149	240,876
	3	2	2	132,039	123,810	255,849
	5	2	1	141,128	141,128	141,128
	6	2	1	126,235	126,235	126,235
	7	2	1	106,318	106,318	106,318
	8	2	1	142,948	142,948	142,948
	9	2	2	180,638	121,873	302,511
	13	2	1	132,828	132,828	132,828
	14	2	2	102,504	87,864	190,368
	15	2	1	84,208	84,208	84,208
	16	2	1	158,108	158,108	158,108
Subtotal	-	22	15	-	-	1,881,377
	1	1	1	107,880	107,880	107,880
	2	2	1	126,094	126,094	126,094
	3	1	1	135,702	135,702	135,702
	16	1	1	115,793	115,793	115,793
	17	3	5	192924	8538	391,863

IACSP93-3046	18	3	4	113852	10714	252,862
	19	3	9	127023	4750	378,918
	20	3	6	183570	4414	482,101
	21	3	8	133902	4038	325,041
	22	3	3	153020	4057	251,633
	23	3	5	181700	4576	506,830
	24	3	5	152382	14478	522,427
	25	3	7	124722	3960	371,939
Subtotal	-	32	56	-	-	3,969,083
Total	**-**	**63**	**71**	**-**	**-**	**5,850,460**

Supplementary Table 2: Results of PacBio® Sequel sequencing and sequence assembly plus annotation.

Pool Name	Contig	Conting length	Coverage	BAC clone	Variety
1	1	107880	61.22	Shy112C03	IACSP93-3046
2	4	126094	42.98	Shy123P01	
	7	138727	22.07	Shy488A19	SP80-3280
	17	102149	30.52	Shy361L04	
3	3	135702	52.33	Shy187N11	IACSP93-3046
	10	132039	64.22	Shy255L20	SP80-3280
	15	123810	140.51	Shy289A21	

5	6	141128	17.41	Shy253P08	
6	08	126235	59.89	Shy486B15	
7	17	106318	129.15	Shy041F06	
8	8	142948	28.61	Shy260G24	
				Shy223J17	
9	1	180638	67.16	Shy378L10	
	5	121873	59.36	Shy378L03	
13	7	132828	58.63	Shy368O04	
14	21	102504	47.01	Shy492F12	
	29	87864	164.54	Shy021C23	
15	15	84208	276.17	Shy486F01	
16	1	158108	109.61	Shy504G20	
	10	115793	53.93	Shy141H03	IACSP93-3046
17	3	192924	110.5	Shy273L13	
	29	102135	126.15	Shy130N20	
	74	45538	48.08	Shy012B04	
18	5	113852	199.27	Shy416D13	
	13	106999	99.82	Shy006P16	
19	1	127023	115	Shy120H04	
	48	54704	31.71	Shy411A07	
	114	36851	25.38		
	2804	126947	134.62	Shy178E18	
20	3	183570	85.32	Shy192F11	

	9	170179	204.58	Shy031N20	
	26	72739	212.27	Shy265N09	
	47	46762	220.47		
21	1	680338	42.32	Shy282B05	
	228	21867	9.18	Shy238L17	
	254	8060	11.17		
22	4	153020	375.13	Shy333K17	
	23	94556	372.38	Shy404J16	
23	2	151700	83.2	Shy188C18	
	3357	159080	103.9	Shy191K12	
24	1	152382	124.36	Shy320P14	
	10	134990	72.59	Shy187N11	
25	17	110471	135.26	Shy406H18	
	19	124722	101.93	Shy397C11	
	24	102547	88.29	Shy116E06	

Supplementary Table 03: Summary of orthologous genes in sorghum and their proteins. Each gene was assigned a number following the order in which it appeared in the sorghum QTL. There are some absences in the column listing the names of the genes in *S. spontaneum*; however, this does not indicate that the genes are absent in the genomic region but rather that they were not detected by automated annotation (Zhang et al., 2018). In fact, only gene 14 was absent from chromosome Sspon.2 in *S. spontaneum.* The other genes (3, 5, 24, 27, 30, 31, 35, 36, 43, 51 and 55) were visualized and annotated manually (Supplementary Figures 1, 2, 3 and 4).

Gene number	*S. bicolor* gene	*S. spontaneum* gene	Protein (Phytozome 13)
01	Sobic.002G223401	Sspon.02G0013480	Similar to a protein of the prolyl oligopeptidase family **(POP)**
02	Sobic.002G223500	Sspon.02G0013470	Predicted **AST**-like aspartate aminotransferase protein
03	Sobic.002G223550	-	Predicted, uncharacterized protein.
04	Sobic.002G223600	Sspon.02G0013460	Predicted phosphatidylinositol-4-phosphate 5-kinase-related protein **(PIP5Ks)**.
05	Sobic.002G223650	-	Hypothetical, uncharacterized protein.
06	Sobic.002G223700	Sspon.02G0013450	Protein inferred by homology. Rhomboid-like protein RBL10.
07	Sobic.002G223800	Sspon.02G0013440	Hypothetical uncharacterized protein.
08	Sobic.002G223900	Sspon.02G0013430	Hypothetical protein. Similar to F-box-like proteins (FBPs).
09	Sobic.002G224000	Sspon.02G0014420	Similar to alpha-carbonic anhydrase domain-containing protein (CA).
10	Sobic.002G224100	Sspon.02G0013410	Similar to **Phosphate Transporter 3, PHT3.**
11	Sobic.002G224200	Sspon.02G0013390	Similar to **Heat shock factor B 4 (HSFB4) protein.**

12	Sobic.002G224300	Sspon.02G0013400	Predicted protein similar to NAM protein, an NAC transcription factor.
13	Sobic.002G224400	Sspon.02G0037420	**Receptor-like serine/threonine-protein kinase, lectin type G**
14	Sobic.002G224450	-	Uncharacterized predicted protein.
15	Sobic.002G224500	Sspon.02G0049360	Similar to **BHLH domain-containing protein** (Basic helix-loop-helix). HEC
16	Sobic.002G224700	Sspon.02G0013380	**Inference by homology, NF - Kappa B activating protein**
17	Sobic.002G224800	Sspon.02G0013370	Protein belonging to the GDXG family, a family of lipolytic enzymes. Similar to alpha/beta hydrolase (with folded domain. Alpha/beta hydrolase fold domain - containing (ABH).
18	Sobic.002G224900	Sspon.02G0037410	Protein belonging to the GDXG family, a family of lipolytic enzymes. Similar to alpha/beta hydrolase (with folded domain. Alpha/beta hydrolase fold domain - containing (ABH).
19	Sobic.002G225000	Sspon.02G0013350	Protein inferred by homology. Similar to C-terminal remorin protein. Remorin C-Terminal like (REM).
20	Sobic.002G225100	Sspon.02G0013320	Similar to Abscisic Acid - insentive 5 - like protein 4, **basic - leucine zipper domain (ABF1).**

21	Sobic.002G225200	Sspon.02G0013280	Similar to putative uncharacterized protein B1342C04.33
22	Sobic.002G225300	Sspon.02G0013260	Similar to **heat shock transcription factor B1 (HSB1) protein.**
23	Sobic.002G225400	Sspon.02G0013240	Similar to **abscisic acid 8'-hydroxylase (AA8' OH) protein.**
24	Sobic.002G225500	-	Similar to a protein containing a **zinc finger domain**
25	Sobic.002G225600/650	Sspon.02G0013220	Similar to alpha-amylase **(AMY).**
26	Sobic.002G225700	Sspon.02G0013210	Similar to **ethylene responsive factor 109 - (ERF109)**
27	Sobic.002G225800	-	E3 ubiquitin ligase involved in syntaxin degradation
28	Sobic.002G225900	Sspon.02G0013200	Similar to fucosyltransferase (FUT). GO:0008417
29	Sobic.002G226000	Sspon.02G0013190	Malectin-like domain-containing protein **(MLD)**
30	Sobic.002G226100	-	Protein disulfide isomerase **(SCO2)**
31	Sobic.002G226200	-	Uncharacterized protein
32	Sobic.002G226300	Sspon.02G0013180	Calmodulin - like protein **(CML).**
33	Sobic.002G226400	Sspon.02G0013170	Calmodulin - like protein **(CML).**
34	Sobic.002G226500	Sspon.02G0013160	Putative uncharacterized protein similar to basic leucine zipper

			(BZIP) domain - containing protein.
35	Sobic.002G226600	-	Uncharacterized protein
36	Sobic.002G226700	-	similar to BRI1 **kinase inhibitor 1 (BKI1)**
37	Sobic.002G226800	Sspon.02G0013130	**PP2C**- protein phosphatase 2C // subfamily not named
38	Sobic.002G226900	Sspon.02G0013140	Similar to Enoyl-CoA reductase (**ECR**)
39	Sobic.002G227000	Sspon.02G37380/37400	Similar to Enoyl-CoA reductase (**ECR**)
40	Sobic.002G227100	Sspon.02G0013150	Cyclin-dependent kinase inhibitor (**CDK-inhibitor**)
41	Sobic.002G227200	Sspon.02G0037390	Similar to DNAJ-proteins or heat shock protein 40 (**HSP40**)
42	Sobic.002G227300	Sspon.02G0037370	Uncharacterized protein - predicted
43	Sobic.002G227400	-	Similar to pentatricopeptide repeat **(PPR)**
44	Sobic.002G227500	Sspon.02G0013040	Similar to putative **SEC23** protein transport protein SEC23
45	Sobic.002G227700	Sspon.02G0013050	Protein belonging to the GDXG family, a family of lipolytic enzymes. Similar to alpha/beta hydrolase (with folded domain. Alpha/beta hydrolase fold domain - containing (**ABH**).

46	Sobic.002G227800	Sspon.02G0013060	Protein belonging to the GDXG family, a family of lipolytic enzymes. Similar to alpha/beta hydrolase (with folded domain. Alpha/beta hydrolase fold domain - containing (**ABH**).
47	Sobic.002G227900	Sspon.02G0013070	Similar to alpha/beta hydrolase fold domain - containing (ABH). Gibberellin receptor (**GID1L2).**
48	Sobic.002G228000	Sspon.02G0013080	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
49	Sobic.002G228100	Sspon.02G0013090	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
50	Sobic.002G228200	Sspon.02G0013090	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
51	Sobic.002G228300	-	Uncharacterized and predicted protein.
52	Sobic.002G228400	Sspon.02G0013100	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
53	Sobic.002G228500	Sspon.02G0013110	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
54	Sobic.002G228600	Sspon.02G0013120	Similar to alpha/beta hydrolase fold domain - containing (ABH). Carboxylesterase (**CXE**)**.**
55	Sobic.002G228700	-	Similar to phospholipase A / patatin-related (**pPLA**)

56	Sobic.002G228800	Sspon.02G0037310	Uncharacterized protein - predicted
57	Sobic.002G228900	Sspon.02G0013030	RNA Binding Protein - **RBP47**

Supplementary Table 4. Differentially expressed genes (DEGs) identified between *Saccharum spontaneum* and *Saccharum officinarum*, IACSP93-3046 and SP80-3280. Log_2(fold change) (Log_2(FC)) values and false discovery rate (FDR)-corrected *p* values are provided for each gene.