Mini Review

Genomic resources for energy cane breeding in the post genomics era

Augusto L. Diniz [a], Sávio S. Ferreira [b], Felipe ten-Caten [a], Gabriel R.A. Margarido [c], João M. dos Santos [d], Geraldo V. de S. Barbosa [d], Monalisa S. Carneiro [e], Glaucia M. Souza [a,*]

[a] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748, São Paulo 05508-000, SP, Brazil
[b] Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, São Paulo 05508-090, SP, Brazil
[c] Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Avenida Pádua Dias, 11, Piracicaba 13418-900, SP, Brazil
[d] Departamento de Fitotecnia e Fitossanidade, Centro de Ciências Agrárias, Universidade Federal de Alagoas, BR 104 Norte, km 85, Rio Largo 571000-000, AL, Brazil
[e] Departamento de Biotecnologia e Produção Vegetal e Animal, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Rodovia Anhanguera km 174, Araras 13600-970, SP, Brazil

ABSTRACT

Sugarcane is one of the most sustainable energy crops among cultivated crops presenting the highest tonnage of cultivated plants. Its high productivity of sugar, bioethanol and bioelectricity make it a promising green alternative to petroleum. Furthermore, the myriad of products that can be derived from sugarcane biomass has been driving breeding programs towards varieties with a higher yield of fiber and a more vigorous and sustainable performance: the energy cane. Here we provide an overview of the energy cane including plant description, breeding efforts, types, and end-uses. In addition, we describe recently published genomic resources for the development of this crop, discuss current knowledge of cell wall metabolism, bioinformatic tools and databases available for the community.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction

Sugarcane (Saccharum spp.) is a perennial, tropical or subtropical non-cereal C4 grass; the major crop for food and bioenergy production and the highest tonnage crop in the world [1]. Mainly grown for sugar production in the tropical and subtropical regions of the world, sugarcane has one of the highest solar energy conversion efficiency and is a crop with one of the highest biomass yields [2–4]. Due to the high yields of both sugar and lignocellulosic biomass it is considered an important feedstock for substituting fossil fuel energy, either as natural biomass or transformed into liquid or gaseous forms [5]. An additional advantage is that this semi-perennial grass allows harvesting for several years (4 to 5 years)

without the need for replanting, which reduces the cost of bioenergy production [6]. The development of conversion processes that use all plant carbon in a Biorefineries approach [7,8] has stimulated the development of plants with several co-products for different applications, including sugars, biofuels and bioelectricity. In this scenario, breeders are designing crosses towards the development and improvement of energy cane, a new type of cane containing a high yield of fiber.

Over the last century, there was a great effort in breeding programs and conventional agricultural research to increase the yield of sugarcane and sugar to reach the current levels [9]. In addition, research groups around the world, have generated a large amount of molecular and physiological data on sugarcane. In addition, sugarcane geneticists have invested significant effort to explore and dissect the complex genome of cane using a wide range of genomic tools. These combined factors make data integration a key step to achieve a broader understanding of cane physiology and its interaction with the environment and climate changes as well as to help the design of more productive varieties.

In this mini review, we focus on bioenergy and energy cane and provide an overview of the current genomic resources and databases for the development of this crop. It is important to note that the theoretical potential of sugarcane dry biomass production is 177 t/(ha yr) or a fresh weight cane yield of 381 t/(ha yr) [10]. Worldwide sugarcane yield averages around 39 and 84 t/(ha yr), respectively. There is great interest and opportunity to decrease this yield gap.

2. The *Saccharum* complex

Sugarcane is the world's leading biomass crop, produced in over 100 countries [1]. Modern sugarcane cultivars are polyploid inter-specific hybrids, typically with 10–13 sets of their 10 basic chromosomes, 80–85% of *Saccharum officinarum* (2n = 80), 10–15% of *S. spontaneum* (2n = 40–128) and ~5% with recombined chromosomes between these two ancestors [11,12]. Gene duplications as a result of polyploidization alter the transcriptional landscape [13] and provide additional flexibility to adapt and evolve new patterns of gene expression for homo(eo)logous gene copies [14]. This flexibility has been suggested to be an important mechanism allowing the diversification of adaptive traits [15,16] through neo-functionalization of duplicated genes [17] and tissue-specific expression [18].

The high productivity cane makes this crop an excellent source of sugar, bioethanol and bioelectricity [19] and a promising green alternative to petroleum [20–22] with vast potential to mitigate climate change without affecting food security [23]. Additionally, the myriad of products that can be derived from sugarcane biomass [24], such cellulosic bioethanol, further enhance opportunities for sugarcane in a portfolio of technologies needed to transition to a low carbon 'bioeconomy'. In this scenario, hybrids obtained through the cross-breeding of commercial varieties of sugarcane with ancestral species, such as *S. spontaneum*, have allowed the production of genotypes characterized by high fiber content, moderate brix levels, fine stalks and higher tillering rate – the energy cane [25–27].

3. The energy cane

Energy cane is an ideal type of sugarcane with high yield of fiber, more vigorous and rustic, i.e. these plants are less demanding in soil, climate, water and nutrients, more resistant to pests and diseases, which brings a series of economic and environmental advantages [25,28,29]. Efforts to develop this crop include inter-specific hybridization between modern sugarcane varieties and

other closely related wild species, such as *S. spontaneum*, which has the greatest potential as a source of genetic variation for a number of important traits for bioenergy production and low-input adaptability [30].

Initially, Tew and Cobil [31] classified energy cane hybrids into two categories: one (Type I) defined as a cane closer to the conventional sugarcane regarding sucrose content but higher fiber content; and other (Type II) with only marginal content of sugar but with fiber content higher than Type I, used exclusively for biomass production. More recently, Kumar et al. [32] proposed the classification of cane varieties considering variation in sucrose and fiber content. According to the authors, cane type I includes traditional commercial sugarcane varieties, with high sugar and commercial yield (13% sucrose and 12% fiber content); cane type II also comprises varieties with high sugar and commercial yield, but with an increase in fiber content (13% sucrose content and >14% fiber content); cane type III (energy cane) includes varieties for multiple use purposes, focusing on high biomass production (sucrose content <12% and >22% fiber content); and cane type IV (energy cane) embraces varieties for energy cogeneration purposes (sucrose content <5% and fiber >22%).

Compared to sugarcane commercial varieties, energy canes have higher ratooning ability and number of tillers [27,29,30,33], which combined are very important in defining total biomass yield (Fig. 1). Because this crop is "vegetative propagation based", this characteristic is also important in overcoming one of the most significant economic constraints in the cane cultivation: clonal multiplication [27]. In addition, there is also a profound difference regarding the root systems; energy cane produced an abundant and vigorous root system, surpassing the conventional cane in

Fig. 1. An energy cane RB hybrid and a sugarcane variety (SP791011) at six months after planting, under field conditions at the experimental site 'Estação de Floração e Cruzamento da Serra do Ouro' (lat 9° 13′ S, long 35° 50′ W, alt 450 m asl) in Alagoas, Brazil.

lateral extension, depth and volume. This trait, shared with the *S. spontaneum* progenitor, which is considered an invasive weed in some countries [34], allows its cultivation in marginal lands because it gives grater rusticity, helps mitigate soil erosion, boost permanent carbon sequestration and extends the crop life cycle up to 10 years; this is an important attribute due to the high cost of replanting sugarcane [27,30].

The energy cane breeding initiatives began in Puerto Rico, with a pioneering commercial project established to conduct an integrated exploration of sugarcane as a biomass feedstock for multiple products, instead of only sugar [35]. The growing interest in bioenergy in recent decades pushed several sugarcane breeding programs world-wide to also produce energy cane commercial varieties. In the United States, the Cultivar L 79-1002 ('CP 52-68' x Tainan, *S. spontaneum* clone) was developed by the Louisiana State University Agricultural Center in cooperation with the USDA-ARS and the American Sugarcane League, Inc. This cultivar has high biomass yield and fiber content, on average of 257 g kg^{-1} [36]. The breeding program in Barbados has vigorous canes with exceptional fiber content (>30%), which are suitable for energy cogeneration [37]. Mauritius developed a similar program, aiming the increasing of biomass and fiber yield [38]. Other genetic improvement initiatives have been conducted in Australia [39,40], Colombia [41], Japan [42–44], and Thailand [45]. Further details about breeding programs using *S. spontaneum* as an integral part of their activities can be found in Matsuoka et al. [27] and da Silva [30].

In Brazil, energy cane hyrids were obtained by Canavialis, a private sugarcane breeding company [27]. In 2011, an initiative was launched to create the first biorefinery in South America to produce cellulosic ethanol from sugarcane residues and energy cane genotypes, the Brazilian Group GranBio, an innovation industry for ethanol generation through biomass conversion [46]. New varieties have been developed by RIDESA (Inter-University Network for the Development of Sugarcane Industry) at the experimental site 'Estação de Floração e Cruzamento da Serra do Ouro' (lat 9° 13′ S, long 35° 50′ W, alt 450 m asl), Federal University of Alagoas (UFAL), Brazil. This institution maintains an important collection of sugarcane germplasm, which holds modern hybrids and a myriad of *Saccharum*, *Erianthus* and *Miscanthus* accessions. In a study conducted by UFAL/RIDESA researchers, six energy cane clones were selected which presented an overall average of 24.7% higher yield of dry biomass/ha as compared to the standard sugarcane variety (RB0442) (Table 1). Plants were grown under field conditions at São Miguel dos Campos – Alagoas – Brazil; and traits were measured at 13 months after planting and in the second harvest year.

There are three major problems in the selection of energy cane clones: high incidence of smut disease (*Sporisorium scitamineum*), high flowering rates and low unit stem mass. Thus, the major challenge of genetic breeding for the commercial cultivation and consolidation of energy cane cultivars is to use more effective strategies to overcome these three problems. In addition, there are other technological (agro-industrial) bottlenecks, such as: (*i*)

in mechanized harvesting, development of harvesting machines for high biomass and high fiber cultivars; and (*ii*) in industrial processing, improve the efficiency of grinding for broth extraction (Geraldo Barbosa, UFAL, personal communication).

4. Sugarcane genomic resources

As mentioned before, modern sugarcane cultivars are polyploid interspecific hybrids and have a large (~10 Gb) and complex genome. Nevertheless, opportunities to accelerate breeding progress and enrich the knowledge of the fundamental biology of this important crop drive efforts to explore and dissect its complex genome using different genomic tools and to develop a high-quality reference genome.

After over a decade of multiple parallel genome sequencing initiatives [47,48], Garsmeur et al. published the first mosaic monoploid genome reference of the modern cultivar R570 [49]. This study relies on 4535 bacterial artificial chromosomes (BACs) sequences that were colinear to the gene-rich portion of the sorghum genome (used as a reference). The final assembly consisted of 382 Mb of high-quality sequence in 3965 contigs, organized as a single tiling path, representing the single copy sugarcane gene space, and includes 25,316 predicted protein-coding gene models.

The next milestone was the publication of the first allele-defined genome reference of a tetraploid *S. spontaneum* genotype (AP85-441) [50]. To this end, authors took advantage of multiple sequencing technologies including high-throughput chromatin conformation capture (Hi-C) to assemble 32 pseudo-chromosomes (2.9 Gbp) comprising 8 homologous groups of 4 members each, bearing 35,525 genes with alleles defined. Subsequently, Nascimento et al. [51] published a new *S. spontaneum* gene space reference (from accession US851008), including 39,234 genes.

Recently, our group completed the assembly (4.26 Gb) of the Brazilian modern cultivar SP80-3280 [52], which includes the complete sequence of 373,869 genes and their upstream regions which may be further explored to identify regulatory promoter elements. This is the largest genomic data set available for the sugarcane researchers' community and includes putative homo(eo)logs (mostly 2–5 copies) for a large fraction of the SP80-3280 gene space [52].

To date, most of the sugarcane RNA-seq initiatives are based on *de novo* transcriptome assembly [53–69]. Nevertheless, a few studies also consider the closely related species *Sorghum bicolor* [70], the *S. officinarum* gene indices (SoGI) v3.0 [71] or the previously *de novo* assembled transcriptome [72] as a reference for transcriptome analysis. Further transcriptome studies are expected to take advantage of the multiple reference genomes now available.

To investigate the value of the three largest public assemblies as a genomic resource, we used RNA-seq data from Kasirajan et al. [71]. These authors created RNA-seq libraries from top and bottom internodes of 20 different genotypes, including commercial cultivars and introgression lines derived from crosses with wild

Table 1
Comparison of yield related traits between six RB energy cane clones and a sugarcane standard variety (RB0442).

	Value	Traits					
		FB	DB	Fiber (%)	TFH	TRS	TSH
Top 6 RB energy cane clones	Min	131.8	43.2	20.2	26.6	51	10.6
	Average	137.6	46.4	22.7	31.2	67.8	12.7
	Max	153.3	50.1	24.1	33.9	96.2	15.1
RB0442	Average	126.5	37.2	13.8	17.4	122.1	18
RB clones vs RB0442	%	8.8	24.7	64.5	79.3	−44.5	−29.4
CV	%	20.94	21.12	16.97	25.00	29.01	28.49

CV = coefficient of variation; FB = fresh biomass (t/ha); DB = dry biomass (t/ha); TFH = fiber (t/ha); TRS = Total Recoverable Sugar (%); TSH = sugar (t/ha)
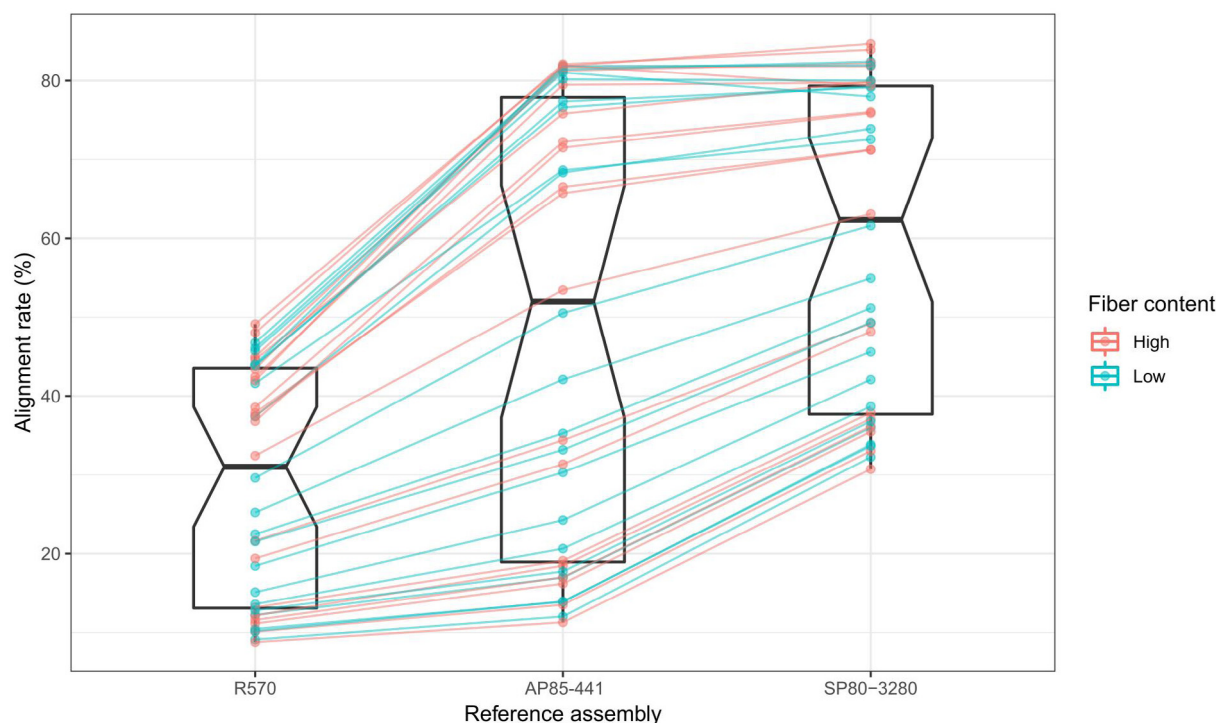
Fig. 2. Alignment rates of RNA-seq libraries from the top and bottom internodes of 20 different genotypes [71], contrasting for fiber content, against the monoploid R570 hybrid assembly [49], the AP85-441 tetraploid *S. spontaneum* genome reference [50] and the SP80-3280 hybrid gene space assembly [52].

S. spontaneum relatives and *Erianthus*. We used HISAT2 [73] with default parameters to align these RNA-seq reads against the monoploid R570 hybrid assembly [49], the AP85-441 tetraploid *S. spontaneum* reference [50] and the SP80-3280 hybrid gene space assembly [52].

For the vast majority of sequenced samples, the SP80-3280 assembly resulted in higher alignment rates than those against the monoploid reference and the *S. spontaneum* assembly (Fig. 2). Only in the very high end of alignment rates did the AP85-441 assembly perform better for three of the 40 samples. These results show that these genome assemblies, and in particular the SP80-3280, can be used as a reference for downstream genomic studies. Combining both *de novo* and reference-guided transcriptome approaches, especially by taking advantage of multiple genome references, may allow the understanding of the extent to which homo(eo)logs resemble or differ from each other in their expression patterns, the spatiotemporal dynamics of these relationships, and how epistatic interactions between individual homo(eo)logs affect biological traits.

5. Linking genomic data to biomass improvement by exploring the plant cell wall metabolism

Plant biomass is composed mainly by secondary cell walls (SCW) and, consequently, achieving tailor-made biomass for bioenergy could benefit from a detailed understanding of SCW biosynthesis. The potential use of currently available sugarcane genomic resources for identifying classes of genes involved in SCW biosynthesis will help to shed light on these aspects. Table 2 indicates the number of cell wall-related genes found in different *Saccharum* genomic resources as well as transcription factors (TFs) from the two main families, NAC and MYB, involved in the gene regulatory network (GRN) controlling SCW biosynthesis.

The SCW-GRN has been elucidated in Arabidopsis and it is comprised by a three-layered structure. At the top level, TFs from the NAC family, named VASCULAR-RELATED NAC-DOMAIN (VNDs) and NAC SECONDARY WALL THICKENING PROMOTING FACTOR (NSTs) genes, act as master switches activating cell differentiation, including programmed cell death to form tracheary elements, and

Table 2
Number of cell wall-related genes and NAC and MYB transcription factors identified in different *Saccharum* ssp. genomic databases: the R570 monoploid genome reference (R570) [49]; the allele-define genome reference of *S. spontaneum* (AP85-441) [50]; the SP80-3280 gene-space assembly; and the SAS (Sugarcane Assembled Sequences) from the SUCEST Project (SUCEST) [94].

| Gene class	R570	AP85-441	SP80-3280	SUCEST
Cell Wall Differentiation	83	386	742	175
Cell Wall Growth/Extension	54	219	434	58
Lignin metabolism	31	168	541	53
Other Glycan Degradation	117	456	870	177
Phenylpropanoid biosynthesis	58	311	511	123
Polysaccharide biosynthesis	40	220	475	112
Structural proteins	15	68	103	27
Unknown	1	7	7	5
MYB	178	999	613	120
NAC	101	412	884	63
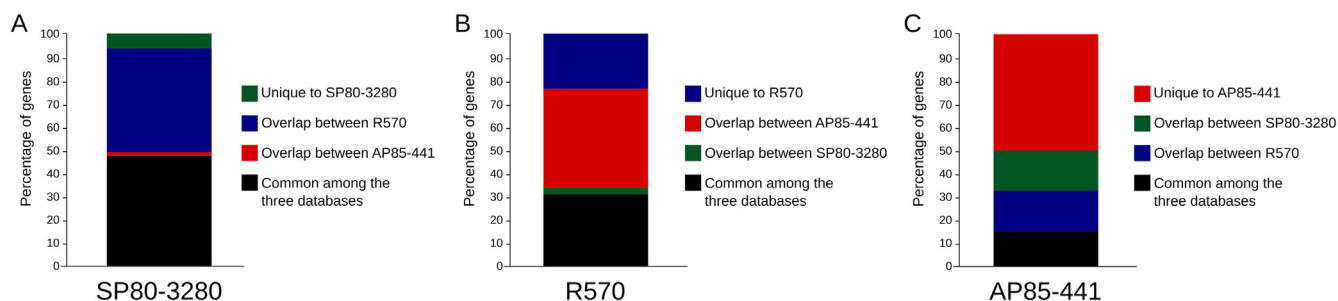
Fig. 3. Overlap of MYB genes among the three sugarcane genomic assemblies (datasets). All genes classified as MYBs (Table 2) were used for reciprocal blastp analysis among all three datasets. A, overlap of SP80-3280 with the other two databases; B, overlap of R570 with the other two databases; C, overlap of AP85-441 with the other two databases. Genes with coverage and identity >=90% were considered "overlapping genes".

SCW deposition in vessels and fibers [74–76]. They activate a second layer of master switches, comprised by TFs from the MYB family (AtMYB46 [77] and AtMYB83 [78] in Arabidopsis). These second level TFs activate biosynthetic genes of cellulose, hemicellulose (xylan) and lignin and a third layer of TFs, turning on SCW deposition. These downstream TFs activate other aspects of SCW deposition, with some redundancies, and includes transcriptional repressors of NAC master switches, such as AtMYB32 [79], establishing a negative feedback loop.

Besides this three-layer core structure and key players of GRN being conserved among vascular plants studied so far [80,81], including grasses like rice, maize, brachypodium, sorghum, miscanthus and switchgrass, considerable divergence of transcription factors target diversities has been reported [82–91]. Even among grasses some divergences of TF target repertoires may exist, such as in MYB SCW repressors orthologs [92]. Within the R570, SP80-3280, AP85 and SUCEST SAS (see below) databases (Table 2), we have found up to 884 and 999 NAC and MYB genes, respectively, which are potential targets for further exploration. For the NAC family, this number is 8-11x higher than in rice (105 genes) and *Arabidopsis thaliana* (75 genes) genomes [93], putting in perspective all the intricacy and diversity found in the *Saccharum* complex.

By comparing the overlap of genes among the three references, we can estimate how these datasets can complement each other. In the transcription factor MYB family, 15–50% of these genes is common to all three genotypes (Fig. 3), suggesting that all three databases have their particularities, given the fact that they are derived from different species (*S. spontaneum*) or varieties (R570 and SP80-3280). This is evident for AP85-441 (*S. spontaneum*) which reflects in approximately 50% of MYB genes not overlapping to sugarcane varieties (Fig. 3C).

As lignin is one of the main causes of biomass recalcitrance, hampering lignocellulosic biofuels production [95–99], much effort has been made to understand its biosynthesis and polymerization and how to engineer it. Lignin is a phenolic polymer composed of three main units, p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S), crosslinked to hemicellulose providing strength and rigidity to the cell wall [100]. Although known for years [100], additional insight into the phenylpropanoid pathway and lignin synthesis have occurred recently. For example, the Caffeoyl-shikimate esterase (CSE) [101], the bifunctional phenylalanine/tyrosine ammonia lyase (PTAL) [102] and a bifunctional cytosolic ascorbate peroxidase functioning as C3H [103] are missing links in the phenylpropanoid pathway that have only recently been discovered. There are ~511 genes in the phenylpropanoid pathway of SP80-3280 (Table 2), whereas other species like *Arabidopsis thaliana* and the model C4 grass *Setaria viridis* have only 26 [104] and 56 [105], respectively. Gaps in our knowledge include how monolignols, the lignin monomers, are transported to the apoplast for polymerization [106]. To date, only one transporter

has been characterized, AtABCG29, which is responsible for transporting the precursor of p-hydroxyphenyl (H-units), a minor component of lignin, but other routes should exist since loss-of-function mutations in AtABCG29 reduce H-units, but do not eliminate it [107].

The majority of studies about plant cell wall metabolism are derived from dicot model plants, particularly Arabidopsis. However, grasses and dicots diverged ~150 million years ago; therefore, considerable differences in their vascular and morpho-anatomic patterns and cell wall structure and composition have emerged. Grasses have several distinct features from eudicots cell walls [108], with different abundance of pectin, structural proteins and phenolic compounds and also hemicellulose structure and composition [109]. Eudicots have only traces of H-units and low levels of other phenolic compounds in their cell walls, whereas grasses have significant amounts of H-units and increased levels of hydroxycinnammic acids [110], especially ferulic and p-coumaric acids esterified to arabinoxylan [108,111,112] and ferulate-monolignol conjugates incorporated to lignin [113]. Furthermore, the flavonoid tricin was discovered in monocot lignin [114,115], acting as a nucleation site for lignification.

Grasses have arabinoxylan as the main hemicellulose, but eudicots do not have arabinosyl substitution in secondary wall xylan, which affects how lignin is crosslinked to hemicellulose [116]. Moreover, mixed-linkage glucan is a monocot-specific hemicellulose, due to the absence in eudicots of the genes responsible for its biosynthesis, cellulose synthase-like F and H (CslF e CslH) [117–119]. Also absent in eudicots, the bifunctional PTAL can use tyrosine as well as phenylalanine as substrate in the first step of phenylpropanoid pathway yielding 4-coumarate, thus bypassing the reaction catalyzed by cinnamate 4-hydroxylase (C4H), giving plasticity to the metabolism [102]. Furthermore, a transcription factor from the MYB family (BdSWAM1) was recently reported as SCW biosynthesis regulator in *Brachypodium distachyon*, although its clade is not found in the Brassicaceae family [120], which includes Arabidopsis. On the other hand, CSE, an essential enzyme in eudicot phenylpropanoid pathway whose down-regulation improves biomass saccharification, does not have a bona-fide ortholog identified in grasses so far [101,121,122]. Given all these differences, it is expected that considerable genetic divergence may be found and much of the knowledge from dicot cell wall cannot be extrapolated to grasses. Therefore, these differences can only be uncovered by studying grass functional genomics.

How SCW biosynthesis is connected to plant growth and biomass accumulation is still less understood. However, unknown factors linking these two processes may exist [123]. For example, one of these factors could be the transcriptional regulatory Mediator complex, since it has been reported to directly control lignin biosynthesis and its disruption rescues dwarfing phenotype in Arabidopsis lignin-deficient mutants [124,125]. Such interactions are

completely unknown for grasses and may be species-specific, raising the need to study crop plant omics [123]. Categorizing these hidden molecular hubs linking SCW biosynthesis to plant growth and other physiological processes are crucial to move forward in developing novel biotechnological strategies to improve plant biomass [123].

Identifying genes of interest, addressing grass specificities and finding the missing links to improve biomass accumulation and quality is a major challenge, especially in a complex genome species such as sugarcane. We expect that the sugarcane researchers can take advantage of genomic databases such as the ones described here to explore cell wall related-genes (Table 2), for example, thus helping to advance sugarcane functional genomics and giving new opportunities for molecular breeding to achieve and improve energy canes.

6. SUCEST-FUN Database: A platform for sugarcane data integration in a genomic context

In addition to the recent genome assemblies, plant genomic databases such as GRASSIUS [126], TropGENE [127], Phytozome [128], Plant TF database [129], MOROKOSHI [130], KBase [131], Gramene [132], PLAZA [133] and Plant GDB [134], are important foundations for molecular breeders to mine candidate genes and to facilitate molecular crop breeding.

Specially for sugarcane and energy cane breeders, the SUCEST-FUN Platform (http://sucest-fun.org/) [135] was developed to allow data analysis on five main aspects: *i*) gene annotation; *ii*) gene expression; iii) integration of public resources; *iv*) sequencing projects; and *v*) functional genomics. The database was initially based on 43,141 SAS (Sugarcane Assembled Sequences) from the SUCEST Project [94] and subsequently the 17,500 ORFeome genes generated using RNA-seq of sugarcane ancestral and hybrid varieties [69], which are useful for protein characterization, single nucleotide polymorphism analysis, splicing variants identification, evolutionary and comparative studies.

An important advantage of the SUCEST-FUN Database is the in-depth automatic and manual annotation conducted by our group and the definition of curated catalogs of transcription factors, cell wall genes, signal transduction genes (including kinases and phosphatases), KEGG metabolic pathways and enzymes, transposable elements, as well as orthologous gene analysis among grasses. For gene expression studies, the SUCEST-FUN Database supports three microarray platforms, including: (*i*) the Signal Transduction (SUCAST) array, composed by 1900 genes with 152 hybridizations; (*ii*) the RNA/Carbohydrate Metabolism and Signal Transduction (SUCAMET) array, composed by 4600 genes with more than 150 hybridizations; and (*iii*) the general regulatory function (CaneReg-

Net) array, composed by 14,522 genes, including sense and anti-sense probes with 122 hybridizations. These transcriptome studies used samples from multiple plant materials such as ancestral genotypes and commercial varieties; multiple tissues, such as leaves, internodes and roots; multiple conditions, such as field and greenhouse; and multiple treatments, such as drought stress, developmental and circadian stages, high CO_2 [136–140]. In this scenario, these experiments, summarized in Table 3, are a valuable data source for co-expression analysis, a promising approach to unravel complex biological processes and regulatory networks, which can be extracted from available tools in the platform [141].

At the genome level, the annotation of public Sugarcane BAC sequences [47] and availability of a genome browser (available at http://sucest-fun.org/cgi-bin/cane_regnet/gbrowse2/gbrowse/microsoft_genome_moleculo_scga7/) with the gene space assembly of SP80-3280 polyploid cultivar [52] enables the survey of sequences and annotation in a global and dynamic way. For instance, we present one example of how we can explore this genomic tool in Fig. 4. Using the 'SCRURZ3080F11.g' SAS ID we searched for SP80-3280 contigs holding this transcript sequence (annotated as a MYB transcription factor). As a result, we found unique matches to nine different contigs, which may represent putative homo(eo)logs for this gene (Fig. 4A). In addition, we further present the genomic features of one of these contigs (scga7_uti_cns_0226226), such as SAS [94] matches, predicted genes [52] and RNA-seq [69] alignment results (Fig. 4B).

7. Final considerations

Increasing plant yields is one of the greatest challenges of biotechnology. Computational tools that link genome sequences, their functions and possible attributes useful for breeding are greatly needed to speed up the process of improving sugarcane and the energy cane. Plant yields are directly impacted by lignocellulosic metabolism. We give an overview of the main genes involved and their regulators. Considering the size and complexity of cane genomes, the fact that many species have been used in breeding and the polyploidization that arose, datamining for genes of interest is a significant bioinformatics challenge for this crop. The SUCEST-FUN Platform comprises a robust infrastructure for storage, continuously updating, annotation, easy and controlled access and integration among the functional catalogs of the sugarcane transcriptome. Through various data-driven clustering analysis tools, crossings and enrichment analysis it allows for a systems biology approach. This will be an important resource considering progenies need to be analyzed in an integrated manner for multiple characteristics (technological, physiological, biochemical, genetic traits) for the construction of gene networks.

Table 3
Public microarray data using the Signal Transduction (SUCAST [137]) array and general regulatory function (CaneRegNet [138]) array.

Platform	GEO accession number		Experiment description	Number of hibridizations	Reference
	Platform	Series			
SUCAST	GPL3799	GSE4966	Phosphate starvation	16	[137]*
		GSE4967	Response to herbivory by Diatraea saccharalis	8	*
		GSE4968	ABA treatment	16	*
		GSE4969	MeJa treatment	12	*
		GSE4970	Response to N_2-fixing endophytic bacteria association	8	*
		GSE4971	Drought response	12	*
		GSE14732	Sucrose content relate to drought and cell wall metabolism	80	*
CaneRegNet	GPL14862	GSE33574	Drought	6	[138]
		GSE42725	Circadian rhythms	22	[139]
		GSE87826	Sugarcane vs *Leifsonia xyli* subsp. *xyli*.	4	[142]
		GSE124990	SP80-3280 growth and maturation	30	[52]
			Sugarcane Ancestral	36	[141]
	GPL22278	GPL22278	Ethephon- and AVG-induced transcriptional changes	24	[143]

Fig. 4. A view of the SUCEST-FUN genome browser, available at http://sucest-fun.org/cgi-bin/cane_regnet/gbrowse2/gbrowse/microsoft_genome_moleculo_scga7/. A: Screen shot of the result of searching for the 'SCRURZ3080F11.g' SAS (Sugarcane Assembled Sequence) derived from the SUCEST Project [94]. This SAS is annotated as a MYB transcription factor and has 9 matches (putative homo(eo)logs) in the SP80-3280 gene space. Yellow bars represent contigs and red diamonds indicate match position. B: Screen shot of the result of searching for the 'scga7_uti_cns_0226226' contig, which contains 'SCRURZ3080F11.g'.

Author contributions

GMS conceived the topic and outline, wrote and critically revised the manuscript. ALD wrote the manuscript and consolidated authors' contribution. MSC, JMS and GVSB provided energy cane breeding information and image. SSF wrote and critically revised the manuscript and identified cell wall-related genes and NAC and MYB transcription factors in different *Saccharum* genomic references. GRAM performed RNA-seq alignment to *Saccharum* genomic references. FTC curated the SUCEST-FUN genome browser. All authors reviewed the manuscript.

Declaration of Competing Interest

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

Acknowledgments

References

[1] FAOSTAT. Production/Crops, Food and Agriculture Organization of the United Nations - Statistics Division 2018. http://www.fao.org/faostat/en/#home.

[2] Henry RJ. Evaluation of plant biomass resources available for replacement of fossil oil. Plant Biotechnol J 2010;8:288–93. https://doi.org/10.1111/j.1467-7652.2009.00482.x.

[3] Byrt CS, Grof CPL, Furbank RT. C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspectivefree access. J Integr Plant Biol 2011;53:120–35. https://doi.org/10.1111/j.1744-7909.2010.01023.x.

[4] Jakob K, Zhou F, Paterson AH. Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. Vitro Cell DevBiol-Plant 2009;45:291–305. https://doi.org/10.1007/s11627-009-9214-x.

[5] Lam E, Shine J, Silva JD, Lawton M, Bonos S, Calvino M, et al. Improving sugarcane for biofuel: engineering for an even better feedstock. GCB Bioenergy 2009;1:251–5. https://doi.org/10.1111/j.1757-1707.2009.01016.x.

[6] Botha FC. Energy yield and cost in a sugarcane biomass system. Proceedings of the 2009 Conference of the Australian Society of Sugar Cane Technologists, 2009, p. 1–10.

[7] Bonomi A, Cavalett O, da Cunha MP, Lima MAP. The Virtual Sugarcane Biorefinery Concept. In: Bonomi A, Cavalett O, Pereira da Cunha M, Lima MAP, editors. Virtual Biorefinery, Cham: Springer International Publishing; 2016, p. 5–11. https://doi.org/10.1007/978-3-319-26045-7_2.

[8] Mariano AP, Dias MOS, Junqueira TL, Cunha MP, Bonomi A, Filho RM. Butanol production in a first-generation Brazilian sugarcane biorefinery: technical aspects and economics of greenfield projects. Bioresour Technol 2013;135:316–23. https://doi.org/10.1016/j.biortech.2012.09.109.

[9] Dal-Bianco M, Carneiro MS, Hotta CT, Chapola RG, Hoffmann HP, Garcia AAF, et al. Sugarcane improvement: how far can we go?. Curr Opin Biotechnol 2012;23:265–70. https://doi.org/10.1016/j.copbio.2011.09.002.

[10] Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, Souza GM. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnol J 2010;8:263–76. https://doi.org/10.1111/j.1467-7652.2009.00491.x.

[11] D'Hont A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenetic and Genome Research 2005;109:27–33. https://doi.org/10.1159/000082378.

[12] Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, et al. Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 1999;99:171–84. https://doi.org/10.1007/s001220051222.

[13] Renny-Byfield S, Wendel JF. Doubling down on genomes: polyploidy and crop plants. Am J Bot 2014;101:1711–25. https://doi.org/10.3732/ajb.1400119.

[14] Feldman M, Levy AA, Fahima T, Korol A. Genomic asymmetry in allopolyploid plants: wheat as a model. J Exp Bot 2012;63:5045–59. https://doi.org/10.1093/jxb/ers192.

[15] Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet 2017;18:411–24. https://doi.org/10.1038/nrg.2017.26.

[16] Mudge SR, Osabe K, Casu RE, Bonnett GD, Manners JM, Birch RG. Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyploid crop species. Planta 2009;229:549–58. https://doi.org/10.1007/s00425-008-0852-8.

[17] Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc R Soc B: Biol Sci 2012;279:5048–57. https://doi.org/10.1098/rspb.2012.1108.

[18] Makova KD. Divergence in the spatial pattern of gene expression between human duplicate genes. Genome Res 2003;13:1638–45. https://doi.org/10.1101/gr.1133803.

[19] Goldemberg J. Ethanol for a sustainable energy future. Science 2007;315:808–10. https://doi.org/10.1126/science.1137013.

[20] Souza GM, Ballester MVR, de Brito Cruz CH, Chum H, Dale B, Dale VH, et al. The role of bioenergy in a climate-changing world. Environ Dev 2017;23:57–64. https://doi.org/10.1016/j.envdev.2017.02.008.

[21] Jaiswal D, De Souza AP, Larsen S, LeBauer DS, Miguez FE, Sparovek G, et al. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. Nat Clim Change 2017;7:788–92. https://doi.org/10.1038/nclimate3410.

[22] Souza GM, Victoria RL, Joly CA, Verdade LM. Bioenergy & sustainability: bridging the gaps. Paris Cedex: Scientific Committee on Problems of the Environment (SCOPE); 2015.

[23] Kline KL, Msangi S, Dale VH, Woods J, Souza GM, Osseweijer P, et al. Reconciling food security and bioenergy: priorities for action. GCB Bioenergy 2017;9:557–76. https://doi.org/10.1111/gcbb.12366.

[24] Souza GM, Filho RM. Industrial biotechnology and biomass: what next for Brazil's future energy and chemicals?. Ind Biotechnol 2016;12:24–5. https://doi.org/10.1089/ind.2015.29016.gms.

[25] Carvalho-Netto OV, Bressiani JA, Soriano HL, Fiori CS, Santos JM, Barbosa GV, et al. The potential of the energy cane as the main biomass crop for the cellulosic industry. Chem Biol Technol Agric 2014;1:20. https://doi.org/10.1186/s40538-014-0020-2.

[26] Kandel R, Yang X, Song J, Wang J. Potentials, Challenges, and genetic and genomic resources for sugarcane biomass improvement. Front Plant Sci 2018;9. https://doi.org/10.3389/fpls.2018.00151.

[27] Matsuoka S, Kennedy AJ, Santos EGD dos, Tomazela A, L, Rubio LCS. Energy cane: its concept, development, characteristics, and prospects. Adv Botany 2014. https://doi.org/10.1155/2014/597275.

[28] Goldemberg J, Coelho ST, Guardabassi P. The sustainability of ethanol production from sugarcane. Energy Policy 2008;36:2086–97. https://doi.org/10.1016/j.enpol.2008.02.028.

[29] Alexander AG. The energy cane alternative. Amsterdam: Elsevier; 1985.

[30] da Silva JA. The importance of the wild cane saccharum spontaneum for bioenergy genetic breeding. Sugar Tech 2017;19:229–40. https://doi.org/10.1007/s12355-017-0510-1.

[31] Tew TL, Cobill RM. Genetic improvement of sugarcane (Saccharum spp.) as an energy crop. In: Vermerris W, editor. Genetic improvement of bioenergy crops, New York, NY: Springer New York; 2008, p. 273–94. https://doi.org/10.1007/978-0-387-70805-8_9.

[32] Kumar A, Kumar N, Baredar P, Shukla A. A review on biomass energy resources, potential, conversion and policy in India. Renew Sustain Energy Rev 2015;45:530–9. https://doi.org/10.1016/j.rser.2015.02.007.

[33] Wang L-P, Jackson PA, Lu X, Fan Y-H, Foreman JW, Chen X-K, et al. Evaluation of sugarcane × saccharum spontaneum progeny for biomass composition and yield components. Crop Sci 2008;48:951–61. https://doi.org/10.2135/cropsci2007.10.0555.

[34] Hammond BW. *Saccharum spontaneum* (Gramineae) in panama: the physiology and ecology of invasion. J Sustainable For 1999;8:23–38. https://doi.org/10.1300/J091v08n03_03.

[35] Samuels A, Alexander AG, Rios C, Garcia M. The production of energy cane in Puerto Rico: the Hatillo project. Proc Cong Int Soc Sugar Cane Technol 1984;3:14–7.

[36] Bischoff KP, Gravois KA, Reagan TE, Hoy JW, Kimbeng CA, LaBorde CM, et al. Registration of 'L 79–1002' sugarcane. J Plant Registrations 2008;2:211. https://doi.org/10.3198/jpr2007.12.0673crc.

[37] Rao PS, Davis H, Simpson S. New sugarcane cultivars and year round sugar and ethanol production with bagassebased cogeneration in Barbados and Guiana. Proc 2007 Congr 2007;26:1169–76.

[38] Santchurn D, Badaloo MGH, Zhou M, Labuschagne MT. Contribution of sugarcane crop wild relatives in the creation of improved varieties in Mauritius. Plant Genet Resour: Characterization Utilization 2019;17:151–63. https://doi.org/10.1017/S1479262118000552.

[39] Jackson P. Utilizing wild germplasm in sugarcane breeding - progress and prospects. ISSCT Joint, Okinawa: 2018, p. 12.

[40] Aitken K, Jingchuan L, Piperidis G, Qing C, Yuanhong F, Jackson P. Worldwide genetic diversity od Saccharum spontaneum and level of diversity captured in a sugarcane breeding program. ISSCT Joint, Okinawa: 2018, p. 44.

[41] López J, Jaimes H, Franco M, Ocampo I, Barrios R, Salazar F, et al. Development of transgenic sugarcane associate with increasing biomass, sugar and stress tolerance in Colombia. ISSCT Joint, Okinawa: 2018, p. 54.

[42] Terajima Y, Matsuoka M, Ujihara K, et al. The simultaneous production of sugar and biomass ethanol using high-biomass sugarcane derived from interspecific and intergeneric cross in Japan. Proceedings of the Biomass Asia Workshop, Tokyo: 2005.

[43] Terajima Y, Sugimoto A, Fukuhara S, et al. The feature of root growth and activity of a high yielding interspecific hybrid between Saccharum hybrid and S. spontaneum L. Proceedings of the 25th Congress of the International Society of Sugar Cane Technologists, 2005, p. 255–8.

[44] Sugimoto A, Terajima Y, Terauchi T, et al. Developing new types of sugarcane by hybridization between commercial sugarcane cultivars and wild relatives. Proceedings of the Symposium FAO RAP-NIAS, vol. 2012/1, Tsukuba: 2012, p. 11–24.

[45] Tippayawat A, Terajima Y, Werapon P, Sansayawichai T, Irei S, Sugimoto A, et al. Agronomic traits and root distribution of intergeneric F1 and BC1 hybrids between Saccharum spp. hybrid and Thai Erianthus Arundianaceus in north-east Thailand. ISSCT Joint, Okinawa 2018:33.

[46] de Morais LK, de Aguiar MS, de Albuquerque e Silva P, Câmara TMM, Cursi DE, Júnior ARF, et al. Breeding of Sugarcane. In: Cruz VMV, Dierig DA, editors. Industrial Crops, vol. 9, New York, NY: Springer New York; 2015, p. 29–42. https://doi.org/10.1007/978-1-4939-1447-0_2.

[47] de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics 2014;15:540. https://doi.org/10.1186/1471-2164-15-540.

[48] Riaño-Pachón DM, Mattiello L. Draft genome sequencing of the sugarcane hybrid SP80-3280. F1000Res 2017;6:861. https://doi.org/10.12688/f1000research10.12688/f1000research.11859.2.

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

[50] Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nat Genet 2018;50:1565–73. https://doi.org/10.1038/s41588-018-0237-2.

[51] Nascimento LC, Yanagui K, Jose J, Camargo ELO, Grassi MCB, Cunha CP, et al. Unraveling the complex genome of Saccharum spontaneum using Polyploid Gene Assembler. DNA Res 2019;26:205–16. https://doi.org/10.1093/dnares/dsz001.

[52] Souza GM, Van Sluys M-A, Lembke CG, Lee H, Margarido GRA, Hotta CT, et al. Assembly of the 373K gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. GigaScience 2019.

[53] Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS ONE 2014;9:. https://doi.org/10.1371/journal.pone.0088462e88462.

[54] Que Y, Su Y, Guo J, Wu Q, Xu L. A Global view of transcriptome dynamics during sporisorium scitamineum challenge in sugarcane by RNA-seq. PLoS ONE 2014;9:. https://doi.org/10.1371/journal.pone.0106476e106476.

[55] Vargas L, Santa Brígida AB, Mota Filho JP, de Carvalho TG, Rojas CA, Vaneechoutte D, et al. Drought tolerance conferred to sugarcane by association with gluconacetobacter diazotrophicus: a transcriptomic view of hormone pathways. PLoS ONE 2014;9:. https://doi.org/10.1371/journal.pone.0114744e114744.

[56] Vicentini R, Bottcher A, Brito M dos S, dos Santos AB, Creste S, Landell MG de A, et al. Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. PLOS ONE 2015;10:e0134909. https://doi.org/10.1371/journal.pone.0134909.

[57] Mattiello L, Riaño-Pachón DM, Martins MCM, da Cruz LP, Bassi D, Marchiori PER, et al. Physiological and transcriptional analyses of developmental stages along sugarcane leaf. BMC Plant Biol 2015. https://doi.org/10.1186/s12870-015-0694-z.

[58] Li M, Liang Z, Zeng Y, Jing Y, Wu K, Liang J, et al. novo analysis of transcriptome reveals genes associated with leaf abscission in sugarcane (Saccharum officinarum L.). BMC Genomics 2016. https://doi.org/10.1186/s12864-016-2552-2.

[59] Huang D-L, Gao Y-J, Gui Y-Y, Chen Z-L, Qin C-X, Wang M, et al. Transcriptome of high-sucrose sugarcane variety GT35. Sugar Tech 2016;18:520–8. https://doi.org/10.1007/s12355-015-0420-z.

[60] Santa Brigida AB, Rojas CA, Grativol C, de Armas EM, Entenza JOP, Thiebaut F, et al. Sugarcane transcriptome analysis in response to infection caused by Acidovorax avenae subsp. avenae. PLoS ONE 2016;11:. https://doi.org/10.1371/journal.pone.0166473e0166473.

[61] Dong M, Cheng G, Peng L, Xu Q, Yang Y, Xu J. Transcriptome analysis of sugarcane response to the infection by sugarcane steak mosaic virus (SCSMV). Trop Plant Biol 2017;10:45–55. https://doi.org/10.1007/s12042-016-9183-2.

[62] Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, et al. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics 2017;18. https://doi.org/10.1186/s12864-017-3757-8.

[63] Belesini AA, Carvalho FMS, Telles BR, de Castro GM, Giachetto PF, Vantini JS, et al. novo transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. Genet Mol Res 2017;16. https://doi.org/10.4238/gmr16028845.

[64] Hoang NV, Furtado A, O'Keeffe AJ, Botha FC, Henry RJ. Association of gene expression with biomass content and composition in sugarcane. PLoS ONE 2017;12:. https://doi.org/10.1371/journal.pone.0183417e0183417.

[65] Xu S, Wang J, Shang H, Huang Y, Yao W, Chen B, et al. Transcriptomic characterization and potential marker development of contrasting sugarcane cultivars. Sci Rep 2018;8. https://doi.org/10.1038/s41598-018-19832-x.

[66] Hoang NV, Furtado A, Thirugnanasambandam PP, Botha FC, Henry RJ. De novo assembly and characterizing of the culm-derived meta-transcriptome from the polyploid sugarcane genome based on coding transcripts. Heliyon 2018;4:. https://doi.org/10.1016/j.heliyon.2018.e00583e00583.

[67] McNeil MD, Bhuiyan SA, Berkman PJ, Croft BJ, Aitken KS. Analysis of the resistance mechanisms in sugarcane during Sporisorium scitamineum infection using RNA-seq and microscopy. PLoS ONE 2018;13:. https://doi.org/10.1371/journal.pone.0197840e0197840.

[68] Telles BR, Carvalho FM de S, Vantini J da S, Belesini AA, de Castro GM, Giachetto PF, et al. Prolonged water deficit reveals new profile of sugarcane gene expression and metabolic pathway related to tolerance. Sugar Tech 2019;21:451–61. https://doi.org/10.1007/s12355-018-0674-3.

[69] Nishiyama MY, Ferreira SS, Tang P-Z, Becker S, Pörtner-Taliana A, Souza GM. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. PLoS ONE 2014;9:. https://doi.org/10.1371/journal.pone.0107351e107351.

[70] Wai CM, Zhang J, Jones TC, Nagai C, Ming R. Cell wall metabolism and hexose allocation contribute to biomass accumulation in high yielding extreme segregants of a Saccharum interspecific F2 population. BMC Genomics 2017;18. https://doi.org/10.1186/s12864-017-4158-8.

[71] Kasirajan L, Hoang NV, Furtado A, Botha FC, Henry RJ. Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. Sci Rep 2018;8. https://doi.org/10.1038/s41598-018-30033-4.

[72] Schaker PDC, Palhares AC, Taniguti LM, Peters LP, Creste S, Aitken KS, et al. RNAseq transcriptional profiling following whip development in sugarcane smut disease. PLoS ONE 2016;11:. https://doi.org/10.1371/journal.pone.0162237e0162237.

[73] HISAT: a fast spliced aligner with low memory requirements n.d. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4655817/ (accessed May 29, 2019).

[74] Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, et al. Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev 2005;19:1855–60. https://doi.org/10.1101/gad.1331305.

[75] Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell 2005;17:2993–3006. https://doi.org/10.1105/tpc.105.036004.

[76] Zhong R, Demura T, Ye Z-H. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell 2006;18:3158–70. https://doi.org/10.1105/tpc.106.047399.

[77] Zhong R, Richardson EA, Ye Z-H. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. Plant Cell 2007;19:2776–92. https://doi.org/10.1105/tpc.107.053678.

[78] McCarthy RL, Zhong R, Ye Z-H. MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell Physiol 2009;50:1950–64. https://doi.org/10.1093/pcp/pcp139.

[79] Wang H, Zhao Q, Chen F, Wang M, Dixon RA. NAC domain function and transcriptional control of a secondary cell wall master switch: NST1 function and regulation. Plant J 2011;68:1104–14. https://doi.org/10.1111/j.1365-313X.2011.04764.x.

[80] Zhong R, Lee C, Ye Z-H. Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. Trends Plant Sci 2010;15:625–32. https://doi.org/10.1016/j.tplants.2010.08.007.

[81] Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. Front Plant Sci 2015;6:288. https://doi.org/10.3389/fpls.2015.00288.

[82] Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. Plant Physiol 2011;155:916–31. https://doi.org/10.1104/pp.110.168641.

[83] Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye Z-H. Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. Plant Cell Physiol 2011;52:1856–71. https://doi.org/10.1093/pcp/pcr123.

[84] Valdivia ER, Herrera MT, Gianzo C, Fidalgo J, Revilla G, Zarra I, et al. Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot Brachypodium distachyon. J Exp Bot 2013;64:1333–43. https://doi.org/10.1093/jxb/ers394.

[85] Yoshida K, Sakamoto S, Kawai T, Kobayashi Y, Sato K, Ichinose Y, et al. Engineering the Oryza sativa cell wall with rice NAC transcription factors regulating secondary wall formation. Front Plant Sci 2013;4. https://doi.org/10.3389/fpls.2013.00383.

[86] Zhong R, Yuan Y, Spiekerman JJ, Guley JT, Egbosiuba JC, Ye Z-H. Functional characterization of NAC and MYB transcription factors involved in regulation of biomass production in switchgrass (Panicum virgatum). PLoS ONE 2015;10:. https://doi.org/10.1371/journal.pone.0134611e0134611.

[87] Noda S, Koshiba T, Hattori T, Yamaguchi M, Suzuki S, Umezawa T. The expression of a rice secondary wall-specific cellulose synthase gene, OsCesA7, is directly regulated by a rice transcription factor, OsMYB58/63. Planta 2015;242:589–600. https://doi.org/10.1007/s00425-015-2343-z.

[88] Scully ED, Gries T, Sarath G, Palmer NA, Baird L, Serapiglia MJ, et al. Overexpression of *SbMyb60* impacts phenylpropanoid biosynthesis and alters

secondary cell wall composition in *Sorghum bicolor*. Plant J 2016;85:378–95. https://doi.org/10.1111/tpj.13112.

[89] Golfier P, Volkert C, He F, Rausch T, Wolf S. Regulation of secondary cell wall biosynthesis by a NAC transcription factor from Miscanthus. Plant Direct 2017;1:. https://doi.org/10.1002/pld3.24e00024.

[90] Xiao Wenhan, Yang Yue, Yu Jingjuan. ZmNST3 and ZmNST4 are master switches for secondary wall deposition in maize (Zea mays L.). Plant Sci 2018;266:83–94. https://doi.org/10.1016/j.plantsci.2017.03.012.

[91] Rao X, Chen X, Shen H, Ma Q, Li G, Tang Y, et al. Gene regulatory networks for lignin biosynthesis in switchgrass *(Panicum virgatum)*. Plant Biotechnol J 2019;17:580–93. https://doi.org/10.1111/pbi.13000.

[92] Agarwal T, Grotewold E, Doseff AI, Gray J. MYB31/MYB42 Syntelogs Exhibit Divergent Regulation of Phenylpropanoid Genes in Maize, Sorghum and Rice. Sci Rep 2016;6:28502. https://doi.org/10.1038/srep28502.

[93] Ooka H. Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana. DNA Res 2003;10:239–47. https://doi.org/10.1093/dnares/10.6.239.

[94] Vettore AL. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 2003;13:2725–35. https://doi.org/10.1101/gr.1532103.

[95] Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 2007;315:804–7. https://doi.org/10.1126/science.1137016.

[96] Chen F, Dixon RA. Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 2007;25:759–61. https://doi.org/10.1038/nbt1316.

[97] Vanholme R, Morreel K, Ralph J, Boerjan W. Lignin engineering. Curr Opin Plant Biol 2008;11:278–85. https://doi.org/10.1016/j.pbi.2008.03.005.

[98] Zhao X, Zhang L, Liu D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels, Bioprod Biorefin 2012;6:465–82. https://doi.org/10.1002/bbb.1331.

[99] Zeng Y, Zhao S, Yang S, Ding S-Y. Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. Curr Opin Biotechnol 2014;27:38–45. https://doi.org/10.1016/j.copbio.2013.09.008.

[100] Bonawitz ND, Chapple C. The genetics of lignin biosynthesis: connecting genotype to phenotype. Annu Rev Genet 2010;44:337–63. https://doi.org/10.1146/annurev-genet-102209-163508.

[101] Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, et al. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in arabidopsis. Science 2013;341:1103–6. https://doi.org/10.1126/science.1241602.

[102] Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. Nat Plants 2016;2:16050. https://doi.org/10.1038/nplants.2016.50.

[103] Barros J, Escamilla-Trevino L, Song L, Rao X, Serrani-Yarce JC, Palacios MD, et al. 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. Nat Commun 2019;10:1994. https://doi.org/10.1038/s41467-019-10082-7.

[104] Fraser CM, Chapple C. The Phenylpropanoid Pathway in Arabidopsis. The Arabidopsis Book 2011;9:. https://doi.org/10.1199/tab.0152e0152.

[105] Ferreira SS, Simões MS, Carvalho GG. de Lima LGA, Svartman RM de A, Cesarino I. The lignin toolbox of the model grass Setaria viridis. Plant Mol Biol 2019. https://doi.org/10.1007/s11103-019-00897-9.

[106] Perkins M, Smith RA, Samuels L. The transport of monomers during lignification in plants: anything goes but how?. Curr Opin Biotechnol 2019;56:69–74. https://doi.org/10.1016/j.copbio.2018.09.011.

[107] Alejandro S, Lee Y, Tohge T, Sudre D, Osorio S, Park J, et al. AtABCG29 is a monolignol transporter involved in lignin biosynthesis. Curr Biol 2012;22:1207–12. https://doi.org/10.1016/j.cub.2012.04.064.

[108] Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol 2008;11:301–7. https://doi.org/10.1016/j.pbi.2008.03.002.

[109] Carpita NC. Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol 1996;47:445–76. https://doi.org/10.1146/annurev.arplant.47.1.445.

[110] Harris PJ, Trethewey JAK. The distribution of ester-linked ferulic acid in the cell walls of angiosperms. Phytochem Rev 2010;9:19–33. https://doi.org/10.1007/s11101-009-9146-4.

[111] Ralph J, Grabber JH, Hatfield RD. Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. Carbohydr Res 1995;275:167–78. https://doi.org/10.1016/0008-6215(95)00237-N.

[112] Lam TBT, Kadoya K, Iiyama K. Bonding of hydroxycinnamic acids to lignin: ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the β-position, in grass cell walls. Phytochemistry 2001;57:987–92. https://doi.org/10.1016/S0031-9422(01)00052-8.

[113] Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan D, Helmich KE, et al. Monolignol ferulate conjugates are naturally incorporated into plant lignins. Sci Adv 2016;2:. https://doi.org/10.1126/sciadv.1600393e1600393.

[114] Lan W, Lu F, Regner M, Zhu Y, Rencoret J, Ralph SA, et al. Tricin, a flavonoid monomer in monocot lignification. Plant Physiol 2015;167:1284–95. https://doi.org/10.1104/pp.114.253757.

[115] Eudes A, Dutta T, Deng K, Jacquet N, Sinha A, Benites VT, et al. SbCOMT (Bmr12) is involved in the biosynthesis of tricin-lignin in sorghum. PLoS ONE 2017;12:. https://doi.org/10.1371/journal.pone.0178160e0178160.

[116] Terrett OM, Dupree P. Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. Curr Opin Biotechnol 2019;56:97–104. https://doi.org/10.1016/j.copbio.2018.10.010.

[117] Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, et al. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science 2006;311:1940–2. https://doi.org/10.1126/science.1122975.

[118] Yin Y, Huang J, Xu Y. The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol 2009;9:99. https://doi.org/10.1186/1471-2229-9-99.

[119] Doblin MS, Pettolino FA, Wilson SM, Campbell R, Burton RA, Fincher GB, et al. A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis. Proc Natl Acad Sci USA 2009;106:5996–6001. https://doi.org/10.1073/pnas.0902019106.

[120] Handakumbura PP, Brow K, Whitney IP, Zhao K, Sanguinet KA, Lee SJ, et al. Secondary wall associated MYB1 is a positive regulator of secondary cell wall thickening in Brachypodium distachyon and is not found in the Brassicaceae. Plant J 2018;96:532–45. https://doi.org/10.1111/tpj.14047.

[121] Ha CM, Escamilla-Trevino L, Yarce JCS, Kim H, Ralph J, Chen F, et al. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in Medicago truncatula. Plant J 2016;86:363–75. https://doi.org/10.1111/tpj.13177.

[122] Saleme M de LS, Cesarino I, Vargas L, Kim H, Vanholme R, Goeminne G, et al. Silencing CAFFEOYL SHIKIMATE ESTERASE Affects Lignification and Improves Saccharification in Poplar. Plant Physiol 2017;175:1040–57. https://doi.org/10.1104/pp.17.00920.

[123] Ohtani M, Demura T. The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model. Curr Opin Biotechnol 2019;56:82–7. https://doi.org/10.1016/j.copbio.2018.10.002.

[124] Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, et al. Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. Nature 2014;509:376–80. https://doi.org/10.1038/nature13084.

[125] Dolan WL, Dilkes BP, Stout JM, Bonawitz ND, Chapple C. Mediator complex subunits MED2, MED5, MED16, and MED23 genetically interact in the regulation of phenylpropanoid biosynthesis. Plant Cell 2017;29:3269–85. https://doi.org/10.1105/tpc.17.00282.

[126] Yilmaz A, Nishiyama MY, Fuentes BG, Souza GM, Janies D, Gray J, et al. GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiol 2009;149:171–80. https://doi.org/10.1104/pp.108.128579.

[127] Ruiz M, Rouard M, Raboin LM, Lartaud M, Lagoda P, Courtois B. TropGENE-DB, a multi-tropical crop information system. Nucleic Acids Res 2004;32:D364–7. https://doi.org/10.1093/nar/gkh105.

[128] Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012;40:D1178–86. https://doi.org/10.1093/nar/gkr944.

[129] Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2017;45:D1040–5. https://doi.org/10.1093/nar/gkw982.

[130] Makita Y, Shimada S, Kawashima M, Kondou-Kuriyama T, Toyoda T, Matsui M. MOROKOSHI: transcriptome database in sorghum bicolor. Plant Cell Physiol 2015;56:e6. https://doi.org/10.1093/pcp/pcu187.

[131] Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: the United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 2018;36:566–9. https://doi.org/10.1038/nbt.4163.

[132] Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, et al. Gramene: a resource for comparative grass genomics. Nucleic Acids Res 2002;30:103–5. https://doi.org/10.1093/nar/30.1.103.

[133] Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, et al. 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res 2018;46:D1190–6. https://doi.org/10.1093/nar/gkx1002.

[134] Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, et al. PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res 2008;36:D959–65. https://doi.org/10.1093/nar/gkm1041.

[135] Nishiyama MY, Vicente FFR, Lembke CG, Sato PM, Dal-Bianco ML, Fandino RA, et al. The SUCEST-FUN regulatory network database: designing an energy grass. Int Sugar J 2012;114:821–6.

[136] Papini-Terzi FS, Rocha FR, Vêncio RZN, Oliveira KC, Felix J de M, Vicentini R, et al. Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Res 2005;12:27–38. https://doi.org/10.1093/dnares/12.1.27.

[137] Rocha FR, Papini-Terzi FS, Nishiyama MY, Vêncio RZ, Vicentini R, Duarte RD, et al. Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. BMC Genomics 2007;8:71. https://doi.org/10.1186/1471-2164-8-71.

[138] Lembke CG, Nishiyama MY, Sato PM, de Andrade RF, Souza GM. Identification of sense and antisense transcripts regulated by drought in sugarcane. Plant Mol Biol 2012;79:461–77. https://doi.org/10.1007/s11103-012-9922-1.

[139] Hotta CT, Nishiyama MY, Souza GM. Circadian rhythms of sense and antisense transcription in sugarcane, a highly polyploid crop. PLoS ONE 2013;8:. https://doi.org/10.1371/journal.pone.0071847e71847.

[140] De Souza AP, Gaspar M, Da Silva EA, Ulian EC, Waclawovsky AJ, Nishiyama MY, et al. Elevated CO2 increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane. Plant Cell Environ 2008;31:1116–27. https://doi.org/10.1111/j.1365-3040.2008.01822.x.

[141] Ferreira SS, Hotta CT, Poelking VG de C, Leite DCC, Buckeridge MS, Loureiro ME, et al. Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. Plant Mol Biol 2016;91:15–35. https://doi.org/10.1007/s11103-016-0434-2.

[142] Cia MC, de Carvalho G, Azevedo RA, Monteiro-Vitorello CB, Souza GM, Nishiyama-Junior MY, et al. Novel insights into the early stages of ratoon stunting disease of sugarcane inferred from transcript and protein analysis. Phytopathology 2018;108:1455–66. https://doi.org/10.1094/PHYTO-04-18-0120-R.

[143] Cunha CP, Roberto GG, Vicentini R, Lembke CG, Souza GM, Ribeiro RV, et al. Ethylene-induced transcriptional and hormonal responses at the onset of sugarcane ripening. Sci Rep 2017;7. https://doi.org/10.1038/srep43364.