Screen of Genes Linked to High-Sugar Content in Stems by Comparative Genomics

Martín Calviño · Rémy Bruggmann · Joachim Messing

Abstract One of the great advantages of the fully sequenced rice genome is to serve as a reference for other cereal genomes in particular for identifying genes linked to unique traits. A trait of great interest is reduced lignocellulose in the stem of related species in favor of fermentable sugars as a source of biofuels. While sugarcane is one of the most efficient biofuel crops, little is known about the underlying gene repertoire involved in it. Here, we take advantage of the natural variation of sweet and grain sorghum to uncover genes that are conserved in rice, sorghum, and sugarcane but differently expressed in sweet versus grain sorghum by using a microarray platform and the syntenous alignment of rice and sorghum genomic regions containing these genes. Indeed, enzymes involved in carbohydrate accumulation and those that reduce lignocellulose can be identified.

Keywords Integrative genetics · Sugar accumulation · Cell wall synthesis · Microarray analysis · Synteny · Biofuel crops

Introduction

Comparisons of genetic maps and sequences of several grass species have shown that there is global conservation of gene content and order [11]. Therefore, grasses have been considered as a "single genetic system" [2]. The practical aspect of such a concept is of great importance for agronomical purposes because a useful trait in one species could be transferred to another. A relevant example could be carbohydrate partitioning and allocation. In cereals such as wheat, sorghum, and rice, the process of grain filling demands carbon from photosynthesis assimilation as well as the remobilization of pre-stored carbohydrates in the stem before and after anthesis [45]. It has been estimated that about 30% of the total yield in rice depends on the carbohydrate content accumulated in the stem before heading [18]. For these reasons, characterization of genes involved in carbohydrate metabolism and accumulation can lead to the development of improved cereal crops.

In recent years, there has been an increasing demand on biomass for the production of ethanol as a renewable resource for fuel. The biggest producers of ethanol in the world are Brazil and the USA [34]. In Brazil, it is derived from sugarcane, while in the USA, ethanol is derived from the grain of corn. Because of the use of the entire plant as a source for fermentable sugars, carbohydrate accumulation and partitioning has been extensively studied in sugarcane, probably more than in any other species [30]. However, genes involved in these processes cannot easily be identified because of the complex genome of sugarcane, with several cultivars differing greatly in their ploidy levels from $2n=100$ to $2n=130$ chromosomes [9, 13]. Even if one could make further improvements to sugarcane, it has the disadvantage of being a crop restricted to tropical growing areas.

On the other hand, the use of corn grain for ethanol production poses a major conflict because of its dual use as food and fuel. Therefore, it has been proposed to use grain solely for food and only the stover as a source for ethanol. A major impediment to this approach is that, in contrast to sugarcane, corn stover consists mainly of lignocellulose, which is more costly to process than fermentable sugars [8].

M. Calviño · R. Bruggmann · J. Messing (✉)
Waksman Institute of Microbiology, Rutgers University,
190 Frelinghuysen Road,
Piscataway, NJ 08854-8020, USA
e-mail: messing@waksman.rutgers.edu

Therefore, it would be attractive to identify corn varieties with reduced lignocellulose. Interestingly, there is extensive natural intra-species variation for sugar content in sorghum with cultivars that do not accumulate sugars (referred to as grain sorghums) in contrast to those that accumulate large amounts of sugars in their stems [16]. Such intra-species variation can serve as a platform to identify genes linked to increased sugar content and reduced lignocellulose [4]. Moreover, if these genes are conserved by ancestry in related species, one could envision the introduction of such a trait by the import of specific regulatory regions. Conservation of gene order between closely related species permits the alignment of orthologous chromosomal segments. Non-collinear genes would constitute paralogous copies [29]. To facilitate such alignments, the use of rice with one of the smallest cereal genomes that has been sequenced [17] increasingly becomes the anchor genome for other grasses [28]. In this sense, we can use rice as a reference genome for biofuel crops such as sugarcane and sorghum.

While rice offers an excellent reference as a compact genome from an evolutionary point of view, it is less suitable as a reference for a phenotype of reduced lignocellulose. Moreover, rice is a bambusoid C3 cereal plant, and sorghum and sugarcane are panicoid C4 cereal plants, which branched out 50 mya [23]. Sorghum and sugarcane belong to the Saccharinae clade and diverged from each other only 8–9 mya [14, 20]. Therefore, sugarcane and its reduced lignocellulose can serve as a trait reference for sorghum varieties that differ in the cellulose content of their stems. Consequently, we took advantage of a GeneChip that was created to study gene expression in sugarcane and its role in the accumulation of sugar in the stem during development [6] for the comparison of grain and sweet sorghum genes. One would expect that sweet sorghum and sugarcane use similar gene products for enhanced sugar accumulation in their stems. Indeed, we not only identified genes involved in sugar accumulation and lignocellulose synthesis, whose expression levels are correlated with the trait, but also demonstrate their ancestry through the alignment of orthologous regions of the rice and sorghum genomes. Therefore, the same genes could also be used to improve other biofuel crops like switchgrass and *Miscanthus*, validating a translational genomic approach.

Results

Sugar accumulation in the stem of grain and sweet sorghum cultivars

Previous reports have indicated that in sorghum stems, sugars start to accumulate at flowering stage [16, 25]. We compared the accumulation of sugars in the stem between six

sweet sorghum lines (Dale, Della, M81-E, Rio, Top76-6, and Simon) and one line from grain sorghum (BTx623). As an estimation of the total amount of sugars present in the juice of sorghum stems, we measured the Brix degree of each internode along the main stem at the time of flowering. We found great variation in flowering time as well as in Brix degree between the sweet sorghum lines when compared to grain sorghum BTx623 (Fig. 1a,b). In general, the Brix degree was lower in the mature and immature internodes of the stem, in contrast to maturing internodes. These findings are in agreement with previous studies [16, 25]. Consistent with the inability of grain sorghum to accumulate significant levels of sugars in the stem, the Brix degree in BTx623 was low and remained fairly constant for all the internodes along the stem. Among the sweet sorghum cultivars, Rio had the highest Brix degree and Simon the lowest. Furthermore, the difference in flowering time between BTx623 and Rio was smaller than in the rest of sweet sorghum lines with high Brix degrees. For this reason, we decided to perform a comparative analysis of transcripts in the stem of the Rio and BTx623 sorghum lines.
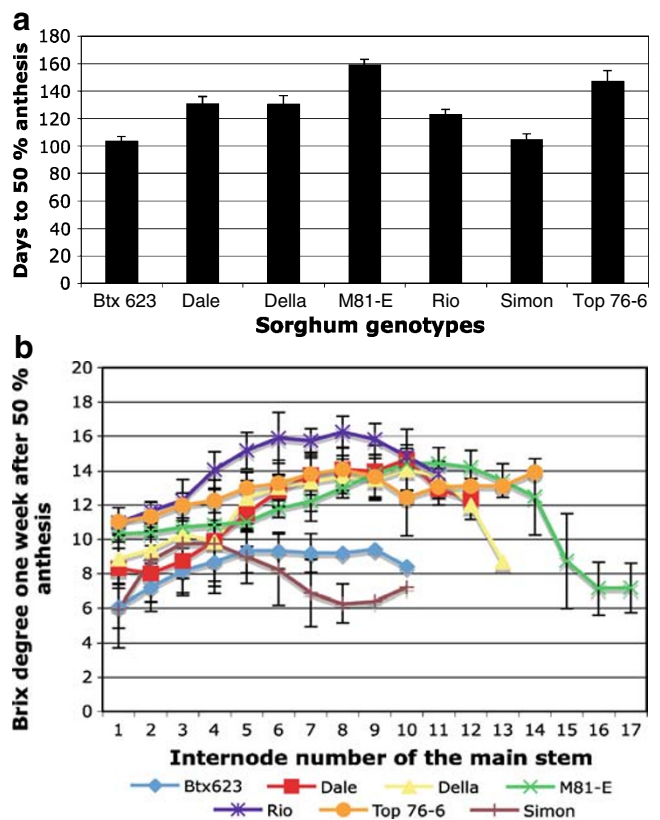

Fig. 1 Variation in flowering time and Brix degree. **a** Comparison of flowering time between grain sorghum BTx623 and six sweet sorghum cultivars. Time to flowering was measured as days required reaching 50% anthesis. **b** Comparison of Brix degree along the main stem between grain sorghum BTx623 and six sweet sorghum genotypes. The Brix degree was measured for each internode, and the average of a triplicate experiment was plotted.

Microarray analysis of transcripts from sorghum stem tissues

In order to identify genes expressed in the stem with a potential role in sugar accumulation and reduced lignocellulose [4], we compared transcript profiles between grain (BTx623) and sweet sorghum (Rio). Such a genome-wide analysis became possible because of the recently designed GeneChip of sugarcane [6]. This array was specifically developed with sequences that were obtained from several cDNA libraries representing distinct tissue types including stem from 15 sugarcane varieties. The use of this GeneChip permitted us to directly compare gene expression data of two different sorghum cultivars with the previously generated data from sugarcane. Three independent plants for each BTx623 and Rio were grown until anthesis and RNA was extracted from the same maturing internode for all six plants. These RNAs were used to prepare biotylinated cRNAs for hybridization, each sample separately hybridized to one array.

The sugarcane array comprised 8,224 probe sets, of which more than 70% (5,900) gave a positive signal with sorghum RNA samples. When a twofold cut-off value was applied as criterion to distinguish differentially expressed transcripts between grain and sweet sorghum, a total of 195 transcripts were identified, with 132 transcripts being down-regulated and 63 transcripts up-regulated in Rio, respectively (Electronic Supplementary Material Tables S1 and S2). Because some probe sets identify the same gene, the number of genes that is down-regulated is 103 and up-regulated 51, respectively. Based on the annotation of the sorghum genes, we were able to infer the possible function for most of the differentially expressed transcripts.

Among the transcripts that were up-regulated in Rio, a saposin-like type B gene displayed the highest differential expression. Saposins are involved in the degradation of sphingolipids [31]. Other transcripts encoding stress-related proteins such as heat shock protein 70 (HSP70) and HSP90 were up-regulated, consistent with an osmotic stress imposed by high concentration of sugars (Electronic Supplementary Material Tables S1 and S2) [3]. Our results show that in Rio, down-regulated genes outnumber those that are up-regulated by a factor of 2. The most reduced transcript has a fasciclin domain. This domain has been shown to be involved in cell adhesion (Table 1) [10, 22].

Genes with altered expression in carbohydrate metabolism in sweet sorghum

Based on Gene Ontology (GO) terms (http://www.geneon tology.org/), the sucrose and starch metabolic pathway from the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/), and the Carbohydrate-Active enzymes (CAZy) database (http://www.cazy.org/), we found that almost 16% of the transcripts that were differentially expressed between BTx623 and Rio corresponded to transcripts affecting carbohydrate metabolism (Tables 1 and 2). Based on the link between hypothetical function and the sweet sorghum trait, we selected 37 candidate genes, of which differential expression could be considered to be the cause or the consequence of increased sugar and decreased lignocellulose in sorghum stems. Clearly such a screen would not detect candidate genes that might play a qualitative rather than a quantitative role. For instance, one might expect that the role of sucrose phosphate synthase in carbohydrate metabolism should be differentially expressed, but it appears to be unchanged under the parameters examined. We therefore can discover mainly genes that are differentially expressed. Extrinsic evidence for the link between these candidate genes and their potential function is described in more detail in "Discussion."

Among these, transcripts that were up-regulated include hexokinase 8 and carbohydrate phosphorylase (starch and sucrose metabolism), nicotinamide adenine dinucleotide phosphate (NADP) malic enzyme (C4 photosynthesis), a D-mannose binding lectin (sugar binding), and a lysin motif (LysM) domain protein possibly involved in cell wall degradation. Transcripts that were down-regulated included sucrose synthase 2 and fructokinase 2 (starch and sucrose metabolism), alpha-galactosidase and beta-galactosidase (hydrolysis of glycosidic bonds), and cellulose synthase 1, 7, and 9 together with cellulose synthase catalytic subunit 12 (cell wall metabolism). In addition, several others transcripts with a cell-wall-related role that were down-regulated included cinnamoyl CoA reductase, cinnamyl alcohol dehydrogenase, 4-coumarate coenzyme A ligase, caffeoyl-CoA O-methyltransferase, xyloglucan endo-transglycosylase/hydrolase, peroxidase and phenylalanine, and histidine ammonia-lyase.

Validation of microarray data by quantitative reverse transcription polymerase chain reaction

To validate the data obtained by microarray analysis, we randomly sampled 14 of the 37 candidate genes and compared their expression levels in both Rio and BTx623 by performing quantitative reverse transcription polymerase chain reaction (qRT-PCR; Fig. 2a). In Rio, the expression of saposin, carbohydrate phosphorylase, hexokinase-8, and NADP malic enzyme is up-regulated in comparison with their expression in BTx623. In contrast, the expression of fasciclin-like protein FLA15, cellulose synthase 1 and 7, fructokinase-2, 4-coumarate coenzyme A ligase, sucrose synthase 2, laccase, cinnamoyl CoA reductase, beta-galactosidase 3 precursor, and alpha-galactosidase precursor were down-regulated in Rio. Although the levels of gene

Table 1 List of "Trait-Specific" Genes that Are Syntenic with Rice

Gene[a]	Rice	Sorghum	Expression[b]
Starch and sucrose metabolism			
Hexokinase 8	Os01g0190400	Sb03g003190.1[c]	2.3
Hexokinase 8	*Os05g0187100*	*Sb09g005840.1*	
Carbohydrate phosphorylase	Os01g0851700	Sb03g040060.1[c]	1.2
Sucrose synthase 2	Os03g0401300	**Sb01g033060.1[c]**	−1.3
Sucrose synthase 2	*Os07g0616800*		
Fructokinase-2	Os08g0113100	Sb07g001320.1[c]	−1.7
Sorbitol dehydrogenase	Os08g0545200	Sb07g025220.1[c]	1.6
Sugar binding			
D-mannose binding lectin	Os06g0165200	Sb10g022730.1[c]	2
CO_2 assimilation			
NADP-dependent malic enzyme	Os01g0723400	**Sb03g033250.1[c]**	2
Cell-wall-related			
LysM domain protein/cell wall catabolism	Os03g0110600	Sb01g049890.1[c]	1.2
Cellulose synthase-7	Os03g0837100	**Sb01g002050.1[c]**	−1
Cellulose synthase-1	Os05g0176100	**Sb09g005280.1[c]**	−1.1
Cellulose synthase-9	Os07g0208500	**Sb02g006290.1[c]**	−1.1
Cellulose synthase-9	*Os03g0808100*	*Sb01g004210.1*	
Cellulose synthase catalytic subunit 12	Os09g0422500	Sb02g025020.1[c]	−4.7
Alpha-galactosidase precursor	Os10g0493600	Sb01g018400.1[c]	−1.8
Beta-galactosidase 3 precursor	Os01g0875500	Sb03g041450.1[c]	−2.4
Beta-galactosidase 3 precursor	*Os05g0428100*	*Sb03g041450.1*	
Cinnamoyl CoA reductase	Os08g0441500	Sb07g021680.1[c]	−2.9
Cinnamoyl CoA reductase	*Os09g0419200*	*Sb10g005700.1*	
Laccase	Os01g0842400	Sb03g039520.1	−3.5
4-Coumarate coenzyme A ligase	Os02g0177600	Sb04g005210.1[c]	−3.7
4-Coumarate coenzyme A ligase	*Os06g0656500*	*Sb10g026130.1*	
Fasciclin domain	Os03g0788600	Sb01g005770.1[c]	−1.75
Fasciclin domain	*Os07g0160600*	*Sb02g003410.1*	
Fasciclin-like protein FLA15	Os05g0563600	Sb09g028490.1[c]	−6.5
Caffeoyl-CoA *O*-methyltransferase 2	Os06g0165800	**Sb10g004540.1[c]**	−2.15
Caffeoyl-CoA *O*-methyltransferase	Os08g0498100	**Sb07g028530.1[c]**	−5.3
Caffeoyl-CoA O-methyltransferase	*Os09g0481400*	*Sb02g027930.1*	

In boldface: sorghum genes that correspond to sugarcane probe set IDs previously reported by [6]

[a] Paralogs in italics

[b] Mean Log2 Ratio of sweet versus grain sorghum

[c] Sorghum gene to which a sugarcane probe set was mapped

expression between the microarray and the qRT-PCR method differ to some extent, there is no difference in the classification of up- or down-regulated genes. This 100% correspondence of microarray with qRT-PCR data illustrates that the microarray platform can be used as an effective method for screening large amounts of genes for a particular trait across closely related species. Before one would embark on any further experimentation, a much smaller candidate gene set can then be tested by more labor-intensive methods for gene expression between cultivars of the same species. In order to see if the expression difference between BTx623 and Rio for the transcripts encoding a saposin-type B protein and a fasciclin-like protein FLA15 also extended to other sweet sorghum lines, we extracted RNA from maturing stems of BTx623, Dale, and Della at

flowering and measured the expression of both genes by qRT-PCR. We found that the saposin-type B gene is also highly expressed in Dale and Della when compared to grain sorghum and that the opposite is true for the expression of fasciclin-like protein FLA15, highly expressed in BTx623 compared to Dale and Della (Fig. 2b).

Genomic location of differentially expressed genes

In order to see if genes that were differentially expressed between grain and sweet sorghum cluster together in a particular region of the sorghum genome, we generated a "transcriptome map" (Fig. 3). We mapped the sequences of all up- and down-regulated sugarcane probes to the recently sequenced sorghum genome (BTx623; http://www.phyto

Table 2 List of "Trait-Specific" Genes that Are Not Syntenic with Rice

In boldface: sorghum genes that correspond to sugarcane probe set IDs previously reported by [6]

[a] Mean Log2 Ratio of sweet versus grain sorghum

Gene	Sorghum	Expression[a]
Cell-wall-related		
Alcohol dehydrogenase	**Sb10g006290**	1
Cinnamyl alcohol dehydrogenase	Sb04g011550	−1.5
Dolichyl-diphospho-oligosaccharide	Sb02g006330	−1.4
Xyloglucan endo-transglycosylase/hydrolase	Sb06g015880	−1.1
Putative Xylanase inhibitor	Sb05g027350	−1.5
Putative Xylanase inhibitor	Sb02g004660	−1.5
Glycoside hydrolase family 1	**Sb02g029640**	−1.1
Phenylalanine and histidine ammonia-lyase	Sb04g026520	−2
Peroxidase	Sb02g037840	−1.5
Similar to Saposin type B protein	Sb09g013990	5.7

zome.net/cgi-bin/gbrowse/sorghum/) using GenomeThreader [12]. From a total of 195 probe sets, 176 could be mapped to the sorghum genome based on their alignment with a sorghum gene ("Materials and methods"). In addition, six probe sets could be mapped to the genome but do not correlate with the current sorghum gene annotation, and for another 13 probe sets, we were not able to map them to the sorghum genome. Genes that were differentially expressed between grain and sweet sorghum do not appear to cluster in any particular region of the genome but rather reflect random distribution (Fig. 3).

Trait-specific syntenic gene pairs between rice and sorghum

It can be considered that important gene functions have been conserved by ancestry and that divergence is mainly due to changes in regulatory control regions of genes. To determine the ancestry of genes, however, requires the alignment of syntenic regions. Because we know now the positions of the sorghum genes in their respective chromosomes, we can align them with the rice genome as a reference [17] and determine whether the aligned regions are collinear between rice and sorghum. Indeed, we found that from a total of 154 differentially expressed sorghum genes, 123 have an orthologous copy in syntenic positions in rice (Electronic Supplementary Material Table S1). This collection includes 28 candidate genes for the sweet sorghum trait (Table 1). Interestingly, one of these candidate genes, sucrose synthase 2, is duplicated in rice but not in grain sorghum. So the question arose whether the gene is duplicated in sweet sorghum thereby explaining the difference in gene expression simply by gene duplication. Because we have only the sequence of grain sorghum, we performed a Southern blot analysis of genomic DNA of sweet sorghum. When genomic DNA from BTx623 and Rio are compared, both possess a single copy of sucrose synthase 2 (data not shown). Therefore, it is unlikely that gene duplication plays a role in changing the level of gene expression.

Discussion

Translational genomics

The non-renewable nature of fossil oil imposes an increasing pressure to develop alternative energies in order to support and secure social and economic growth in the near future [34]. Currently, there is a worldwide interest to develop biofuel crops, the best example being sugarcane, used in Brazil since the 1970s. Besides sugarcane, other grasses such as *Brachypodium distachyon*, *Miscanthus*, maize, rice, sweet sorghum, and switchgrass are considered as crops for biofuel research and production. Recently, the entire gene cluster of ten sorghum kafirin genes contained within a chromosomal segment of 45 kb was intact and stably inserted into the maize genome. Expression analysis then has shown that kafirins accumulated in maize endosperm in a developmental and tissue-specific manner [38]. Such transfer of genomic DNA between species that cannot be crossed could then be used in advanced breeding techniques to introduce desirable traits from one species to another. Here, we integrate the traits of sugar accumulation and lignocellulose content with genomic and expression data of the three species, sugarcane, sorghum, and rice. We used the recently developed Affymetrix sugarcane genome array [6] as a tool for the identification of genes differentially expressed in maturing stems of grain and sweet sorghum. The intra-species variation for sugar content in sorghum is more pronounced than between sugarcane varieties, making sorghum a more suitable model to study this trait. On the other hand, because we can map sorghum genes to their chromosomal positions, we can use rice as a reference genome to identify genes by their ancestry.

Cross-referencing tissue-specific transcripts

Sorghum and sugarcane belong to the *Saccharinae* clade and diverged from each other only 8 to 9 mya [20], while rice is a more distant relative and separated from this clade
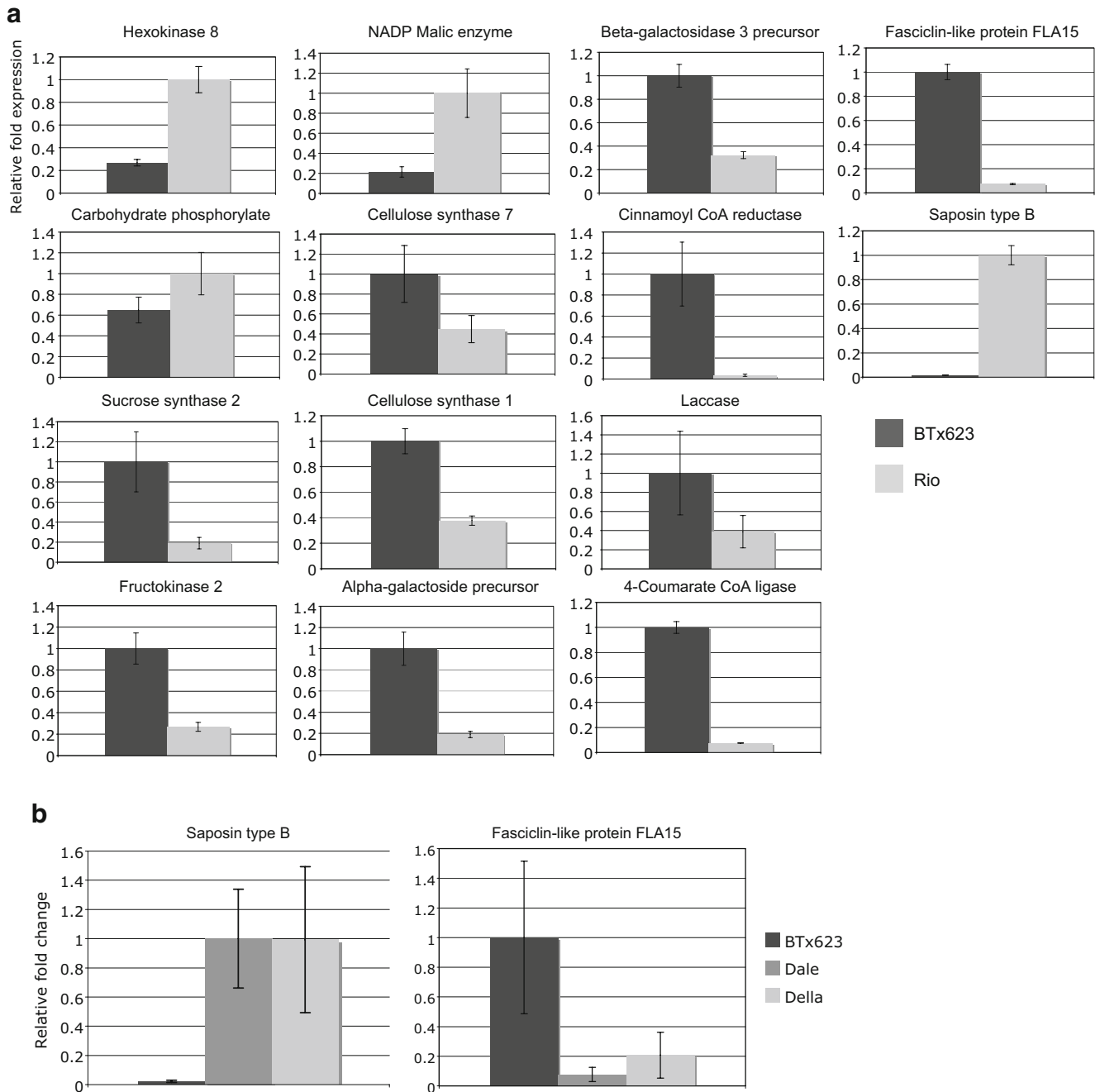
a

b

Fig. 2 Validation of microarray data by quantitative RT-PCR. **a** The expression of 14 genes selected from Table 1 was analyzed through quantitative RT-PCR. The results of three independent experiments for both BTx623 and Rio are shown. The quantification of the mRNA abundance for each gene is presented as relative fold expression. **b** qRT-PCR comparing the expression of saposin type B and fasciclin-like protein FLA15 in BTx623 and two sweet sorghum lines Della and Dale.

50 mya [23]. Because sorghum and sugarcane belong to the same clade, we reasoned that, by hybridizing RNA from grain and sweet sorghum onto the sugarcane GeneChip, we could correlate changes in transcript levels with traits from sweet sorghum such as sugar content and reduced ligno-cellulose. Given the tissue specificity and the rather small gene set of the sugarcane GeneChip, the positive hybridization of stem-derived RNAs from sorghum to 5,900 sugarcane probes of a GeneChip comprising 8,224 probe sets in total is a good indication of such cross-referencing. By applying a twofold cut-off value as a parameter to filter out differentially expressed transcripts, a total of 195 probe sets were identified, of which 63 corresponded to transcripts that were up-regulated (51 genes) and 132 (103 genes)
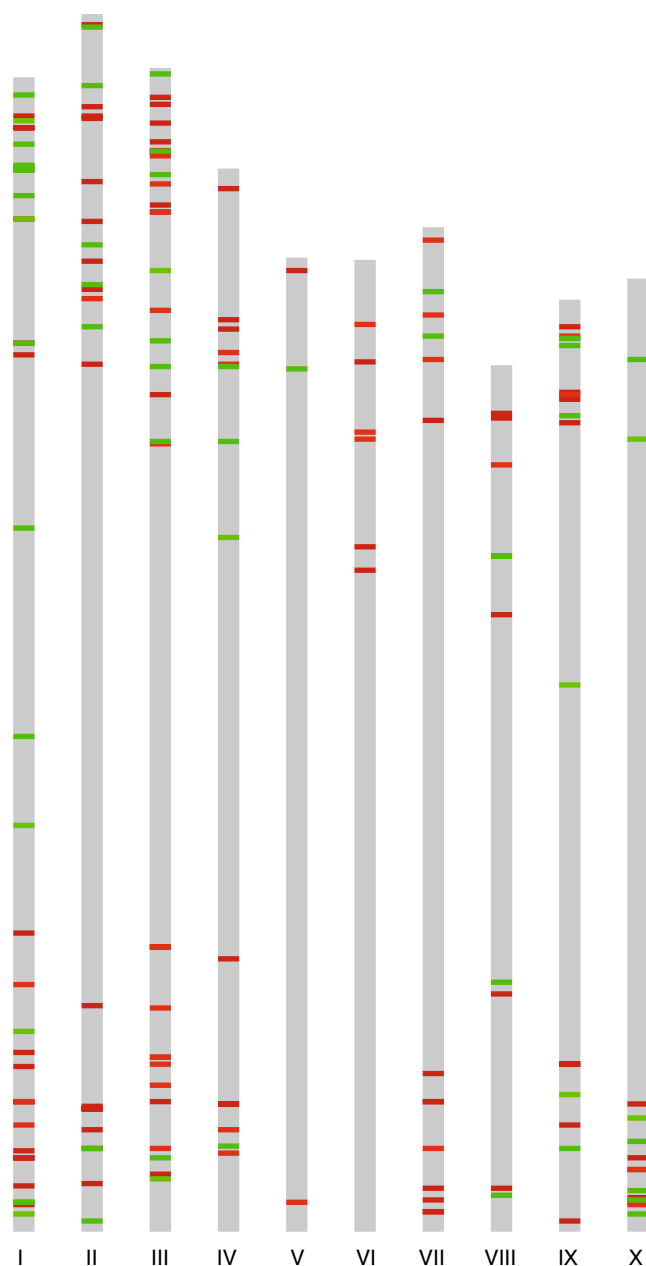
Fig. 3 Localization of differentially expressed genes on the physical map of sorghum. Each sugarcane probe set representing a differentially expressed gene between BTx623 and Rio with a fold change of two or higher was mapped to the sorghum genome and plotted on the physical map. Up-regulated genes are in *green* and down-regulated genes are in *red*.

corresponded to transcripts that were down-regulated in the sweet sorghum Rio line, respectively. Each differentially expressed sorghum transcript was classified based on the Pfam domains of their encoded proteins and their GO term ("Materials and methods").

Based on the sucrose and starch metabolic pathway from the KEGG (http://www.genome.jp/kegg/) and the CAZy database (http://www.cazy.org/), we found that almost 16%

of the transcripts involved in sucrose and starch metabolism and in cell-wall-related processes were differentially expressed between BTx623 and Rio. This is particularly interesting because a previous study with cDNAs from immature and maturing stem of sugarcane identified only 2.4% of the transcripts related to carbohydrate metabolism [7]. Furthermore, because sorghum stems are fully elongated at the anthesis stage, tissue samples from maturing internodes were also more suitable in profiling changes in gene expression associated with carbohydrate metabolism. The implication is that screening of differentially expressed genes can greatly be enhanced by genetic variability and selection of tissue.

Function of genes with elevated expression
in sweet sorghum

The highest elevated transcript identified in our study encodes a saposin-like type B domain. Increased expression has also been validated and tested in other sweet sorghum lines by qRT-PCR. We also found a higher expression in Dale and Della compared to that in BTx623 (Fig. 2b). Saposins are water soluble proteins that interact with the lysosomal membrane and are involved in the catabolism of glycosphingolipids in animals [31, 39]. Although it was unexpected that such a function could be related to a role in sugar accumulation, it underscores the value of a microarray-based screen to detect possible new network effects. For instance, we could hypothesize that the removal of sugars from glycosphingolipids in the membrane alters its structure in such a way that it constitutes an early step in carbohydrate partitioning. Additional transcripts that were increased in sweet sorghum included hexokinase 8, sorbitol dehydrogenase, and carbohydrate phosphorylase (starch phosphorylase). Hexokinase has a role not only in glycolysis but also as a glucose sensor that controls gene expression [19]. Sorbitol dehydrogenase is an enzyme involved in carbohydrate metabolism that converts the sugar alcohol form of glucose (sorbitol) into fructose [47]. Increased transcript levels of carbohydrate phosphorylase suggest that enhanced starch degradation in Rio may contribute to sugar accumulation. Another increased transcript encodes a NADP-malic enzyme suggesting that carbon fixation is enhanced in the stems of sweet versus grain sorghum. Indeed, the activity of enzymes involved in photosynthesis and the expression of their transcripts are modulated by sink strength. In sugarcane, the accumulation of sucrose in the maturing and mature internodes of the stem contribute greatly to sink strength [27]. Kinetic models have been proposed to explain sucrose accumulation in sugarcane [35, 41]. These models support the notion that sucrose accumulates in the vacuole against a concentration gradient. Indeed, we found that a transcript encoding a vacuolar adenosine triphosphate (ATP) synthase

catalytic subunit A had an increased expression in sweet sorghum, consistent with the role of this ATP synthase in the generation of an electrochemical gradient across the vacuolar membrane to propel the transport of sucrose.

The only cell-wall-related transcript that was up-regulated in sweet sorghum encodes a lysine motif containing protein. The LysM domain is widespread in bacterial proteins that degrade cell walls but is also present in eukaryotes. They are assumed to have a general role in peptidoglycan binding [1].

Mobilization of sugars in the stems of sweet sorghum

Interestingly, genes with reduced transcript levels outpaced those with increased levels by a 2:1 margin. Down-regulated transcripts involved in the starch and sucrose metabolic pathway found in our study included alpha-galactosidase, beta-galactosidase, sucrose synthase 2, and fructokinase 2. Alpha and beta-galactosidase enzymes are *O*-glycosyl hydrolases that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety [15]. Sucrose synthase is involved in the reversible conversion of sucrose to uridine diphosphate (UDP)-glucose and -fructose [24]. UDP-glucose can then be used as a substrate for starch and cell wall synthesis. Fructose instead is converted into fructose-6-phosphate by fructokinase and further metabolized through glycolysis [33]. Our findings are in agreement with previous reports showing that the onset of sucrose accumulation in Rio was accompanied by a decrease in sucrose synthase activity in stem tissue [25]. Similarly, Tarpley et al. [40] proposed that a decline in the levels of sucrose synthase may be necessary for sucrose accumulation at stem maturity in sorghum. Consistent with our findings, Xue et al. [44] have recently reported the down-regulation in the expression of both sucrose synthase and fructokinase genes in the stems of wheat genotypes with high water-soluble carbohydrates.

Reduced expression of cellulose and lignocellulose-related genes in sweet sorghum stems

Several transcripts involved in cell-wall-related processes were identified as down-regulated in sweet sorghum. These included cellulose synthase 1, 7, and 9 as well as cellulose synthase catalytic subunit 12 in cellulose synthesis. In the case of lignin biosynthesis, we found transcripts such as phenylalanine and histidine ammonia-lyase, cinnamoyl CoA reductase, 4-coumarate coenzyme A ligase, and caffeoyl-CoA *O*-methyltransferases. Interestingly, the expression of two transcripts encoding for xylanase inhibitors were also down-regulated in sweet sorghum. Xylanase inhibitor proteins belong to the group of protein inhibitors of cell wall degrading enzymes. Xylan is the major

hemicellulose polymer in cereals and is degraded by plant endoxylanases [21]. This suggests that, in sweet sorghum, the degradation of hemicellulose is promoted by suppressing the expression of xylanases inhibitors.

In other cases, a decrease in the expression of cellulose synthase genes in wheat genotypes with high water-soluble carbohydrate content has also been observed [44]. In addition, Casu et al. [6] have recently characterized the expression of several cellulose synthase and cellulose synthase-like genes in sugarcane stem and found that their expression is highly variable depending on internode maturity [6].

Reduced higher-order components in sweet sorghum stems

In addition to cellulose synthesis, the geometric deposition of cellulose fibrils generally perpendicular to the axis of cell elongation is a critical step in cell wall formation. There is evidence that the orientation of cellulose deposition is somehow assisted by microtubules [37]. An example of this is the *fiber fragile* mutant *fra1* encoding a kinesin-like protein. In this mutant, cellulose deposition displayed an abnormal orientation [5]. Consistent with these observations, the expression of two transcripts encoding tubulin alpha-2/alpha-4 chain and tubulin folding cofactor A, in conjunction with a transcript encoding a protein with kinesin motor domain, were all down-regulated in sweet sorghum.

Less clear, but also related to cell wall formation, is fasciclin. Interestingly, the most strongly down-regulated transcript in sweet sorghum encodes a protein with a fasciclin domain. Fasciclin domains are found in animal arabinogalactan proteins that have a role in cell adhesion and communication [22]. These proteins are structural components that mediate the interaction between the plasma membrane and the cell wall. However, their specific role in plants is still unknown [10]. A loss-of-function mutant in the *Arabidopsis* gene fasciclin-like arabinogalactan 4 (*AtFLA4*) displayed thinner cell walls and increased sensitivity to salinity [46].

Reduced cross-linking in sweet sorghum stems

Other transcripts that were also down-regulated encode a peroxidase and a laccase. It has been shown that peroxidases have an important role in cell wall modification [32]. By controlling the abundance of H_2O_2 in the cell wall, a necessary step for the cross linking of phenolic compounds, peroxidases act to inhibit cell elongation and, in conjunction with laccases, are assumed to be involved in monolingol unit oxidation, a reaction necessary for lignin assembly. Furthermore, it is known that peroxidase activity can be controlled by ascorbate. Indeed, the expression of a transcript encoding a protein similar to guanosine diphosphate (GDP)-mannose

3, 5-epimerase was increased in sweet sorghum. This protein catalyzes the reversible conversion of GDP-mannose either into GDP-L-galactose or a novel intermediate, GDP-gulose, a step necessary for the biosynthesis of vitamin C in plants [43]. In addition, GDP-mannose is used to incorporate mannose residues into cell wall polymers [26]. For these reasons, it is considered that GDP-mannose 3,5 epimerase could modulate the carbon flux into the vitamin C pathway as well as the demand for GDP-mannose into the cell wall biosynthesis [43]. Indeed, it is known that the stem of high-sucrose-accumulating genotypes of sugarcane are high in moisture content and low in fiber, whereas the stem of low-sucrose-accumulating genotypes are low in moisture content, thin, and fibrous [4].

Compensation of osmotic shock in sweet sorghum stems

Consistent with the idea that high concentration of sugars imposes osmotic stress to the cell, we found increased transcripts encoding heat shock proteins HSP70 and HSP90. Additionally, a transcript encoding a poly adenosine diphosphate (ADP)-ribose polymerase 2 (*PARP 2*) was significantly down-regulated in sweet sorghum. This is in agreement with a recent report in which *Arabidopsis* and *Brassica napus* transgenic plants with reduced levels of *PARP 2* displayed resistance to various abiotic stresses [42]. Poly ADP-ribosylation involves the tagging of proteins with long-branched poly ADP-ribose polymers and is mediated by PARP enzymes [36]. Poly ADP-ribosylation has important roles in the cellular response to genotoxic stress, influence DNA synthesis and repair, and is also involved in chromatin structure and transcription.

Mapping genes linked to sugar content and cell wall metabolism in sorghum and rice

Although sugarcane has not been sequenced yet, we can use the sequenced genome of sorghum to construct a "transcriptome map" with the genes found in our study. Assuming that gene order has been largely conserved between these closely related species, the "transcriptome map" of sorghum serves as a valuable reference for sugarcane. We could not find any particular clustering of these genes but did observe that most of the genes are located towards the telomeres and only a few of them near the centromeres. We also could not find any of these genes in the telomeric region on the long arm of chromosome six.

Comparing this map with the rice genome demonstrated that, out of 154 differentially expressed genes, 123 were in syntenic positions. With respect to the subset of genes involved in the accumulation of fermentable sugars and reduced lignocellulose, 21 genes were also found in syntenic regions, whereas nine genes appeared to be paralogous copies (Tables 1 and 2).

Outlook

Given the synteny of these genes between rice and sorghum, one can assume that they are allelic between different sorghum cultivars. Therefore, future genetic mapping experiments should provide a direct link of allelic variation and the sweet sorghum trait. Most likely, such allelic variations extend to the control regions of these genes because of their differential expression. Transgenic experiments can then be used to verify such functional aspects for biofuel properties. Moreover, gain of function experiments could be used to import desirable traits such as accumulation of fermentable sugars from sweet sorghum into maize. The generation of "sweet sorghum-like transgenic corn" will alleviate in part the increasing pressure of growing corn either for food or for biofuel since it would then be possible to use the grain for food and at the same time to extract fermentable sugars from the stem to use in ethanol production.

Materials and methods

Plant materials and growth conditions

Seeds from both grain and sweet sorghum (*Sorghum bicolor* (L.) Moench) were sown in pro-mix soil (Premiere Horticulture Inc., USA) and grown in our greenhouse with a day length of 15 h light: 9 h dark at constant temperature of 23°C. The genotype representing grain sorghum in our study was BTx623, whereas the genotypes representing sweet sorghum were Dale, Della, M81-E, Rio, Simon, and Top76-6. The seeds from sweet sorghum were kindly provided by Dr. William L. Rooney of Texas A&M, College Station, TX, USA.

Measurement of "Brix degree" from sorghum stem's juice

The juice from internodes of the main stem in both grain and sweet sorghum was harvested at the time of anthesis. A section of approximately 6 cm long was dissected from the middle of each internode, and 300 μl of juice was extracted by pressing each internode with a garlic squeezer. The concentration of total soluble sugars in the juice was measured with a pocket refractometer (Atago Inc., Japan).

Isolation of total RNA from stem tissue

Both grain sorghum BTx623 and sweet sorghum Rio were grown until anthesis and total RNA from internode 8 for each genotype (internodes were numbered from the base

towards the apex of the stem) was extracted using the RNeasy Plant Mini Kit (QIAGEN Inc., USA).

GeneChip sugarcane genome array hybridization

Sorghum RNA from internode 8 was hybridized to the Affymetrix GeneChip Sugarcane Genome Array (Affymetrix Inc., USA). Probe set information can be found at NetAffx Analysis Center's web page (http://www.affymetrix.com/analysis/index.affx). The One-Cycle Eukaryotic Target Labeling Assay protocol was used. The labeling, hybridization, and data collection were done at the Transcription Profiling Facility, Cancer Institute of New Jersey, Department of Pediatrics, Robert Wood Johnson Medical School.

Data analysis

Probe sets that were absent in all chips were eliminated. About 5,900 out of the original 8,300 probe sets passed this test. Next, a t test was applied to BTx623 and Rio groups (three replicates for each) with an alpha value of 0.001, and the Benjamini–Hochberg multiple-testing correction was applied. From the probe sets that passed the criteria, only those with a fold change of at least two were considered.

Validation of microarray data through quantitative RT-PCR

cDNA synthesis and PCR amplification was performed in the same tube from 50 ng of total RNA using the iScript One-Step RT-PCR Kit with SYBR Green (BIO-RAD Laboratories, Inc.). The reaction condition used was as specified in the kit, with an annealing temperature of 55°C and 45 cycles for the data collection step. The qRT-PCR reaction was done using the MyiQ Real-Time PCR Detection System (BIO-RAD Laboratories, Inc.). Total RNA was accurately measured for each sample with the extremely sensitive NanoDrop 1000 spectrophotometer (Thermo Scientific, Inc.). A relative quantification normalized against unit mass (50 ng of total RNA) was used to analyze the expression data with the equation: Ratio (test/calibrator) $= 2^{\Delta CT}$, as suggested in Real-Time Applications Guide from Bio-Rad. The primers for each gene were designed based on the region of homology (usually in the last exon or 3′untranslated region) between the sugarcane probe set sequence and the sorghum gene sequence and are listed in Electronic Supplementary Material Table S3. The sequence for each sugarcane probe set is freely available at the Affymetrix website: http://www.affymetrix.com/analysis/index.affx. In addition, the genomic location of each sugarcane probe set in sorghum (BTx623) identified in our work has been up-loaded to our Sorghum Genome Browser

and is available at http://genlisea-rs1.waksman.rutgers.edu/cgi-bin/gbrowse/sbic/.

Physical location of differentially expressed transcripts in the sorghum genome

The sugarcane probe sets that were up- and down-regulated in Sorghum, respectively, were mapped to the genome by using GenomeThreader [12]. Spliced alignments were only considered if 75% (score >0.75) or more bases could be aligned between the genomic sequence and a probe set. If a probe could be mapped to the genome and if it also overlapped with a sorghum gene, we assigned the annotation of the sorghum gene to the probe.

References

1. Bateman A, Bycroft M. The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD). J Mol Biol. 2000;299:1113–9.
2. Bennetzen JL, Freeling M. Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet. 1993;9:259–61.
3. Buchanan CD, Lim S, Salzman RA, Kagiampakis I, Morishige DT, Weers BD, et al. Sorghum bicolors transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol. 2005;58:699–720.
4. Bull T, Glasziou K. The evolutionary significance of sugar accumulation in Saccarhum. Aust J Biol Sci. 1963;16:737–42.
5. Burk DH, Ye ZH. Alteration of oriented deposition of cellulose microfibrils by mutation of a katanin-like microtubule-severing protein. Plant Cell. 2002;14:2145–60.
6. Casu RE, Jarmey JM, Bonnett GD, Manners JM. Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. Funct Integr Genomics. 2007;7:153–67.
7. Casu RE, Grof CP, Rae AL, McIntyre CL, Dimmock CM, Manners JM. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol. 2003;52:371–86.
8. Chapple C, Carpita N. Plant cell walls as targets for biotechnology. Curr Opin Plant Biol. 1998;1:179–85.
9. D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC. Characterization of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Mol Gen Genet. 1996;250:405–13.

10. Faik A, Abouzouhair J, Sarhan F. Putative fasciclin-like arabino-galactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): identification and bioinformatic analyses. Mol Genet Genomics. 2006;276:478–94.

11. Gale MD, Devos KM. Comparative genetics in the grasses. Proc Natl Acad Sci U S A. 1998;95:1971–4.

12. Gremme G, Brendel V, Sparks ME, Kurtz S. Engineering a software tool for gene structure prediction in higher organisms. Inf Softw Technol. 2005;47:965.

13. Grivet L, Arruda P. Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol. 2002;5:122–7.

14. Guimaraes CT, Sills GR, Sobral BW. Comparative mapping of andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci U S A. 1997;94:14261–6.

15. Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc Natl Acad Sci U S A. 1996;93:5674.

16. Hoffman-Thoma G, Hinkel K, Nicolay P, Willenbrink J. Sucrose accumulation in sweet sorghum stem internodes in relation to growth. Physiol Plant. 1996;97:277–84.

17. International Rice Genome Sequencing P. The map-based sequence of the rice genome. Nature 2005;436:793–800.

18. Ishimaru K, Hirotsu N, Madoka Y, Kashiwagi T. Quantitative trait loci for sucrose, starch, and hexose accumulation before heading in rice. Plant Physiol Biochem. 2007;45:799–804.

19. Jang JC, Leon P, Zhou L, Sheen J. Hexokinase as a sugar sensor in higher plants. Plant Cell. 1997;9:5–19.

20. Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, et al. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J. 2007;50:574–85.

21. Juge N, Nohr J, Le Gal-Coeffet MF, Kramhoft B, Furniss CS, Planchot V, et al. The activity of barley alpha-amylase on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus niger* glucoamylase. Biochim Biophys Acta. 2006;1764:275–84.

22. Kawamoto T, Noshiro M, Shen M, Nakamasu K, Hashimoto K, Kawashima-Ohya Y, et al. Structural and phylogenetic analyses of RGD-CAP/beta ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. Biochim Biophys Acta. 1998;1395:288–92.

23. Kellogg EA. Evolutionary history of the grasses. Plant Physiol. 2001;125:1198–205.

24. Koch K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. Curr Opin Plant Biol. 2004;7:235–46.

25. Lingle S. Sucrose metabolism in the primary culm of sweet sorghum during development. Crop Science. 1987;27:1214–9.

26. Lukowitz W, Nickle TC, Meinke DW, Last RL, Conklin PL, Somerville CR. Arabidopsis cyt1 mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. Proc Natl Acad Sci U S A. 2001;98:2262–7.

27. McCormick AJ, Cramer MD, Watt DA. Sink strength regulates photosynthesis in sugarcane. New Phytol. 2006;171:759–70.

28. Messing J, Llaca V. Importance of anchor genomes for any plant genome project. Proc Natl Acad Sci U S A. 1998;95:2017–20.

29. Messing J, Bennetzen J. Grass genome structure and evolution. Genome Dynamics. 2008;4:41–56.

30. Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 2001;11:2075–84.

31. Munford RS, Sheppard PO, O, Hara PJ. Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. J Lipid Res. 1995;36:1653–63.

32. Passardi F, Penel C, Dunand C. Performing the paradoxical: how plant peroxidases modify the cell wall. Trends Plant Sci. 2004;9:534–40.

33. Pego JV, Smeekens SC. Plant fructokinases: a sweet family get-together. Trends Plant Sci. 2000;5:531–6.

34. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The path forward for biofuels and biomaterials. Science 2006;311:484–9.

35. Rohwer JM, Botha FC. Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. Biochem J. 2001;358:437–45.

36. Schreiber V, Dantzer F, Ame JC, de Murcia G. Poly(ADP-ribose): novel functions for an old molecule. Nat Rev Mol Cell Biol. 2006;7:517–28.

37. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, et al. Toward a systems approach to understanding plant cell walls. Science 2004;306:2206–11.

38. Song R, Segal G, Messing J. Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic Acids Res. 2004;32:e189.

39. Stokeley D, Bemporad D, Gavaghan D, Sansom MS. Conformational dynamics of a lipid-interacting protein: MD simulations of saposin B. Biochemistry 2007;46:13573–80.

40. Tarpley L, Lingle S, Vietor DM, Andrews D, Miller F. Enzymatic control of nonstructural carbohydrate concentrations in stems and panicles of sorghum. Crop Science. 1994;34:446–52.

41. Uys L, Botha FC, Hofmeyr JH, Rohwer JM. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. Phytochemistry 2007;68:2375–92.

42. Vanderauwera S, De Block M, Van de Steene N, van de Cotte B, Metzlaff M, Van Breusegem F. Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. Proc Natl Acad Sci U S A. 2007;104:15150–5.

43. Wolucka BA, Van Montagu M. GDP-mannose 3′,5′-epimerase forms GDP-L-gulose, a putative intermediate for the de novo biosynthesis of vitamin C in plants. J Biol Chem. 2003;278:47483–90.

44. Xue GP, McIntyre CL, Jenkins CL, Glassop D, van Herwaarden AF, Shorter R. Molecular dissection of variation in carbohydrate metabolism related to water-soluble carbohydrate accumulation in stems of wheat. Plant Physiol. 2008;146:441–54.

45. Yang J, Zhang J. Grain filling of cereals under soil drying. New Phytol. 2006;169:223–36.

46. Yang J, Sardar HS, McGovern KR, Zhang Y, Showalter AM. A lysine-rich arabinogalactan protein in Arabidopsis is essential for plant growth and development, including cell division and expansion. Plant J. 2007;49:629–40.

47. Zhou R, Cheng L, Dandekar AM. Down-regulation of sorbitol dehydrogenase and up-regulation of sucrose synthase in shoot tips of the transgenic apple trees with decreased sorbitol synthesis. J Exp Bot. 2006;57:3647–57.