Genome-wide association study of multiple yield components in a diversity panel of polyploid sugarcane (*Saccharum* spp.)

Xiping Yang^{*}, Ziliang Luo^{*}, James Todd[†], Sushma Sood[‡], Jianping Wang^{*, §, 1}

^{*} Agronomy Department, University of Florida, Gainesville, FL 32610, USA

[†] Sugarcane Research Unit, USDA-ARS, Houma, LA 70360

[‡] Sugarcane Field Station, USDA, ARS, Canal Point, FL 33438

[§] Center for Genomics and Biotechnology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Corps, Ministry of Education; Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

¹Corresponding authors:

Jianping Wang, Ph. D, Associate Professor

Agronomy Department

Room 337 Cancer/Genetics Research Complex

2033 Mowry Road, University of Florida, Post Office Box 103610, Gainesville, FL 32610-3610

Phone: 352-273-8104 (O); 352-273-8105 (L)

e-mail: wangjp@ufl.edu

Running title: GWAS of yield components in sugarcane

Abstract

Sugarcane (*Saccharum* spp.) is an important economic crop, contributes up to 80% of sugar and approximately 60% bio-fuel globally. To meet the increased demand for sugar and bio-fuel supplies, it is critical to breed sugarcane cultivars with robust performance in yield components. Therefore, dissection of causal DNA sequence variants is of great importance by providing genetic resources and fundamental information for crop improvement. In this study, we evaluated and analyzed nine yield components in a sugarcane diversity panel consisting of 308 accessions primarily selected from the “world collection of sugarcane and related grasses”. By genotyping the diversity panel using target enrichment sequencing, we identified a large number of sequence variants. Genome-wide association study between the markers and traits were conducted with dosages and gene actions taken into consideration. In total, 217 non-redundant markers and 225 candidate genes were identified to be significantly associated with the yield components, which can serve as a comprehensive genetic resource database for future gene identification, characterization, and selection for sugarcane improvement. We further investigated runs of homozygosity (ROH) in the sugarcane diversity panel. We characterized 282 ROHs, and found that the occurrence of ROH in the genome were non-random and probably under selection. ROHs were associated with total weight and dry weight, and high ROHs resulted in decrease of the two traits. This study approved that genomic inbreeding has led to negative impacts on sugarcane yield.

Keywords: Sugarcane germplasm; polyploidy; yield components; target enrichment sequencing; GWAS

Introduction

As an important cash crop, sugarcane (*Saccharum* spp.) contributes up to 80% of sugar and approximately 60% bio-fuel globally (Dahlquist 2013). To meet the increasing global demand for sugar and bio-fuel, sugarcane harvest area in the world has increased from 22.7 to 26.8 million hectares in the last ten years (FAO 2016). However, cropping areas for sugarcane could not be increased indefinitely due to the limited farming land areas and competition with food crops. In order to sustainably improve sugarcane production, breeding sugarcane cultivars with high performance for yield components is critical. Thus, evaluating yield components in sugarcane germplasm and dissecting the genetic basis of causal sequence variants are of great importance to fulfill this strategy by providing tools and genetic resources.

The *Saccharum* genus consists of two wild species (*S. spontaneum* and *S. robustum*) and four cultivated species (*S. officinarum*, *S. barberi*, *S. sinence*, and *S. edule*) (Daniels and Roach 1987). All the *Saccharum* species are readily intercrossed except sterile *S. edule*. Modern sugarcane hybrids were primarily derived from inter-species hybridization between *S. spontaneum* ($2n = 40-128$, $x = 8$) and *S. officinarum* ($2n = 80$, $x = 10$) followed by several backcrosses with *S. officinarum*. Therefore, modern sugarcane hybrids have a chromosome number of 100-130 with approximately 80% chromosomes from *S. officinarum*, 10% chromosomes from *S. spontaneum*, and 10% recombinant chromosomes (D'Hont et al. 1996), and an estimated genome size of approximately 10 Gbp (D'Hont 2005). Due to the nature of huge genome size and high polyploidy level (Primarily aneuploidy), genetic studies of sugarcane have been extremely challenging with relatively slow progress compared with many other crops.

Modern sugarcane cultivars could be traced back to only a handful of *S. spontaneum* and *S. officinarum* ancestor clones (Acevedo et al. 2017; Deren 1995; Lima et al. 2002). This

narrow genetic base hampered the sugarcane cultivar improvement on yield and tolerance. Therefore, maintenance, characterization, and utilization of genetic diversity in sugarcane germplasm collection are critical to broaden the genetic background of sugarcane for crop improvement. The “world collection of sugarcane and related grasses” (WCSRG), which assembled a significant amount of genetic diversity, is an essential resource for sustained sugarcane cultivar improvement (Nayak et al. 2014; Todd et al. 2014). The collection has ~ 1,200 accessions from 45 countries, including *Saccharum* germplasm and closely related grass species with the most abundant being *S. spontaneum*, *S. officinarum*, and interspecific hybrids. This collection preserves gene resources that could be used to presumably improve yield, fiber and abiotic and biotic stresses in sugarcane breeding programs. A core germplasm collection was created by selecting representative accessions from the WCSRG (Nayak et al. 2014; Todd et al. 2014), which showed significant amount of natural phenotypic variations, and can be valuable resources for identifying desirable alleles contributing to yield and tolerance for sugarcane improvement.

Genome-wide association study (GWAS) has been one of the most powerful tools to discover causal variants for complex traits in human, animal and plant systems. Compared with traditional linkage mapping strategy performed in bi-parental populations, GWAS explores historical and evolutionary recombination events at the population level, and thus achieves a high resolution to the gene level (Zhu et al. 2008). In addition, multiple genes and alleles involved in the same traits could be investigated and identified in a natural diversity panel. With continually decreasing cost for the next generation sequencing (NGS) technology, GWAS has been widely applied in complex trait dissection in the genomic era. However, GWAS might not be as promising in polyploid sugarcane (With up to 12 sets of chromosomes) as in diploids. There are

several primary challenges that could hinder the use of GWAS in sugarcane. For example, 1) relatively high sequencing depth needed to call sequence variants due to their large genome sizes (~ 10 Gbp); 2) dosages of markers were difficult to be determined especially for single dose markers; and 3) most GWAS models and software are designed for diploids, in which dosages and gene actions could not be accounted for analyses. Hence in most cases, GWAS in polyploids treated molecular markers as diploids, which might not reflect the actual genetic effects. Since the first GWAS was conducted (Haines et al. 2005), a few GWAS have been performed in sugarcane (Débibakas et al. 2014; Gouy et al. 2015; Racedo et al. 2016; Wei et al. 2006, 2010). However, all of them were challenged by the difficulties mentioned above. GWASpoly in R is a software tailored for polyploids, which could model different types of polyploid gene actions in GWAS, including additive, simplex dominant, and duplex dominant (Rosyara et al. 2016). This software has been tested in a simulated tetraploid population, and successfully applied in potato, sunflower, orchardgrass and wheat for GWAS (Berdugo-Cely et al. 2017; Bock et al. 2018; Phan et al. 2018; Rosyara et al. 2016; Zhao et al. 2017). However, this software has not been utilized for GWAS in sugarcane.

A run of homozygosity (ROH) is a continuous homozygous segment of the genome in an individual due to haplotypes inherited from common ancestors in the past (Ceballos et al. 2018). ROHs capture population history of a species, such as inbreeding, change of population size, admixture and so on. ROH calling relies on high-density genome-wide markers (Ceballos et al. 2018). With the development of NGS technologies, large numbers of single polymorphisms (SNPs) can be generated at a relatively low price, making it achievable to perform ROH analysis to capture the genomic regions contributing to inbreeding and thus to assess the breeding history and to identify the genetic components for trait selection. ROHs are widely applied in human and

animal genetics (Ceballos et al. 2018; Christofidou et al. 2015; Peripolli et al. 2018; Sud et al. 2015; Yang et al. 2015) for detecting inbreeding and loci associated with quantitative trait or disease. However, there is no report of ROHs in sugarcane genetic research.

In this study, we assembled a diversity panel consisted of 299 accessions based on phenotypic and genotypic data collected from the WCSRG, and nine sugarcane cultivars and breeding materials (Nayak et al. 2014; Todd et al. 2014). Nine yield components were evaluated in the diversity panel in three replicates. We deeply sequenced the coding regions of the 308 accessions in the diversity panel, and identified large amounts of sequence variants with a confident dosage resolution. We performed GWAS using GWASpoly in sugarcane with marker dosages and gene actions taken into consideration. A total of 217 non-redundant sequence variants and 225 genes were significantly associated with the traits evaluated. In addition, for the first time, we identified 280 ROHs, indicators of inbreeding in the diversity panel, and found that ROHs were associated with two yield components. The results of this study not only provided the genetic bases of multiple yield components in sugarcane but also suggested important tools and deposited valuable genomics resources for the sugarcane community. The new methods and concepts in this study shed light on research in sugarcane and polyploid species.

Materials and methods

Plants in the association panel

A core sugarcane diversity panel consisting of 299 accessions collected were selected from the WCSRG, maintained at the USDA-ARS Subtropical Horticulture Research Station, Miami, Florida, to represent the wide genetic diversity of the WCSRG (Nayak et al. 2014; Todd et al. 2014). Nine modern sugarcane hybrids frequently used as parental materials in the USA

Florida sugarcane breeding program (Table S1) were also included in this study. In total, 308 accessions were established at USDA-ARS Sugarcane Field Station at Canal Point, FL in 2013 in a field nursery in a randomized complete block design with three replicates. Regular fertilizer and daily water were applied to maintain healthy plants in the nursery.

Phenotyping and data treatment

During August 2014, nine yield components were measured for the sugarcane diversity with three replicates on ratoon plants including stalk diameter, leaf width, leaf length, stalk number, internode length, brix, total weight, dry weight and water content (Todd et al. 2017). In brief, brix, stalk number, stalk diameter, internode length, leaf width, and leaf length of the leaf at the top visible dewlap were recorded for each plant before harvesting. Immediately prior to harvest, three or more stalks (To roughly equal 1 kg dry mass) were shredded, and the total weight was recorded per pot. Shredded cane was subsequently dried at 60 °C, and re-weighted to determine dry weight. Water content was calculated by subtracting dry weight from total weight.

The normality of the trait distribution was checked by their phenotype distribution. For the traits not following normal distribution, a Box Cox transformation was performed using MASS package in R (Ripley et al. 2013) to satisfy model assumptions. A pairwise Pearson correlation among yield components was calculated using the ‘Hmisc’ package in R3.0.2 (Harrell 2018). The total correlation coefficients of yield components to sugarcane dry weight and sugarcane content (Brix) were participated into direct and indirect effects following the method (Dewey and Lu, 1959), and the computation was conducted using ‘lavaan’ and ‘semPlot’ package in R (Epskamp 2015; Rosseel 2012). Analysis of variance was performed using the ‘nlme’ package in R (Pinheiro et al. 2009). Broad sense heritability (H^2) was estimated using the formula:

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \frac{\sigma_e^2}{r})$$

Where σ_g^2 is the genetic variance, σ_e^2 is the residual variance and r is the number of replicates. For each trait, the best linear unbiased predictor (BLUP) was estimated by modeling genotypes as random effects and replicates as fixed effects using *lme4* package in R (Bates et al. 2014), and the computed BLUPs were used for GWAS.

Genotyping

Young leaves were collected from each accession, and the DNA was extracted by using the cetyltrimethyl ammonium bromide (CTAB) method (Wang et al. 2010). The target enrichment sequencing (TES) of each accession was conducted (Song et al. 2016). In brief, a set of 60 thousand 120-bp oligonucleotides probes were designed, and used for TES. The captured DNA libraries were prepared following the product manual of Agilent SureSelectXT reagents (Agilent Technologies, Santa Clara, CA), and then were indexed and cleaned using Ampure beads, quantified by qPCR using KAPA Library Quantification Kits (KAPA Biosystems.com) and sequenced in paired-end mode (2 x 100 bp) by Illumina HiSeq 2000.

The raw reads were sorted into reads per library according to their barcode sequences and then the barcodes were removed from raw data. The paired-end reads were trimmed from ends in pairs based on PHRED score of 20 using Trimomatic v0.36 (Bolger et al. 2014). Reads longer than 50 bp were aligned to sorghum genome v3.0 (Paterson et al. 2009). BWA-mem (Li 2013) with default settings was used for alignment. Uniquely mapped reads were extracted from sequence alignment BAM files, and used for sequence variant calling. SNPs were called by using Unified Genotyper implemented in Genome Analysis Tool Kit v3.30 (McKenna et al. 2010). The raw SNPs were subjected to a series of filtering for each individual accession: 1) mapping

quality for SNPs > 30; 2) base quality for SNP sites > 20; 3) for a heterozygous genotype, at least two reads for minor allele were required; otherwise, missing genotypes were assigned; 4) for a given SNP locus with homozygous genotype, minimum reads were calculated according to ploidy of accessions to ensure single dose SNPs were not called as homozygotes (Yang et al. 2017, 2018). Based on the calculation, minimum reads of 5, 11, 17, 23, 29 and 35 were required for ploidy at 2, 4, 6, 8, 10 and 12 for homozygous genotype calling, respectively; otherwise, missing genotypes were assigned; 5) for a given SNP locus with heterozygous genotype, sequencing depth required to separate all SNP genotypes with a probability higher than 95% at determined ploidy, which were 31, 47, 70, 90 and 108 for ploidy at 4, 6, 8, 10 and 12, respectively. After the stringent filtering, genotypes of SNPs were scaled to the highest ploidy (12) according to ratio of reads for respective alleles. InDels were called using the same procedures as SNPs described above.

Genome wide association study

We used discriminant analysis of principal components (DAPC) implemented in the adegent package for R to infer the population structure of the accessions in this study (Jombart, 2008, 2010). *k*-means clustering was used to identify groups, where the best *k* minimized the Bayesian Information Criterion (BIC) and was thus used for GWAS. GWASpoly is tailored for autopolyploids based on the Q + K mixed model, and could model gene actions for polyploids (Rosyara et al. 2016). Six models were used for GWAS by using GWASpoly, including general model, additive model, two simplex dominant models (1-dom-ref and 1-dom-alt), diplo-general model, and diplo-additive model. In additive model, effect of markers was proportional to the dosage of minor allele, while heterozygotes were equivalent to one of the homozygotes for simplex model (See (Rosyara et al. 2016) for detail). Diplo- means that genotypes of markers

were diploidized for the model. Sequence variants for GWAS were further filtered with a call rate $\geq 95\%$ and minor allele frequency (MAF) ≥ 0.05 . Only samples with less than 20% missing data were included for the GWAS. We performed GWAS runs for nine yield components in the full set of the sugarcane diversity panel (with 300 accessions) in addition to several subsets separately, including a subpopulation 1 (Sacc) with 273 accessions from *S. spontaneum*, *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*, a subpopulation 2 (Spon) with 100 accessions from *S. spontaneum*, a subpopulation 3 (Hybrid) with 173 accessions from *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*, and a subpopulation 4 (P8) with 162 accessions at the same ploidy level of 8. The annotation of candidate genes with GWAS hits were obtained from the sorghum genome annotation (Paterson et al. 2009), and their domains were annotated using InterProScan 5.0 (Jones et al. 2014).

Identification of runs of homozygosity

SNPs with confidence separating homozygote and heterozygote (After step 1, 2 and 3 filtering in ‘genotyping’ part) were used for identification of ROH. Additional quality control for SNPs included: individual missing rate $< 10\%$, SNP call rate $\geq 90\%$, and minor allele frequency $< 1\%$. ROHs were identified in each individual using Plink v1.90b3 using default settings by fixed scanning windows (Purcell et al. 2007). Specifically, the parameters used to define and identify a ROH were: 1) a sliding window contained 50 sequence variants with at most 1 heterozygous call and 5 missing calls; 2) for a variant to be eligible in a ROH, the hit rate of all scanning windows containing the variant must be at least 0.05; 3) maximum gap between two consecutive homozygous variants was 1,000 Kbp; 4) a ROH had a minimum length of 1,000 Kbp, and contained at least 100 sequence variants, and at least one variant per 50 Kbp on average. Linear regression was used to study association between ROHs and phenotypes with

control for population structures. Associations were performed for total number of ROHs (NROHs), sum total length of ROHs (SROHs), and average length of ROHs (AROHs), separately.

Data availability

The cleaned target enrichment sequencing reads generated in this study were deposited into NCBI Short Reads Archive with an accession number of SRP132365. The phenotype, genotype and population structure used in GWAS were deposited into Gatorcloud (<https://bit.ly/2OcwsN4>). The rest intermediate analysis data and the plant materials will be available upon request.

Results

Correlation and path coefficient analysis of yield components

All the traits were normally distributed except stalk number (Figure S1), which was further transformed. According to pairwise Pearson's correlation coefficients for the nine traits, five pairs of traits were found to be highly correlated (Correlation coefficients $r \geq 0.6$ or $r \leq -0.6$) ($P < 0.0001$) (Table 1). Specifically, leaf width was positively correlated with stalk diameter ($r = 0.82$) but negatively correlated with stalk number ($R = -0.62$), and stalk diameter had negative correlation with stalk number ($R = -0.69$), reflecting there was a competition for carbon resources between stalk diameter and stalk number. Based on the correlations, leaf width seems to be more important to determine sugarcane stalk diameter and stalk number compared with leaf length. As expected, sugarcane total weight, dry weight and water content were highly correlated (With a correlation coefficient 0.78 between dry weight and total weight, and 0.92 between total weight and water content) (Table 1). We further partitioned the correlation into direct effects and indirect effects to understand the inter-relationships among these yield components. Stalk number and internode length had high direct effects on sugarcane yield (Dry weight) (Table 2),

which were consistent with the results of overall correlation coefficients (Table 1). Moreover, indirect effects were relatively low for both stalk number and internode length. The results implied that selection on high stalk number and internode length would be beneficial for increasing sugarcane yield. On contrary, leaf width had negative direct and indirect effects on sugarcane yield, while stalk diameter had negative indirect effects (Table 2), which should be selected against for yield improvement. The situation is much complex for sugar content (Brix) because all the yield related traits, such as stalk diameter, stalk number and leaf width, had both relatively high negative and positive effects (Direct or indirect) on sugar content (Table 3). The overall correlation should be considered in breeding material selection. To breed sugarcane cultivars with high sugar content, the results suggested that genotypes with high stalk diameter, leaf width but low stalk number should be selected.

The broad sense heritability (H^2) ranged from 0.42 (Dry weight) to 0.91 (Stalk diameter) for the nine yield components (Table 1). The total weight, dry weight and water content had relative low broad sense heritability, probably because they were complex traits combining impacts from multiple yield components with significant environmental effects. Overall, the estimated broad sense heritability indicated that these yield components were largely controlled by genetic factors. Therefore, the nine yield components were used for GWAS analysis. Though some traits were highly correlated, GWAS may identify the same markers to be associated with the correlated traits. To maximum discovery of associations, we conducted GWAS for each of the nine traits separately without considering the correlation first. However, after the GWAS, we summarized and presented a unique set of associated markers subsequently.

Sequence variants captured by target enrichment sequencing

Overall, 8.0 billion clean reads were generated after sequencing the target regions of the

308 accessions with 7.4 billion (92.1%) of the clean reads mapped to the sorghum genome, and 7.3 billion (90.7%) uniquely mapped. The average sequence depth of the whole diversity panel was 99x in a range of 9x to 375x. Using the uniquely mapped reads, 25.3 million raw SNPs were identified, of which 9.9 million (39.2%) were SNPs with high quality after the stringent sample-by-sample filtering (Table 4). The 9.9 million SNPs included 8.6 million bi-allelic SNPs and 1.3 million multiple-allelic SNPs. Majority of the bi-allelic SNPs (90.2%) were rare variants (MAF < 0.05), and thus were removed from further analysis. With a SNP call rate $\geq 95\%$, 65 thousand SNPs (0.7% of the 9.9 million high quality SNPs) were included as markers for GWAS. Similar patterns were observed for InDels with 880.0 K high quality InDels after sample-by-sample filtering, in which 130.2 K (14.8%) were multiple allelic, and majority of the bi-allelic InDels (562.9 K, 75.1% of the 880.0 K) were rare variants (Table 4). With the same call rates as applied to SNPs, 9.9 K InDels were included as markers for GWAS. Therefore, a total of 74.9 K DNA markers, including 65.0 K SNPs and 9.9 K InDels, were used for subsequent GWAS.

According to the sorghum genome, the majority of the DNA markers used for GWAS were in genic regions (67.5 K, 90.1%) targeting 10,458 gene models. The number of markers per chromosome ranged from 2,758 (Chr05) to 12,836 (Chr01), with an average marker distance of 9,116 bp. Given the large block of linkage disequilibrium in sugarcane (~ 5 cM, (Raboin et al. 2008)), the marker density in this study should be high enough for an effective GWAS.

Inferring population structure

We carried out population structure analysis with DAPC using 13.8 K SNPs, which were evenly distributed according to the sorghum genome (One SNP per 10 Kbp) and had call rates $\geq 95\%$ in the sugarcane diversity panel. We selected $k = 6$ representing number of population groups in the diversity panel, which was within the curve minimizing the BIC value (Figure 1A).

Among the 308 accessions, four groups belonging to *Saccharum* were identified with *S. spontaneum* group including 101 accessions, *S. officinarum* group including 47 accessions, modern *S. hybrids* group including 114 accessions, and ancient *S. hybrids* group (*S. barberi* and *S. sinence*) including 17 accessions (Figure 1B; Table S1). In addition to *Saccharum* spp. groups, two non-*Saccharum* groups were identified in the diversity panel as non-*Saccharum* 1 and non-*Saccharum* 2 with 10 and 19 accessions, respectively (Figure 1B, Table S1). Based on WCSR records, non-*Saccharum* 1 most likely belonged to *Miscanthus*, while non-*Saccharum* 2 belonged to *Erianthus* (Table S1). Due to low sequencing depth (On average 24.3), eight accessions, including five accessions from modern *S. hybrids*, two from non-*Saccharum* 2, and one from *S. spontaneum*, with missing sequence variants > 20%, were removed from further analysis. As a result, 300 accessions from the sugarcane diversity panel were included for subsequent GWAS.

Genome-wide association study

Using a stringent threshold ($P < 0.05$ after Bonferroni correction), 217 non-redundant markers (191 SNPs and 26 InDels) were identified to be significantly associated (Figure 2, Figure S2, S3, S4, S5, S6, and Table S3). The 217 associated markers included 128, 77, 30, 13 and 58 from the full population, subpopulation 1, 2, 3, and 4, respectively (See method part), with some associations overlapped (Table 5). The genome-wide association rate was 0.55% ($217/74,935 = 0.29\%$). The number of associated markers according to the sorghum gene models was 33 in intergenic regions, 60 in introns, 48 in untranslated regions, and 66 in coding regions. For the markers in coding regions, three had high effects (Stop gain or frameshift variant), 25 had moderate effects (Non-synonymous mutations) and 38 had low effects (Synonymous mutations). Only 28 markers (12.9%) led to new alleles with likely different functions, while the

rest did not change the original gene function, which might perform the roles as regulatory elements or associate with other causal genes. Among the 217 associated sequence variants, 109 were detected in at least two GWAS runs or two models reflecting certain level of consistence in the trait-marker association. There were nine markers associated with more than two different traits. Specifically, six markers (2P8663239, 3P66737730, 3P66744355, 4P2535140, 5ind60876916, 10P55668582) were associated with both stalk diameter and leaf width, two markers (3P61470337 and 7P2769299) with water content and total weight, and one marker (3P63471814) with dry weight, water content, and total weight (Table S3). These single markers associated with multiple traits indicated that either these loci located in or be associated with pleiotropic genes, or these traits were highly correlated biochemically.

Among the 217 non-redundant associated sequence variants, 184 were in genic regions targeting 169 gene models, and 33 in intergenic regions according to the sorghum genome (Table S3). To further investigate the genes involved in the associations, a total 225 non-redundant genes were obtained by retrieving the gene models hit by the associated markers and flanked by associated markers, including 167 genes with GWAS direct hits in gene models, 56 were extracted from genomic regions flanked by GWAS hit markers, and two genes were shared between the two categories (Table S3). The average number of candidate genes was 25 for the evaluated yield components with a range of two (Leaf length) to 89 (Leaf width). Details of associated markers and genes for the nine evaluated yield components were deposited in Tables S2 and S3. Based on annotations and domain analysis, among the associated genes, 10 were involved in sugar metabolism, including one (*Sobic.009G245000*) significantly associated with stalk diameter, one (*Sobic.006G278500*) with dry weight, two (*Sobic.002G279900* and *Sobic.009G063400*) with leaf width, and six (*Sobic.001G466900*, *Sobic.001G529600*,

Sobic.002G241100, *Sobic.003G180100*, *Sobic.009G233200* and *Sobic.009G073400*) with internode length (Table S3). These genes encoded enzymes such as sugar-related synthase, hydrolases and transferase, important for carbon participation regulation, thus contribute to the biomass growth. We further searched enriched Gene Ontology (GO) categories for the 225 non-redundant genes, and genes associated for each component separately. The biological process for signal transduction by phosphorylation was significantly enriched in the genes associated with stalk diameter ($P < 0.05$ adjusted with false discovery rate (FDR)) (Yi et al. 2013), indicating signal transduction might be critical for determining stalk diameter sizes. We specifically focused on 27 associated genes with markers in coding regions that caused large effects (Frameshift or stop gained) (Three genes) or moderate effects (missense or disruptive in-frame deletion) (24 genes). The mRNA surveillance pathway was significantly enriched in the 27 associated genes ($P < 0.05$ adjusted with FDR) (Yi et al. 2013), indicating that ensuring fidelity and quality of messenger RNA has significant impacts on sugarcane yield components.

Investigation of ROHs in the sugarcane diversity panel

With stringent filtering, 1.07 million SNPs and 211 accessions were used for ROH identification. The total genotyping rate of the SNPs was 97.2% in the 211 accessions. In total, 282 ROHs were characterized in 46 of the 211 accessions (20.4%). The distribution of ROHs was uneven and not correlated with SNP density, with a minimum of 12 ROHs on chromosome 1 and 5, and 57 ROHs on chromosome 7. There were 270 ROHs forming 50 pools of overlapping ROHs (The pool size ranging from 2 to 20 ROHs with 4 ROHs as the most popular pool size) (Figure 4A). The non-random distribution of ROHs indicated the occurrence of ROHs was most likely under selection during domestication and breeding processes. The sizes of ROHs ranged from 1.0 Mbp to 4.5 Mbp with an average of 1.5 Mbp. The most common ROH

(Detected in 20 accessions) was located at chr3: 37153462-39398419 bp, harboring 25 genes. By searching gene ontology and pathway analysis (Lyne et al. 2015), genes were enriched in five biological processes and three cellular components in photosynthesis ($P < 0.05$ adjusted with FDR).

The ROH detection rate was 29.8% for *S. spontaneum* group (Out of 74 accessions), 27.8% for *S. officinarum* group (Out of 36 accessions), 0 for ancient *S. hybrids* group (out of 14 accessions), 2.7% for modern *S. hybrids* group (Out of 75 accessions), and 100% for both non-*Saccharum* groups (Table 6). The average NROH and SROH was 26.9 and 45.1 Mbp for non-*Saccharum* 2, which were significantly higher than that of any other sub-populations in this study ($P < 0.001$) (Figure 4B, 4C and Table 6). No difference was observed for the AROHs (Figure 4D, Table 6). The association analysis between the ROHs and the nine yield related traits showed that two yield components, total weight and dry weight, were associated with NROHs and SROHs significantly ($P < 0.05$). The negative coefficients of ROHs for both traits indicated that increase of genomic inbreeding in sugarcane had negative influence on these two yield components.

Discussion

Dissection of genetic loci controlling yield components is of great importance for sugarcane improvement. In this study, we performed GWAS analyses in a large representative diversity panel using high-density NGS-based DNA markers with dosages and allele interactions taken into consideration. We identified 217 non-redundant markers and 225 genes significantly associated with the nine yield components evaluated. In addition, for the first time, we investigated ROHs in sugarcane, and found that ROHs were negatively associated with total weight and dry weight. This research provided valuable and novel genomic resources and tools

for sugarcane breeding programs, and empowered GWAS analyses in this polyploid species.

High throughput genotyping sugarcane clones empowered by TES

Sugarcane is a highly polyploidy species with up to 12 sets of chromosomes, and has an estimated genome size of approximately 10 Gbp (D'Hont 2005). Due to the nature of huge and highly polyploidy genomes of sugarcane, intensive sequencing depth is required to accurately call genotypes with allele dosage resolution, which is the prerequisite to perform GWAS with allele dosages under consideration. In this study, we applied TES in a representative diversity panel for a high throughput genotyping. TES could capture a subset of sugarcane genome for sequencing, and therefore effectively investigate sequence variants at limited sequencing resources. This strategy to reduce genome complexity is very helpful for genomic studies in species like sugarcane with complex genomes. By selectively sequencing mostly the coding regions, we achieved an average sequencing depth of 99x, which is difficult to obtain using other type of NGS-enabled methods. At this sequencing depth, we could differentiate all genotypes of sequence variants in sugarcane, even for dodeca-ploid species with up to 13 possible genotypes (Yang et al., 2017). After applying a stringent sample-by-sample filtering, we obtained 74.9 K markers with an average marker distance of 9 Kbp, which is much higher than any of other published GWAS study in sugarcane. This powerful NGS-based technology has maximized the identification of markers associated with the traits evaluated. However, we still had only 39.2% of the raw SNPs pass our stringent quality filtering (Mainly due to sequence depth), and only 0.7% of the SNPs had a decent call rate across the diversity panel. This significant reduction of SNP numbers during filtering informed us that the sequence depth is still the limiting factor for NGS application in this polyploid species.

Population structure in sugarcane

Population structure is an important factor needed to be controlled for GWAS. Several software are available for population structure analysis specifically for diploid species, such as Structure (Pritchard et al. 2000) and Admixture (Alexander et al. 2009), which were not suitable for sugarcane due to its nature of polyploids, outcrossing and clonal propagation, thus not fitting the assumptions in these software. We applied DAPC to capture population structures in this sugarcane diversity panel, which is a multivariate method for the analysis of the population genetic structure not relying on assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium (Jombart et al. 2010). Compared to principal component analysis, DAPC can partition genetic variation into a between-group and a within-group components, and thus achieves clear discrimination of individuals into pre-defined groups. Using 13.8 K high quality and evenly distributed SNPs, we have inferred six groups existed in this diversity panel, including two non-*Saccharum* (*Miscanthus* and *Erianthus*) and four *Saccharum* groups (*S. spontaneum*, *S. officinarum*, modern *S. hybrids*, and ancient *S. hybrids*), which aligned perfectly well with biological classification of this diversity panel according to the records of WCSRG. Moreover, DAPC could clearly separate *S. officinarum* from *S. hybrids*, and modern *S. hybrids* from ancient *S. hybrids*, which are highly genetically similar, supporting the sensitivity of this method to capture miniature population structures in sugarcane.

GWAS on yield components in a sugarcane diversity panel

We evaluated nine yield components, and performed correlation and path coefficients analyses in a representative sugarcane diversity panel. The path coefficients analyses could suggest appropriate selection strategies to breeding programs by revealing inter-relationship among yield components (Dewey and Lu 1959). For example, we recommended selecting large stalk diameter and leaf width for improving sugar content but for different reasons: a positive

direct effect has been estimated for stalk diameter but an indirect effect for leaf width. The net direct effects come from the balance of contributions and competitions among these components. The results of path coefficients analyses informed us that the sizes of stalk diameter might be critical for accumulating high sugar content through improving sugar storage and supporting functions, while only competition for carbon resources existed for leaf width and sugar content (Reflected by negative direct effect). The indirect effect of leaf width on sugar content derived from its influence on stalk diameter. Our results also demonstrated that the same yield component can be favored or against depends on the purpose of breeding programs. For example, genotypes with high stalk number had a high biomass yield but relatively low sugar content. Therefore, comprehensively evaluating and analyzing inter-relationships among yield components in a representative sugarcane diversity panel provided key information for sugarcane breeding programs.

Current studies had three major improvements compared with previous GWAS performed in sugarcane (Débibakas et al. 2014; Gouy et al. 2015; Racedo et al. 2016; Wei et al., 2006, 2010). First, a representative diversity panel with a large number of accessions derived from WCSRG was assembled and used in sugarcane GWAS. This diversity panel consisted of 308 accessions including 279 *Saccharum* accessions and 29 non-*Saccharum* accessions (*Miscanthus* and *Erianthus*). The panel was not limited in sugarcane elite germplasm as previous studies, primarily parental clones and sugarcane hybrids in breeding programs, and had a decent number of accessions from *Saccharum* and non-*Saccharum* species. Therefore, the statistical power of GWAS has been improved in this diversity panel. Moreover, our study has the capability to identify novel genetic resources, which might be not existed in modern sugarcane cultivars due to their narrow genetic basis. More importantly, these novel robust traits could be

quickly introgressed into sugarcane breeding programs through controlled crosses, especially with help of associated DNA markers. Second, we applied TES, an NGS-enabled genome complexity reduction sequencing technology, to genotype this diversity panel, which was proven to be very efficient for investigation of sequence variants in sugarcane (Song et al. 2016). With a stringent sample-by-sample and population-level filtering, 74.9 K markers were retained for GWAS, which was a much larger number of markers than any of other GWAS in sugarcane. Also, TES achieved an average sequencing depth of 99x in this study, which allowed us to differentiate all dosages, even for dodeca-ploid species with up to 13 possible genotypes (Yang et al. 2017). The accurate genotypes were the prerequisite to perform GWAS with allele dosages under consideration in this polyploidy species, which were usually simplified as diploids in previous studies. Third, thanks to the development of GWASpoly (Rosyara et al. 2016), GWAS was performed in sugarcane with six models regarding different gene actions. Depending on the traits and loci analyzed, different gene actions may exist. Without further experimental support, we couldn't evaluate which models are more suitable. To be inclusive, all the six models were tested to maximize the discovery of associated sequence variants.

We identified 217 non-redundant markers and 225 genes associated with the nine yield components. We compared our results with GWAS or linkage mappings in sugarcane. Unfortunately, comparisons were relatively limited since only researches on the same traits and with the associated markers or genes aligned to the sorghum genome can be compared, and only a few such studies were published previously. For the sugar content (Brix), two (*2P5849859* and *2P65250224*) out of three associated markers were found with a distance of 333.1 Kbp and 1,631.3 Kbp, respectively from the associated gene (*Sb02g004780* and *Sb02g028450*) identified previously (Racedo et al. 2016). Recently, quantitative trait loci (QTLs) mapping has been

conducted in a bi-parental population of 173 progeny with multiple QTLs detected for Brix, water content and stalk diameter (Islam et al. submitted). We had associated markers with a distance of 1,615.6 Kbp, 6,556.7 Kbp, 3,165.5 Kbp and 591.1 Kbp from their closest markers (3SNP463 for stalk diameter, 2SNP3040 for Brix, 1SNPUN315 and 3SNP3118 for water content), respectively. However, there were several QTLs from this QTL mapping not detected in our study. Multiple reasons could led to this discrepancy, including plant materials (bi-parental population vs. diversity panel), statistical methods (linkage mapping vs. GWAS), and stringent threshold for significance declared (Logarithm of odd < 3 vs. $P < 0.05$ after Bonferroni correction) etc. We also compared our results with selection genes identified in another study (Yang et al 2018). Eight of our associated genes were also under selection in sugarcane breeding programs. Interestingly, the candidate gene (*Sobic.009G000100*), also under selection and encoding SIN3 histone deacetylase protein, contained four sequence variants (one InDel in intron, and three SNPs in coding region causing one synonymous mutation and two non-synonymous mutations), which were identified for dry weight at least in two GWAS runs or models with $-\log(P)$ ranging from 5.71 to 8.15. The combination of different haplotypes and dosages of this gene might explain the segregation of dry weight in the diversity panel. The comparisons corroborated the reliability of the results. These markers and genes associated with yield components should be fully utilized in sugarcane breeding programs.

ROHs in sugarcane

For the first time, we assessed ROHs in this outcrossing sugarcane using high-density markers. The results of ROHs revealed several interesting phenomena in this polyploidy species. First, the investigation of ROHs could reflect population migration, mating history and structure of sugarcane. Non-*Saccharum* 2 had high NROHs and SROHs but the same AROHs as other

sub-populations. SROHs were correlated with NROHs in non-*Saccharum* 2 (Correlation coefficient 0.99), suggesting increase of the number of ROHs leading to a large SROH in this species. A sudden reduction of population size might happen in non-*Saccharum* 2 and crosses in a population with limited number of individuals could result in a drastic increase of NROHs. In addition, ROHs were not detected in ancient *S.* hybrids and only two ROHs in modern *S.* hybrids. Ancient *S.* hybrids and modern *S.* hybrids were derived from the hybridization between the wild species *S. spontaneum* and the cultivated *S. officinarum* (D'Hont et al. 2002). Compared with their ancestors, the ROH results showed the inter-species crosses broadened the genetic diversity and then reduce ROHs in sugarcane hybrids. The two ROHs in modern *S.* hybrids (One for each accession) were located at Chr07:15.9 Mbp according to the sorghum genome, and were the common ROHs detected in 12 accessions in this diversity panel. The results supported the increase of genomic inbreeding in modern sugarcane cultivars due to multiple backcrosses to *S. officinarum*, and probably their narrow genetic basis. Second, the occurrence of ROHs were likely under selection because the majority of ROHs were common (96.4% were detected in at least two accessions). More interestingly, the genes landing in ROHs were enriched for certain categories with specific gene functions. In our case, the enriched genes in the most common ROHs were involved in plant photosynthesis, suggesting that increase of genomic inbreeding of the regions harboring genes related with photosynthesis would have negative impacts on sugarcane yield. Third, we observed that ROHs were significantly associated with total weight and dry weight, and increase of ROHs had negative influence on these traits. The results further emphasize the importance to broad genetic basis in order to maintain sugarcane production specifically the biomass.

Author contribution statement

JW conceived the experiment. XY performed the research and analyzed the data. JT and SS measured the yield components. ZL helped prepare marker for GWAS. XY and JW wrote the manuscript. All authors reviewed and approved this submission.

Acknowledgments

This research is financially supported by Florida Sugar Cane League and the Office of Science, U.S. Department of Energy (Award number: DE-SC0006995).

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Acevedo A, Tejedor MT, Erazzú LE, Cabada S, Sopena R (2017) Pedigree comparison highlights genetic similarities and potential industrial values of sugarcane cultivars. *Euphytica* 213:121
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655-1664
- Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*
- Berdugo-Cely J, Valbuena RI, Sánchez-Betancourt E, Barrero LS, Yockteng R (2017) Genetic diversity and association mapping in the Colombian Central Collection of *Solanum tuberosum* L. Andigenum group using SNPs markers. *PloS one* 12:e0173039

549 Bock DG, Kantar MB, Caseys C, Matthey-Doret R, Rieseberg LH (2018) Evolution of
550 invasiveness by genetic accommodation. *Nat Ecol Evol* 2(6):991

551 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence
552 data. *Bioinformatics* 30:2114-2120

553 Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF (2018) Runs of homozygosity:
554 windows into population history and trait architecture. *Nat Rev Genet* 19(4):220.

555 Christofidou P, Nelson CP, Nikpay M, Qu L, Li M, Loley C, Debiec R, Braund PS, Denniff M,
556 Charchar FJ (2015) Runs of homozygosity: association with coronary artery disease and gene
557 expression in monocytes and macrophages. *AM J HUM GENET* 97:228-237

558 Dahlquist, E. (2013). *Biomass as energy source: resources, systems and applications*. CRC Press,
559 Taylor & Francis: London, UK

560 Daniels J, Roach BT (1987) Taxonomy and evolution. In: Heinz DJ (ed) *Sugarcane improvement*
561 through breeding. Elsevier Press, Amsterdam, pp 7-84

562 Débibakas S, Rocher S, Garsmeur O, Toubi L, Roques D, D'Hont A, Hoarau J, Daugrois J
563 (2014) Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide
564 association. *Theor Appl Genet* 127:1719-1732

565 Deren CW (1995) Genetic base of US mainland sugarcane. *Crop Sci* 35:1195-1199

566 Dewey DR, Lu K (1959) A Correlation and Path-Coefficient Analysis of Components of Crested
567 Wheatgrass Seed Production 1. *Agron J* 51:515-518

568 D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH;
569 examples of sugarcane and banana. *Cytogenet Genome Res* 109:27-33

570 D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N (1996) Characterisation of
571 the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular
572 cytogenetics. *Mol Gen Genet* 250:405-413

573 D'Hont A, Paulet F, Glaszmann JC (2002) Oligoclonal interspecific origin of 'North Indian' and
574 'Chinese' sugarcane. *Chromosome Res* 10:253-262

575 Epskamp S (2015) semPlot: Unified visualizations of structural equation models. *Struct Equ*
576 *Modeling* 22:474-483

577 FAO (2016) FAOSTAT, FAO Statistical Databases. <http://faostat3.fao.org>. Accessed 7th May
578 2018.

579 Gouy M, Rousselle Y, Chane AT, Anglade A, Royaert S, Nibouche S, Costet L (2015) Genome
580 wide association mapping of agro-morphological and disease resistance traits in sugarcane.
581 *Euphytica* 202:269-284

582 Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY,
583 Nouredine M, Gilbert JR (2005) Complement factor H variant increases the risk of age-related
584 macular degeneration. *Science* 308:419-421

585 Harrell Jr FE (2018) Hmisc: Harrell miscellaneous. [https://cran.r-](https://cran.r-project.org/web/packages/Hmisc/index.html)
586 [project.org/web/packages/Hmisc/index.html](https://cran.r-project.org/web/packages/Hmisc/index.html). Accessed 10th July 2018.

587 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new
588 method for the analysis of genetically structured populations. *BMC Genet* 11:94

589 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.
590 Bioinformatics 24:1403-1405

591 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell
592 A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R,
593 Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics
594 30:1236-1240

595 Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
596 arXiv preprint arXiv:1303.3997

597 Lima M, Garcia A, Oliveira KM, Matsuoka S, Arizono H1, de Souza Jr, C L, De Souza AP
598 (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among
599 genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104:30-38

600 Lyne R, Sullivan J, Butano D, Contrino S, Heimbach J, Hu F, Kalderimis A, Lyne M, N Smith
601 R, Štěpán R, Balakrishnan R (2015) Cross-organism analysis using InterMine. Genesis 53:547-
602 560

603 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
604 Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework
605 for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303

606 Nayak SN, Song J, Villa A, Pathak B, Ayala-Silva T, Yang X, Todd J, Glynn NC, Kuhn DN,
607 Glaz B (2014) Promoting utilization of *Saccharum* spp. genetic resources through genetic
608 diversity analysis and core collection construction. PloS one 9:e110856

609 Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G,
610 Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti
611 AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania
612 A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle
613 AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-
614 ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum
615 bicolor genome and the diversification of grasses. *Nature* 457:551-556

616 Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, do Carmo Panetto,
617 João Cláudio, Ventura RV, Baldi F, da Silva, Marcos Vinícius Gualberto Barbosa (2018)
618 Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos*
619 *indicus*) dairy cattle. *BMC Genomics* 19:34

620 Phan HT, Rybak K, Bertazzoni S, Furuki E, Dinglasan E, Hickey LT, Oliver RP, Tan K (2018)
621 Novel sources of resistance to *Septoria nodorum* blotch in the Vavilov wheat collection
622 identified by genome-wide association studies. *Theor Appl Genet* 1-16

623 Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2009) nlme: Linear and nonlinear mixed
624 effects models. R package version 3:96

625 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
626 genotype data. *Genetics* 155:945-959

627 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De
628 Bakker PI, Daly MJ (2007) PLINK: a tool set for whole-genome association and population-
629 based linkage analyses. *Am J Hum Genet* 81:559-575

630 Raboin L, Pauquet J, Butterfield M, D'Hont A, Glaszmann J (2008) Analysis of genome-wide
631 linkage disequilibrium in the highly polyploid sugarcane. *Theor Appl Genet* 116:701-714

632 Racedo J, Gutierrez L, Perera MF, Ostengo S, Pardo EM, Cuenya MI, Welin B, Castagnaro AP
633 (2016) Genome-wide association mapping of quantitative traits in a breeding population of
634 sugarcane. *BMC Plant Biol* 16:142

635 Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, Ripley MB (2013) Package
636 'mass'. <https://cran.r-project.org/web/packages/MASS/index.html>. Accessed 10th July 2018.

637 Rosseel Y (2012) Lavaan: An R package for structural equation modeling and more. Version
638 0.5-12 (BETA). *J Stat Softw* 48:1-36

639 Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide
640 association studies in autopolyploids and its application to potato. *Plant genome* 9

641 Song J, Yang X, Resende Jr MF, Neves LG, Todd J, Zhang J, Comstock JC, Wang J (2016)
642 Natural allelic variations in highly polyploidy *Saccharum* complex. *Front Plant Sci* 7

643 Sud A, Cooke R, Swerdlow AJ, Houlston RS (2015) Genome-wide homozygosity signature and
644 risk of Hodgkin lymphoma. *Sci Rep* 5:14315

645 Todd J, Sandhu H, Hale A, Glaz B, Wang J (2017) Phenotypic evaluation of a diversity panel
646 selected from the World Collection of Sugarcane (*Saccharum* spp.) and Related Grasses.
647 *Maydica* 62:M19

648 Todd J, Wang J, Glaz B, Sood S, Ayala-Silva T, Nayak SN, Glynn NC, Gutierrez OA, Kuhn DN,
649 Tahir M (2014) Phenotypic characterization of the Miami World Collection of sugarcane

650 (*Saccharum* spp.) and related grasses for selecting a representative core. Genet Resour Crop Evol
651 61:1581-1596

652 Wang J, Roe B, Macmil S, Yu Q, Murray JE, Tang H, Chen C, Najar F, Wiley G, Bowers J
653 (2010) Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. BMC
654 Genomics 11:261

655 Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E (2010)
656 Simultaneously accounting for population structure, genotype by environment interaction, and
657 spatial variation in marker–trait associations in sugarcane. Genome 53:973-981

658 Wei X, Jackson PA, McIntyre CL, Aitken KS, Croft B (2006) Associations between DNA
659 markers and resistance to diseases in sugarcane and effects of population substructure. Theor
660 Appl Genet 114:155-164

661 Yang T, Guo Y, Zhang J, Xu C, Tian Q, Deng H (2015) Genome-Wide Survey of Runs of
662 Homozygosity Identifies Recessive Loci for Bone Mineral Density in Caucasian and Chinese
663 Populations. J Bone Miner Res 30:2119-2126

664 Yang X, Song J, Todd J, Peng Z, Paudel D, Luo Z, Ma X, You Q, Hanson E, Zhao Z, Zhao Y,
665 Zhang J, Ming R, Wang J (2018) Target enrichment sequencing of 307 germplasm accessions
666 identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in
667 sugarcane (*Saccharum* spp.), a “sweet” crop with “bitter” genomes. Plant Biotechnol J doi:
668 10.1111/pbi.12992

669 Yang X, Song J, You Q, Paudel DR, Zhang J, Wang J (2017) Mining sequence variations in
670 representative polyploid sugarcane germplasm accessions. BMC Genomics 18:594

Yi X, Du Z, Su Z (2013) PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Res* 41:W98-103

Zhao X, Bushman BS, Zhang X, Robbins MD, Larson SR, Robins JG, Thomas A (2017) Association of candidate genes with heading date in a diverse *Dactylis glomerata* population. *Plant Science* 265:146-153

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant genome* 1:5-20

Table 1 Pair-wise Pearson's correlation coefficients and broad sense heritability (H^2) for the nine yield components evaluated in the sugarcane diversity panel.

	H^2	Stalk number	Stalk diameter	Internode length	Leaf width	Leaf length	Brix	Dry weight	Total Weight
Stalk number	0.88								
Stalk diameter	0.91	-0.69***							
Internode length	0.68	0.24***	-0.31***						
Leaf width	0.82	-0.62***	0.82***	-0.37***					
Leaf length	0.67	-0.38***	0.53***	-0.13	0.39***				
Brix	0.68	-0.38***	0.54***	-0.18*	0.35***	0.2**			
Dry weight	0.42	0.31***	-0.24***	0.34***	-0.34***	-0.06	-0.05		
Total Weight	0.45	-0.03	0.18*	0.19**	0.00	0.19**	0.23***	0.78***	
Water content	0.51	-0.20**	0.38***	0.07	0.19*	0.29***	0.34***	0.58***	0.92***

Note: * $P < 0.01$; ** $P < 0.001$; *** $P < 0.0001$.

Table 2 Path coefficients showing direct (Diagonal) and indirect effects (Non-diagonal) of six yield components on sugarcane dry weight.

	Stalk number	Stalk diameter	Internode length	Leaf width	Leaf length	Brix
Stalk number	0.30	-0.21	0.07	-0.19	-0.11	-0.11
Stalk diameter	-0.10	0.14	-0.04	0.11	0.07	0.08
Internode length	0.06	-0.08	0.25	-0.02	-0.03	-0.05
Leaf width	0.14	-0.18	0.08	-0.22	-0.09	-0.08
Leaf length	-0.03	0.04	-0.01	0.03	0.08	0.02
Brix	-0.03	0.04	-0.01	0.03	0.02	0.08

Table 3 Path coefficients showing direct (Diagonal) and indirect effects (Non-diagonal) of five yield components on sugar content (Brix).

	Stalk number	Stalk diameter	Internode length	Leaf width	Leaf length
Stalk number	-0.05	0.03	-0.01	0.03	0.02
Stalk diameter	-0.57	0.83	-0.26	0.68	0.44
Internode length	-0.02	0.02	-0.07	0.01	0.01
Leaf width	0.20	-0.27	0.12	-0.33	-0.13
Leaf length	0.06	-0.08	0.02	-0.06	-0.15

Table 4 Summary of single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) identified in the sugarcane diversity panel.

Sequence variants	SNPs	InDels
Raw variants	25.3 M	2.1 M

Sample-by-sample filtering	9.9 M	880.0 K
Bi-allelic variants	8.6 M	749.7 K
Remove rare variants (Minor Allele frequency < 0.05)	840.3 K	186.9 K
Remove variants with low call rates (< 95%)	65.0 K	9.9 K
Total markers for genome-wide association study		74.9 K

Note: M, million; K, thousand.

Table 5 Number of associated sequence variants identified in different runs and models of genome-wide association study.

Runs ^a /Models	General	Additive	1-dom- alt	1-dom- ref	diplo- general	diplo- additive	Total
Full	99	9	25	14	35	27	128
Sacc	71	2	2	2	10	2	77
Spon	31	1	0	0	0	0	30
Hybrid	13	0	0	0	0	0	13
P8	45	6	9	7	10	7	58

Note: ^a GWAS runs in Full set and several subests. Full = all accessions in the diversity panel; Sacc = 273 accessions from *S. spontaneum*, *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; Spon = 100 accessions from *S. spontaneum*; Hybrid = 173 accessions from *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; and P8 = 162 accessions with the same ploidy level (Ploidy = 8).

Table 6 Statistical summary of runs of homozygosity (ROHs) identified in the sugarcane diversity panel.

	Total number of accessions	Accessions with ROHs (%)	NROHs (Std)	SROHs (Kbp) (Std)	Average ROHs (Kbp) (Std)
<i>Saccharum</i> <i>spontaneum</i>	74	22 (29.8)	1.5 (0.7)	2,132.2 (908.1)	1,394.3 (245.9)
<i>Saccharum</i> <i>officinarum</i>	36	10 (27.8)	1.7 (0.8)	2,476.8 (1,244.8)	1,460.1 (315.5)
Ancient <i>Saccharum</i> hybrids	14	0 (0)	NA	NA	NA
Modern <i>Saccharum</i> hybrids	75	2 (2.7)	1.0 (0)	1,162.1 (0.2)	1,162.1 (0.2)
Non- <i>Saccharum</i> 1	4	4 (100)	3.5 (2.4)	5,396.0 (3,267.5)	1,649.7 (237.1)
Non- <i>Saccharum</i> 2	8	8 (100)	26.9	45,117.7	1,677.3 (144.1)

			(11.5)	(20,738.3)	
Total	211	46 (21,8)	6.1 (10.7)	9,924.5	1,469.9 (269.5)
				(18,323,3)	

Note: NROHs, number of runs of homozygosity; SROHs, sum of total length of runs of homozygosity.

Figure legends:

Figure 1 Population structure analysis of the sugarcane diversity panel based on 13.8 K single nucleotide polymorphisms using discriminant analysis of principal components. (A) Bayesian Information Criteria (BIC) vs. number of clusters in k-means clustering. (B) Projection of the sugarcane diversity panel using the first three linear discriminants (LDs). Spon = *S. spontaneum*; Off = *S. officinarum*; Hybrid = modern *S. hybrid*; Barb = *S. barberi*; Sine = *S. sinence*; Non sacc = Non *saccharum*.

Figure 2 Example of significant marker-trait associations identified in the sugarcane diversity panel (Full set) for stalk diameter (Diameter) using general model (A), additive model (B), 1-dom-alt model (C), 1-dom-ref model (D), diplo-general model (E), and diplo-additive model (F).

Figure 3 Summary of runs of homozygosity (ROHs) in the sugarcane diversity panel. A) Distribution of common ROHs according to the sorghum genome; B) Total number of ROHs; C) sum total length of ROHs; D) average length of ROHs. Spon = *S. spontaneum*; Off = *S.*

officinarum; Hybrid = modern *S. hybrid*; Misc = *Miscanthus*; Eria = *Erianthus*.

Online supporting information:

Figure S1 Distribution of the nine yield components in the sugarcane diversity panel, (A) Stalk number (Stalk); (B) Internode length (Internode); (C) Stalk diameter (Diameter); (D) Leaf width (Width); (E) Leaf length (length); (F) Brix; (G) Total weight (Weight); (H) Dry weight (Dry); (I) Water content (Water).

Figure S2 Significant marker-trait associations identified in full set of the sugarcane diversity panel with 300 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

Figure S3 Significant marker-trait associations identified in *Saccharum* set of the sugarcane diversity panel with 273 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

Figure S4 Significant marker-trait associations identified in *Saccharum spontaneum* set of the sugarcane diversity panel with 100 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

Figure S5 Significant marker-trait associations identified in sugarcane hybrid set of the sugarcane diversity panel with 173 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

Figure S6 Significant marker-trait associations identified in the sugarcane diversity panel with 162 accessions with ploidy = 8. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

Table S1 Details of genotypes included in the sugarcane diversity panel.

Table S2 Associated DNA markers identified by genome-wide association study (GWAS) in sugarcane diversity panel.

Table S3 Non-redundant associated markers and candidate genes with genome-wide associations study (GWAS) hits.

Figure 1 Population structure analysis of the sugarcane diversity panel based on 13.8 K single nucleotide polymorphisms using discriminant analysis of principal components. (A) Bayesian Information Criteria (BIC) vs. number of clusters in k-means clustering. (B) Projection of the sugarcane diversity panel using the first three linear discriminants (LDs). Spon = *S. spontaneum*;

768 Off = *S. officinarum*; Hybrid = modern *S. hybrid*; Barb = *S. barberi*, Sine = *S. sinence*; Non sacc
769 = Non *saccharum*.

770

771

772 **Figure 2** Example of significant marker-trait associations identified in the sugarcane diversity
773 panel (Full set) for stalk diameter (Diameter) using general model (A), additive model (B), 1-
774 dom-alt model (C), 1-dom-ref model (D), diplo-general model (E), and diplo-additive model (F).

775

776

777

Figure 3 Summary of runs of homozygosity (ROHs) in the sugarcane diversity panel. A) Distribution of common ROHs according to the sorghum genome; B) Total number of ROHs; C) sum total length of ROHs; D) average length of ROHs. Spon = *S. spontaneum*; Off = *S. officinarum*; Hybrid = modern *S. hybrid*; Misc = *Miscanthus*; Eria = *Erianthus*.

785 **Table S1** Details of genotypes included in the sugarcane diversity panel.

Sample ID	Genus^a	Species^b	Clone name^c	Group
53-39	<i>Saccharum</i>	<i>barberi</i>	Pathri	Ancient <i>S.</i> hybrids
53-17	<i>Saccharum</i>	<i>barberi</i>	Ketari	Ancient <i>S.</i> hybrids
53-8	<i>Saccharum</i>	<i>barberi</i>	Ganapathy	Ancient <i>S.</i> hybrids
53-32	<i>Saccharum</i>	<i>barberi</i>	Matna	Ancient <i>S.</i> hybrids
54-9	<i>Saccharum</i>	<i>barberi</i>	Sunnabile	Ancient <i>S.</i> hybrids
54-8	<i>Saccharum</i>	<i>barberi</i>	Sonabile	Ancient <i>S.</i> hybrids
10-2	<i>Saccharum</i>	<i>officinarum</i>	NG 77-016	Ancient <i>S.</i> hybrids
12-8	<i>Saccharum</i>	<i>officinarum</i>	Pundia	Ancient <i>S.</i> hybrids
41-146	pending	pending	SM 8116	Ancient <i>S.</i> hybrids
67-21	<i>Saccharum</i>	<i>robustum</i>	Unknown (1995: R65-P35)	Ancient <i>S.</i> hybrids
71-32	<i>Saccharum</i>	<i>sinense</i>	Tekcha Okinawa	Ancient <i>S.</i> hybrids
71-29	<i>Saccharum</i>	<i>sinense</i>	Tekcha	Ancient <i>S.</i> hybrids
71-5	<i>Saccharum</i>	<i>sinense</i>	China	Ancient <i>S.</i> hybrids
71-6	<i>Saccharum</i>	<i>sinense</i>	Chuk Che	Ancient <i>S.</i> hybrids
71-1	<i>Saccharum</i>	<i>sinense</i>	Agaul	Ancient <i>S.</i> hybrids
45-89	Unknown	Unknown	Unknown (2008: R45-P089)	Ancient <i>S.</i> hybrids
71-28	Unknown	Unknown	Tanzhou Bamboo Cane	Ancient <i>S.</i> hybrids
53-36	<i>Saccharum</i>	<i>barberi</i>	Nargori	Modern <i>S.</i> hybrids
53-28	<i>Saccharum</i>	<i>barberi</i>	Manga Coimbatore	Modern <i>S.</i> hybrids
67-38	<i>Saccharum</i>	<i>edule</i>	IJ 76-551	Modern <i>S.</i> hybrids
69-14	<i>Saccharum</i>	<i>formerly off.</i>	IM 76-249	Modern <i>S.</i> hybrids
69-16	<i>Saccharum</i>	<i>formerly rob.</i>	IM 76-249	Modern <i>S.</i> hybrids
57-43	<i>Saccharum</i>	hybrid	Q 050	Modern <i>S.</i> hybrids
55-17	<i>Saccharum</i>	hybrid	CO 0513	Modern <i>S.</i> hybrids
57-23	<i>Saccharum</i>	hybrid	POJ 2803	Modern <i>S.</i> hybrids
56-5	<i>Saccharum</i>	hybrid	Diamond 10	Modern <i>S.</i> hybrids
55-29	<i>Saccharum</i>	hybrid	CP 44-0101	Modern <i>S.</i> hybrids
58-26	<i>Saccharum</i>	hybrid	Unknown (1995: R55-P07)	Modern <i>S.</i> hybrids
55-23	<i>Saccharum</i>	hybrid	COS 321	Modern <i>S.</i> hybrids
57-10	<i>Saccharum</i>	hybrid	POJ 1228	Modern <i>S.</i> hybrids
55-7	<i>Saccharum</i>	hybrid	CO 0301	Modern <i>S.</i> hybrids
69-25	<i>Saccharum</i>	hybrid	CO 0727	Modern <i>S.</i> hybrids
56-22	<i>Saccharum</i>	hybrid	LCP 86-454	Modern <i>S.</i> hybrids
55-16	<i>Saccharum</i>	hybrid	CO 0475	Modern <i>S.</i> hybrids
58-29	<i>Saccharum</i>	hybrid	US 67-037-01	Modern <i>S.</i> hybrids
57-29	<i>Saccharum</i>	hybrid	POJ 2946	Modern <i>S.</i> hybrids
2-11	<i>Saccharum</i>	hybrid	BO 11	Modern <i>S.</i> hybrids
55-43	<i>Saccharum</i>	hybrid	CR 00440	Modern <i>S.</i> hybrids
55-13	<i>Saccharum</i>	hybrid	CO 0421	Modern <i>S.</i> hybrids

56-11	<i>Saccharum</i>	hybrid	F 148	Modern <i>S.</i> hybrids
4-5	<i>Saccharum</i>	hybrid	Q 058	Modern <i>S.</i> hybrids
54-30	<i>Saccharum</i>	hybrid	B 54142	Modern <i>S.</i> hybrids
58-11	<i>Saccharum</i>	hybrid	Ragnar	Modern <i>S.</i> hybrids
55-35	<i>Saccharum</i>	hybrid	CP 66-0346	Modern <i>S.</i> hybrids
56-14	<i>Saccharum</i>	hybrid	H 32-8560	Modern <i>S.</i> hybrids
55-34	<i>Saccharum</i>	hybrid	CP 65-0357	Modern <i>S.</i> hybrids
57-1	<i>Saccharum</i>	hybrid	Pepe Cuca	Modern <i>S.</i> hybrids
56-42	<i>Saccharum</i>	hybrid	Ogles Selection	Modern <i>S.</i> hybrids
56-36	<i>Saccharum</i>	hybrid	NCO 293	Modern <i>S.</i> hybrids
58-28	<i>Saccharum</i>	hybrid	Unknown (1995: R65-P18)	Modern <i>S.</i> hybrids
CP88-1762	<i>Saccharum</i>	hybrid	CP88-1762	Modern <i>S.</i> hybrids
CP95-1039	<i>Saccharum</i>	hybrid	CP95-1039	Modern <i>S.</i> hybrids
CP89-2143	<i>Saccharum</i>	hybrid	CP89-2143	Modern <i>S.</i> hybrids
CP96-1252	<i>Saccharum</i>	hybrid	CP96-1252	Modern <i>S.</i> hybrids
CP03-1912	<i>Saccharum</i>	hybrid	CP03-1912	Modern <i>S.</i> hybrids
CP01-2390	<i>Saccharum</i>	hybrid	CP01-2390	Modern <i>S.</i> hybrids
CP00-1101	<i>Saccharum</i>	hybrid	CP00-1101	Modern <i>S.</i> hybrids
CP72-2086	<i>Saccharum</i>	hybrid	CP72-2086	Modern <i>S.</i> hybrids
CP78-1628	<i>Saccharum</i>	hybrid	CP78-1628	Modern <i>S.</i> hybrids
43-13	<i>Saccharum</i>	<i>kanashiroi</i>	<i>S. kanashiroi</i>	Modern <i>S.</i> hybrids
12-5	<i>Saccharum</i>	<i>officinarum</i>	P-MAG-84-28	Modern <i>S.</i> hybrids
14-15	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R65-P13)	Modern <i>S.</i> hybrids
14-2	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R45-P120)	Modern <i>S.</i> hybrids
3-1	<i>Saccharum</i>	<i>officinarum</i>	IA 3330	Modern <i>S.</i> hybrids
4-14	<i>Saccharum</i>	<i>officinarum</i>	Javari Kabbu	Modern <i>S.</i> hybrids
2-21	<i>Saccharum</i>	<i>officinarum</i>	Hinahina 18	Modern <i>S.</i> hybrids
1-9	<i>Saccharum</i>	<i>officinarum</i>	Bamboo Amarilla	Modern <i>S.</i> hybrids
13-17	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R43-P023)	Modern <i>S.</i> hybrids
1-18	<i>Saccharum</i>	<i>officinarum</i>	Black Innes	Modern <i>S.</i> hybrids
7-20	<i>Saccharum</i>	<i>officinarum</i>	NG 51-105	Modern <i>S.</i> hybrids
3-9	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-418A PINK STALK	Modern <i>S.</i> hybrids
13-15	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R37-P070)	Modern <i>S.</i> hybrids
12-21	<i>Saccharum</i>	<i>officinarum</i>	Spaansch	Modern <i>S.</i> hybrids
10-14	<i>Saccharum</i>	<i>officinarum</i>	NG 77-081	Modern <i>S.</i> hybrids
14-16	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R65-P39)	Modern <i>S.</i> hybrids
13-14	<i>Saccharum</i>	<i>officinarum</i>	UM 68-015	Modern <i>S.</i> hybrids
2-22	<i>Saccharum</i>	<i>officinarum</i>	Hind's Special	Modern <i>S.</i> hybrids
11-16	<i>Saccharum</i>	<i>officinarum</i>	Oi Kung	Modern <i>S.</i> hybrids
14-4	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R45-P123)	Modern <i>S.</i> hybrids
12-1	<i>Saccharum</i>	<i>officinarum</i>	P-MAG-84-02	Modern <i>S.</i> hybrids

3-8	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-418	Modern <i>S.</i> hybrids
11-14	<i>Saccharum</i>	<i>officinarum</i>	<i>S. officinarum</i> (8284)	Modern <i>S.</i> hybrids
6-5	<i>Saccharum</i>	<i>officinarum</i>	NG 28-033	Modern <i>S.</i> hybrids
2-15	<i>Saccharum</i>	<i>officinarum</i>	Guam A	Modern <i>S.</i> hybrids
2-4	<i>Saccharum</i>	<i>officinarum</i>	F 36-819	Modern <i>S.</i> hybrids
4-3	<i>Saccharum</i>	<i>officinarum</i>	IK 76-108	Modern <i>S.</i> hybrids
11-13	<i>Saccharum</i>	<i>officinarum</i>	<i>S. officinarum</i> (8276)	Modern <i>S.</i> hybrids
6-3	<i>Saccharum</i>	<i>officinarum</i>	NG 28-020	Modern <i>S.</i> hybrids
13-23	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R45-P109)	Modern <i>S.</i> hybrids
11-20	<i>Saccharum</i>	<i>officinarum</i>	Padangsche Dark Red	Modern <i>S.</i> hybrids
13-21	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R45-P072)	Modern <i>S.</i> hybrids
14-8	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R56-P50)	Modern <i>S.</i> hybrids
69-24	<i>Saccharum</i>	<i>officinarum</i>	BADILA	Modern <i>S.</i> hybrids
7-19	<i>Saccharum</i>	<i>officinarum</i>	NG 51-099	Modern <i>S.</i> hybrids
6-11	<i>Saccharum</i>	<i>officinarum</i>	NG 28-062	Modern <i>S.</i> hybrids
13-6	<i>Saccharum</i>	<i>officinarum</i>	Sylva	Modern <i>S.</i> hybrids
4-1	<i>Saccharum</i>	<i>officinarum</i>	IK 76-069	Modern <i>S.</i> hybrids
2-6	<i>Saccharum</i>	<i>officinarum</i>	IK 76-065	Modern <i>S.</i> hybrids
4-23	<i>Saccharum</i>	<i>officinarum</i>	Manjav	Modern <i>S.</i> hybrids
10-21	<i>Saccharum</i>	<i>officinarum</i>	NG 77-135	Modern <i>S.</i> hybrids
13-16	<i>Saccharum</i>	<i>officinarum</i>	Unknown (1995: R37-P071)	Modern <i>S.</i> hybrids
41-150	pending	pending	SM 8136	Modern <i>S.</i> hybrids
41-145	pending	pending	B 69691	Modern <i>S.</i> hybrids
54-43	<i>Saccharum</i>	hybrid	CL 41-0223	Modern <i>S.</i> hybrids
70-17	pending	pending	RSB 90-22	Modern <i>S.</i> hybrids
69-22	pending	pending	B 73348	Modern <i>S.</i> hybrids
70-9	pending	pending	NG 26-011	Modern <i>S.</i> hybrids
69-32	pending	pending	CP 88-1762	Modern <i>S.</i> hybrids
69-38	pending	pending	F 154	Modern <i>S.</i> hybrids
69-29	pending	pending	CP 73-1547	Modern <i>S.</i> hybrids
66-37	<i>Saccharum</i>	<i>robustum</i>	NG 77-107	Modern <i>S.</i> hybrids
66-41	<i>Saccharum</i>	<i>robustum</i>	NG 77-196	Modern <i>S.</i> hybrids
66-6	<i>Saccharum</i>	<i>robustum</i>	IJ 76-547	Modern <i>S.</i> hybrids
65-44	<i>Saccharum</i>	<i>robustum</i>	IJ 76-435	Modern <i>S.</i> hybrids
71-8	<i>Saccharum</i>	<i>sinense</i>	Desi Paunda	Modern <i>S.</i> hybrids
71-40	<i>Saccharum</i>	<i>sinense</i>	Yuegsen	Modern <i>S.</i> hybrids
71-7	<i>Saccharum</i>	<i>sinense</i>	Chynia	Modern <i>S.</i> hybrids
39-28	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-073	Modern <i>S.</i> hybrids
45-17	<i>Saccharum</i>	<i>spontaneum</i>	SES 264	Modern <i>S.</i> hybrids
45-62	<i>Saccharum</i>	<i>spontaneum</i>	Timor Wild	Modern <i>S.</i> hybrids
39-138	<i>Saccharum</i>	<i>spontaneum</i>	SM 7916	Modern <i>S.</i> hybrids

69-13	<i>Erianthus</i> contamination	Unknown	Erianthus contamination	Modern <i>S.</i> hybrids
68-20	<i>Saccharum</i>	Unknown	MOL 6125	Modern <i>S.</i> hybrids
68-21	<i>Saccharum</i>	Unknown	MOL 6446	Modern <i>S.</i> hybrids
5-13	Unknown	Unknown	R 469	Modern <i>S.</i> hybrids
5-14	Unknown	Unknown	R 570	Modern <i>S.</i> hybrids
7-11	Unknown	Unknown	NG 28-285	Modern <i>S.</i> hybrids
68-35	Unknown	Unknown	Unknown (1995: R65-P47)	Modern <i>S.</i> hybrids
45-107	Unknown	Unknown	Unknown (2008: R45-P107)	Modern <i>S.</i> hybrids
43-92	Unknown	Unknown	Unknown (2009: R43-P092)	Modern <i>S.</i> hybrids
45-59	Unknown	Unknown	Unknown (2008: R45-P059)	Modern <i>S.</i> hybrids
37-1	<i>Miscanthus</i>	<i>floridulus</i>	US 56-0022-03	Non <i>Saccharum</i> 1
56-17	<i>Saccharum</i>	hybrid	H 53-4596	Non <i>Saccharum</i> 1
37-14	<i>Sorghum</i>	<i>plumosum</i>	US 71-0017	Non <i>Saccharum</i> 1
43-35	<i>Saccharum</i>	<i>procerum</i>	Kalimpong	Non <i>Saccharum</i> 1
43-95	<i>Saccharum</i>	<i>rufipilum</i>	US 57-060-02	Non <i>Saccharum</i> 1
37-5	<i>Miscanthus</i>	<i>sinensis</i>	US 64-007-01	Non <i>Saccharum</i> 1
37-13	<i>Miscanthus</i>	<i>sinensis</i>	US 64-004-02	Non <i>Saccharum</i> 1
39-61	<i>Saccharum</i>	<i>spontaneum</i>	Karenko	Non <i>Saccharum</i> 1
41-54	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-161	Non <i>Saccharum</i> 1
37-47	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P047)	Non <i>Saccharum</i> 1
37-55	<i>Erianthus</i>	Unknown	Unknown (2009: R37-P055)	Non <i>Saccharum</i> 2
43-39	<i>Saccharum</i>	<i>arundinaceum</i>	US 57-003-03	Non <i>Saccharum</i> 2
53-33	<i>Saccharum</i>	<i>barberi</i>	Matna Shahj	Non <i>Saccharum</i> 2
43-44	<i>Saccharum</i>	<i>bengalense</i>	US 67-009-02	Non <i>Saccharum</i> 2
43-20	<i>Saccharum</i>	<i>bengalense</i>	IND 81-047	Non <i>Saccharum</i> 2
43-42	<i>Saccharum</i>	<i>brevifolius</i> var. <i>brevifolius</i>	US 63-017	Non <i>Saccharum</i> 2
43-40	<i>Saccharum</i>	<i>kanashiroi</i>	US 57-005-02	Non <i>Saccharum</i> 2
43-6	<i>Saccharum</i>	<i>ravennae</i>	SES 305	Non <i>Saccharum</i> 2
43-91	<i>Saccharum</i>	<i>rufipilum</i>	US 61-012-01	Non <i>Saccharum</i> 2
37-54	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P054)	Non <i>Saccharum</i> 2
43-74	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R43-P074)	Non <i>Saccharum</i> 2
41-72	<i>Saccharum</i>	<i>spontaneum</i>	Isiolo	Non <i>Saccharum</i> 2
41-73	<i>Saccharum</i>	<i>spontaneum</i>	Karenko Large	Non <i>Saccharum</i> 2
45-31	<i>Saccharum</i>	<i>spontaneum</i>	SES 288	Non <i>Saccharum</i> 2
41-48	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-062	Non <i>Saccharum</i> 2
37-166	<i>Saccharum</i>	<i>spontaneum</i>	US 58-005-03	Non <i>Saccharum</i> 2
37-50	<i>Erianthus</i>	Unknown	Unknown (2009: R37-P050)	Non <i>Saccharum</i> 2
43-43	<i>Saccharum</i>	Unknown	US 63-052	Non <i>Saccharum</i> 2
37-75	Unknown	Unknown	Unknown (2009: R37-P075)	Non <i>Saccharum</i> 2

67-33	<i>Miscanthus</i>	hybrid	Fiji 59	<i>S. officinarum</i>
11-8	<i>Saccharum</i>	<i>officinarum</i>	NG 96-024	<i>S. officinarum</i>
2-7	<i>Saccharum</i>	<i>officinarum</i>	Fiji	<i>S. officinarum</i>
9-22	<i>Saccharum</i>	<i>officinarum</i>	NG 57-259	<i>S. officinarum</i>
5-6	<i>Saccharum</i>	<i>officinarum</i>	Nanahu 48	<i>S. officinarum</i>
5-3	<i>Saccharum</i>	<i>officinarum</i>	MOL 1238	<i>S. officinarum</i>
3-6	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-324	<i>S. officinarum</i>
1-17	<i>Saccharum</i>	<i>officinarum</i>	Black Fiji	<i>S. officinarum</i>
6-21	<i>Saccharum</i>	<i>officinarum</i>	NG 28-212	<i>S. officinarum</i>
7-17	<i>Saccharum</i>	<i>officinarum</i>	NG 51-058	<i>S. officinarum</i>
6-12	<i>Saccharum</i>	<i>officinarum</i>	NG 28-080	<i>S. officinarum</i>
1-10	<i>Saccharum</i>	<i>officinarum</i>	Bamboo Blanca	<i>S. officinarum</i>
6-18	<i>Saccharum</i>	<i>officinarum</i>	NG 28-099	<i>S. officinarum</i>
3-15	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-475	<i>S. officinarum</i>
2-1	<i>Saccharum</i>	<i>officinarum</i>	IN 84-068	<i>S. officinarum</i>
2-12	<i>Saccharum</i>	<i>officinarum</i>	Fiji 47	<i>S. officinarum</i>
7-6	<i>Saccharum</i>	<i>officinarum</i>	NG 28-266	<i>S. officinarum</i>
9-1	<i>Saccharum</i>	<i>officinarum</i>	NG 57-143	<i>S. officinarum</i>
11-2	<i>Saccharum</i>	<i>officinarum</i>	NG 77-154	<i>S. officinarum</i>
5-21	<i>Saccharum</i>	<i>officinarum</i>	NG 21-049	<i>S. officinarum</i>
8-18	<i>Saccharum</i>	<i>officinarum</i>	NG 57-099	<i>S. officinarum</i>
3-16	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-478	<i>S. officinarum</i>
2-3	<i>Saccharum</i>	<i>officinarum</i>	EK 02	<i>S. officinarum</i>
3-14	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-470	<i>S. officinarum</i>
11-11	<i>Saccharum</i>	<i>officinarum</i>	NH 70-069-3	<i>S. officinarum</i>
2-18	<i>Saccharum</i>	<i>officinarum</i>	Hawaiian Original	<i>S. officinarum</i>
10-6	<i>Saccharum</i>	<i>officinarum</i>	NG 77-042	<i>S. officinarum</i>
2-9	<i>Saccharum</i>	<i>officinarum</i>	Fiji 24	<i>S. officinarum</i>
6-4	<i>Saccharum</i>	<i>officinarum</i>	NG 28-032	<i>S. officinarum</i>
3-13	<i>Saccharum</i>	<i>officinarum</i>	IJ 76-429	<i>S. officinarum</i>
5-10	<i>Saccharum</i>	<i>officinarum</i>	NC 042	<i>S. officinarum</i>
9-15	<i>Saccharum</i>	<i>officinarum</i>	NG 57-219	<i>S. officinarum</i>
10-3	<i>Saccharum</i>	<i>officinarum</i>	NG 77-018	<i>S. officinarum</i>
10-7	<i>Saccharum</i>	<i>officinarum</i>	NG 77-043	<i>S. officinarum</i>
7-18	<i>Saccharum</i>	<i>officinarum</i>	NG 51-095	<i>S. officinarum</i>
12-15	<i>Saccharum</i>	<i>officinarum</i>	<i>S. officinarum</i> (8275)	<i>S. officinarum</i>
4-4	<i>Saccharum</i>	<i>officinarum</i>	IM 76-245	<i>S. officinarum</i>
37-78	pending	pending	US 56-271-02	<i>S. officinarum</i>
66-28	<i>Saccharum</i>	<i>robustum</i>	NG 57-054	<i>S. officinarum</i>
66-16	<i>Saccharum</i>	<i>robustum</i>	IN 84-076	<i>S. officinarum</i>
65-42	<i>Saccharum</i>	<i>robustum</i>	IJ 76-412	<i>S. officinarum</i>

66-12	<i>Saccharum</i>	<i>robustum</i>	IN 84-045	<i>S. officinarum</i>
67-17	<i>Saccharum</i>	<i>robustum</i>	Unknown (1995: R65-P27)	<i>S. officinarum</i>
66-17	<i>Saccharum</i>	<i>robustum</i>	IS 76-184	<i>S. officinarum</i>
43-178	<i>Saccharum</i>	<i>robustum</i>	NG 57-208	<i>S. officinarum</i>
66-25	<i>Saccharum</i>	<i>robustum</i>	NG 57-012	<i>S. officinarum</i>
8-16	Unknown	Unknown	Unknown	<i>S. officinarum</i>
43-41	<i>Saccharum</i>	<i>arundinaceum</i>	Teboe Glonggong	<i>S. spontaneum</i>
37-30	<i>Sorghum</i>	<i>arundinaceum</i>	US 47-0017	<i>S. spontaneum</i>
41-148	pending	pending	Glagah Kloet	<i>S. spontaneum</i>
41-171	pending	pending	SES 011	<i>S. spontaneum</i>
66-9	<i>Saccharum</i>	<i>robustum</i>	IM 76-232	<i>S. spontaneum</i>
41-51	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-142	<i>S. spontaneum</i>
39-120	<i>Saccharum</i>	<i>spontaneum</i>	SES 196	<i>S. spontaneum</i>
37-53	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P053)	<i>S. spontaneum</i>
45-11	<i>Saccharum</i>	<i>spontaneum</i>	SES 220	<i>S. spontaneum</i>
39-176	<i>Saccharum</i>	<i>spontaneum</i>	IN 76-086	<i>S. spontaneum</i>
41-176	<i>Saccharum</i>	<i>spontaneum</i>	SLC 92-78	<i>S. spontaneum</i>
41-117	<i>Saccharum</i>	<i>spontaneum</i>	S 66-097	<i>S. spontaneum</i>
39-68	<i>Saccharum</i>	<i>spontaneum</i>	NG 51-025	<i>S. spontaneum</i>
41-18	<i>Saccharum</i>	<i>spontaneum</i>	IK 76-067	<i>S. spontaneum</i>
45-75	<i>Saccharum</i>	<i>spontaneum</i>	US 4625	<i>S. spontaneum</i>
39-162	<i>Saccharum</i>	<i>spontaneum</i>	US 78-508	<i>S. spontaneum</i>
39-49	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-191	<i>S. spontaneum</i>
41-92	<i>Saccharum</i>	<i>spontaneum</i>	PAL 84-07	<i>S. spontaneum</i>
37-66	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P066)	<i>S. spontaneum</i>
39-154	<i>Saccharum</i>	<i>spontaneum</i>	Tongza	<i>S. spontaneum</i>
39-39	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-138	<i>S. spontaneum</i>
41-44	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-088	<i>S. spontaneum</i>
39-177	<i>Saccharum</i>	<i>spontaneum</i>	SES 197 A	<i>S. spontaneum</i>
39-169	<i>Saccharum</i>	<i>spontaneum</i>	PPGN 84-08	<i>S. spontaneum</i>
39-140	<i>Saccharum</i>	<i>spontaneum</i>	Shoaguan	<i>S. spontaneum</i>
39-63	<i>Saccharum</i>	<i>spontaneum</i>	MOL 1032	<i>S. spontaneum</i>
45-34	<i>Saccharum</i>	<i>spontaneum</i>	SES 517	<i>S. spontaneum</i>
39-96	<i>Saccharum</i>	<i>spontaneum</i>	PTAR 84-07	<i>S. spontaneum</i>
41-136	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R41-P136)	<i>S. spontaneum</i>
41-125	<i>Saccharum</i>	<i>spontaneum</i>	SES 014	<i>S. spontaneum</i>
39-21	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-007	<i>S. spontaneum</i>
37-65	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P065)	<i>S. spontaneum</i>
37-56	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P056)	<i>S. spontaneum</i>
39-133	<i>Saccharum</i>	<i>spontaneum</i>	SES 519	<i>S. spontaneum</i>
39-45	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-170	<i>S. spontaneum</i>

39-112	<i>Saccharum</i>	<i>spontaneum</i>	S. spont # 89	<i>S. spontaneum</i>
39-43	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-157	<i>S. spontaneum</i>
39-93	<i>Saccharum</i>	<i>spontaneum</i>	PPGN 84-05	<i>S. spontaneum</i>
39-38	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-136	<i>S. spontaneum</i>
43-170	<i>Saccharum</i>	<i>spontaneum</i>	SLC 92-81	<i>S. spontaneum</i>
39-27	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-055	<i>S. spontaneum</i>
41-122	<i>Saccharum</i>	<i>spontaneum</i>	SES 004 A	<i>S. spontaneum</i>
39-72	<i>Saccharum</i>	<i>spontaneum</i>	Okinawa # 13	<i>S. spontaneum</i>
39-163	<i>Saccharum</i>	<i>spontaneum</i>	US 78-511	<i>S. spontaneum</i>
41-74	<i>Saccharum</i>	<i>spontaneum</i>	Kepandjen	<i>S. spontaneum</i>
39-14	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-022	<i>S. spontaneum</i>
41-52	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-144	<i>S. spontaneum</i>
45-97	<i>Saccharum</i>	<i>spontaneum</i>	US 78-500	<i>S. spontaneum</i>
39-11	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-012 F14	<i>S. spontaneum</i>
45-27	<i>Saccharum</i>	<i>spontaneum</i>	SES 365	<i>S. spontaneum</i>
39-113	<i>Saccharum</i>	<i>spontaneum</i>	S. spont Iran	<i>S. spontaneum</i>
43-68	<i>Saccharum</i>	<i>spontaneum</i>	17-95	<i>S. spontaneum</i>
43-168	<i>Saccharum</i>	<i>spontaneum</i>	SLC 92-94	<i>S. spontaneum</i>
45-93	<i>Saccharum</i>	<i>spontaneum</i>	US 60-004-06	<i>S. spontaneum</i>
41-112	<i>Saccharum</i>	<i>spontaneum</i>	PTAR 84-02	<i>S. spontaneum</i>
39-24	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-030	<i>S. spontaneum</i>
39-30	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-081	<i>S. spontaneum</i>
39-156	<i>Saccharum</i>	<i>spontaneum</i>	US 56-015-08	<i>S. spontaneum</i>
39-147	<i>Saccharum</i>	<i>spontaneum</i>	THA 83-149	<i>S. spontaneum</i>
41-96	<i>Saccharum</i>	<i>spontaneum</i>	PCAV 84-20	<i>S. spontaneum</i>
37-44	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P044)	<i>S. spontaneum</i>
39-65	<i>Saccharum</i>	<i>spontaneum</i>	Mandalay # 1	<i>S. spontaneum</i>
39-115	<i>Saccharum</i>	<i>spontaneum</i>	SES 072	<i>S. spontaneum</i>
39-135	<i>Saccharum</i>	<i>spontaneum</i>	SES 602	<i>S. spontaneum</i>
45-30	<i>Saccharum</i>	<i>spontaneum</i>	SES 275	<i>S. spontaneum</i>
43-67	<i>Saccharum</i>	<i>spontaneum</i>	15-95	<i>S. spontaneum</i>
45-115	<i>Saccharum</i>	<i>spontaneum</i>	US 78-526	<i>S. spontaneum</i>
41-86	<i>Saccharum</i>	<i>spontaneum</i>	P DAV 84-01	<i>S. spontaneum</i>
41-11	<i>Saccharum</i>	<i>spontaneum</i>	Holes 1	<i>S. spontaneum</i>
45-58	<i>Saccharum</i>	<i>spontaneum</i>	Taiwan spont. 072	<i>S. spontaneum</i>
37-68	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P068)	<i>S. spontaneum</i>
45-13	<i>Saccharum</i>	<i>spontaneum</i>	SES 234	<i>S. spontaneum</i>
45-22	<i>Saccharum</i>	<i>spontaneum</i>	SES 294	<i>S. spontaneum</i>
37-48	<i>Saccharum</i>	<i>spontaneum</i>	Unknown (2009: R37-P048)	<i>S. spontaneum</i>
39-48	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-180	<i>S. spontaneum</i>
41-17	<i>Saccharum</i>	<i>spontaneum</i>	IK 76-049	<i>S. spontaneum</i>

43-52	<i>Saccharum</i>	<i>spontaneum</i>	03-95	<i>S. spontaneum</i>
41-30	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-021	<i>S. spontaneum</i>
39-37	<i>Saccharum</i>	<i>spontaneum</i>	IND 81-132	<i>S. spontaneum</i>
39-1	<i>Saccharum</i>	<i>spontaneum</i>	Coimbatore, 2n=64	<i>S. spontaneum</i>
39-125	<i>Saccharum</i>	<i>spontaneum</i>	SES 289	<i>S. spontaneum</i>
41-116	<i>Saccharum</i>	<i>spontaneum</i>	S 66-088	<i>S. spontaneum</i>
41-98	<i>Saccharum</i>	<i>spontaneum</i>	PCA NOR 84-01	<i>S. spontaneum</i>
45-112	<i>Saccharum</i>	<i>spontaneum</i>	US 78-519	<i>S. spontaneum</i>
39-99	<i>Saccharum</i>	<i>spontaneum</i>	S. spont # 03	<i>S. spontaneum</i>
39-18	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-066	<i>S. spontaneum</i>
39-79	<i>Saccharum</i>	<i>spontaneum</i>	PCA NOR 84-03	<i>S. spontaneum</i>
39-164	<i>Saccharum</i>	<i>spontaneum</i>	Yacheng # 12	<i>S. spontaneum</i>
45-26	<i>Saccharum</i>	<i>spontaneum</i>	SES 341	<i>S. spontaneum</i>
45-67	<i>Saccharum</i>	<i>spontaneum</i>	UM 69-015	<i>S. spontaneum</i>
39-121	<i>Saccharum</i>	<i>spontaneum</i>	SES 208	<i>S. spontaneum</i>
39-119	<i>Saccharum</i>	<i>spontaneum</i>	SES 184 B	<i>S. spontaneum</i>
41-127	<i>Saccharum</i>	<i>spontaneum</i>	Iran spont.	<i>S. spontaneum</i>
45-24	<i>Saccharum</i>	<i>spontaneum</i>	SES 297 B	<i>S. spontaneum</i>
41-10	<i>Saccharum</i>	<i>spontaneum</i>	Gugu	<i>S. spontaneum</i>
39-19	<i>Saccharum</i>	<i>spontaneum</i>	IN 84-089	<i>S. spontaneum</i>
39-86	<i>Saccharum</i>	<i>spontaneum</i>	PCAV 84-01	<i>S. spontaneum</i>
39-139	<i>Saccharum</i>	<i>spontaneum</i>	Saudi Arabia	<i>S. spontaneum</i>
45-101	Unknown	Unknown	Unknown (2009: R45-P101)	<i>S. spontaneum</i>
37-39	Unknown	Unknown	Unknown (2009: R37-P039)	<i>S. spontaneum</i>
45-37	Unknown	Unknown	Unknown (2009: R45-P037)	<i>S. spontaneum</i>

Note: ^{a,b,c} shows records from the “world collection of sugarcane and related grasses” database.

Table S2 Associated DNA markers identified by genome-wide association study (GWAS) in sugarcane diversity panel.

Marker	Trait	Model	Run ^a	-log (P-value)
2P5849859	Brix	1-dom-ref	Full	6.43
2P65250224	Brix	1-dom-ref	Full	6.20
3P8744861	Brix	General	Spon	5.90
3P74306004	Dry weight	General	Full	7.91
4P5522303	Dry weight	General	Full	7.15
4P42706832	Dry weight	General	Full	6.29
4P42731578	Dry weight	General	Full	7.56
4P65854015	Dry weight	General	Full	7.79
6ind1921629	Dry weight	General	Full	6.05
7P6235402	Dry weight	General	Full	6.58
7P51411398	Dry weight	General	Full	6.95
9ind14840	Dry weight	General	Full	8.15
9P14902	Dry weight	General	Full	7.72
9P14943	Dry weight	General	Full	7.52
4P42706832	Dry weight	Additive	Full	7.03
4P42731578	Dry weight	Additive	Full	7.36
4P42706832	Dry weight	Diplo-general	Full	6.49
4P42731578	Dry weight	Diplo-general	Full	8.70
6P59188088	Dry weight	Diplo-general	Full	6.07
9ind14840	Dry weight	Diplo-general	Full	7.00
9P14909	Dry weight	Diplo-general	Full	6.69
9P14943	Dry weight	Diplo-general	Full	6.56
4P42706832	Dry weight	Diplo-additive	Full	7.30
4P42731578	Dry weight	Diplo-additive	Full	9.38
4P42706832	Dry weight	1-dom-ref	Full	7.13
4P42731578	Dry weight	1-dom-ref	Full	9.61
1ind8950292	Dry weight	General	Sacc	5.48
3P74306004	Dry weight	General	Sacc	7.42
4P5522303	Dry weight	General	Sacc	7.19
4P65854015	Dry weight	General	Sacc	7.24
6ind1921629	Dry weight	General	Sacc	5.58
6P2929777	Dry weight	General	Sacc	5.48
9ind14840	Dry weight	General	Sacc	7.05
9P14902	Dry weight	General	Sacc	6.73
9P14943	Dry weight	General	Sacc	6.76
10ind4226205	Dry weight	General	Sacc	5.46
10P54834195	Dry weight	General	Sacc	6.46
9ind14840	Dry weight	Diplo-general	Sacc	5.94
9P14902	Dry weight	Diplo-general	Sacc	5.94

9P14909	Dry weight	Diplo-general	Sacc	5.71
9P14943	Dry weight	Diplo-general	Sacc	6.55
1P75446476	Dry weight	General	Spon	5.50
3P63471814	Dry weight	General	Spon	6.09
2P65567954	Dry weight	General	P8	5.45
3P74306004	Dry weight	General	P8	6.96
4P64402962	Dry weight	General	P8	5.90
6P60911181	Dry weight	General	P8	5.69
7P51411398	Dry weight	Additive	P8	5.64
7P51411398	Dry weight	Diplo-general	P8	5.63
7P51411398	Dry weight	Diplo-additive	P8	6.46
7P51411398	Dry weight	1-dom-alt	P8	6.40
3P6442205	Dry weight	1-dom-ref	P8	5.31
4P6158926	Dry weight	1-dom-ref	P8	5.01
1ind10881500	Internode length	General	Full	7.48
1P27173085	Internode length	General	Full	6.30
1P57661367	Internode length	General	Full	10.01
1P59991637	Internode length	General	Full	7.95
1ind69782978	Internode length	General	Full	10.13
1P74010368	Internode length	General	Full	9.45
1P77825524	Internode length	General	Full	6.67
1P79325131	Internode length	General	Full	6.61
2P10400743	Internode length	General	Full	6.71
2P23105630	Internode length	General	Full	6.30
2P63085415	Internode length	General	Full	6.91
2P63159953	Internode length	General	Full	9.29
2P63843655	Internode length	General	Full	7.26
2P69863652	Internode length	General	Full	6.36
3P6323064	Internode length	General	Full	6.40
3P12136508	Internode length	General	Full	6.04
3P46952220	Internode length	General	Full	6.66
3P50031763	Internode length	General	Full	6.60
3P62738654	Internode length	General	Full	6.42
3P67226374	Internode length	General	Full	6.26
3P67957169	Internode length	General	Full	6.10
3P73713218	Internode length	General	Full	6.42
4P3106357	Internode length	General	Full	6.71
4P6585299	Internode length	General	Full	6.30
4P59750250	Internode length	General	Full	6.56
4P60209923	Internode length	General	Full	6.74
4P60825826	Internode length	General	Full	6.31

4ind67150492	Internode length	General	Full	11.20
5P5611035	Internode length	General	Full	7.29
5P41257174	Internode length	General	Full	6.43
6P46131193	Internode length	General	Full	6.20
6ind49957559	Internode length	General	Full	7.65
7P59358345	Internode length	General	Full	10.06
8P4222233	Internode length	General	Full	8.54
9P8930087	Internode length	General	Full	6.37
9P9107915	Internode length	General	Full	6.14
9P47153694	Internode length	General	Full	5.99
9P49501084	Internode length	General	Full	6.42
9P57288733	Internode length	General	Full	6.18
10P9053698	Internode length	General	Full	6.14
10P49024161	Internode length	General	Full	6.61
10P52292222	Internode length	General	Full	6.29
1P18164757	Internode length	Additive	Full	7.90
1P74010368	Internode length	Additive	Full	6.21
3P3400683	Internode length	Additive	Full	10.83
7P2868380	Internode length	Additive	Full	7.90
1P5406884	Internode length	Diplo-general	Full	6.38
1ind10881500	Internode length	Diplo-general	Full	7.83
1P27173085	Internode length	Diplo-general	Full	6.30
2P23105630	Internode length	Diplo-general	Full	6.30
2P49257930	Internode length	Diplo-general	Full	6.10
2P63843655	Internode length	Diplo-general	Full	6.90
2P69011550	Internode length	Diplo-general	Full	6.05
2P69863652	Internode length	Diplo-general	Full	7.00
3P4235169	Internode length	Diplo-general	Full	7.00
3P4235202	Internode length	Diplo-general	Full	6.67
3P46952220	Internode length	Diplo-general	Full	6.30
3P58486015	Internode length	Diplo-general	Full	6.54
4P3106357	Internode length	Diplo-general	Full	7.06
4P6585299	Internode length	Diplo-general	Full	6.30
4P58275825	Internode length	Diplo-general	Full	6.09
4P60825826	Internode length	Diplo-general	Full	6.31
7P9127781	Internode length	Diplo-general	Full	6.25
7P59358345	Internode length	Diplo-general	Full	10.64
9P9107915	Internode length	Diplo-general	Full	6.14
10P52292222	Internode length	Diplo-general	Full	6.50
1P5406884	Internode length	Diplo-additive	Full	7.21
1P18164757	Internode length	Diplo-additive	Full	7.90

1P27173085	Internode length	Diplo-additive	Full	7.15
1ind80028141	Internode length	Diplo-additive	Full	6.63
2P23105630	Internode length	Diplo-additive	Full	7.15
2P49257930	Internode length	Diplo-additive	Full	6.06
2P63843655	Internode length	Diplo-additive	Full	7.51
2P69863652	Internode length	Diplo-additive	Full	6.91
3P3400683	Internode length	Diplo-additive	Full	10.83
3P4235169	Internode length	Diplo-additive	Full	7.87
3P4235202	Internode length	Diplo-additive	Full	7.53
3P46952220	Internode length	Diplo-additive	Full	7.15
3P58486015	Internode length	Diplo-additive	Full	7.14
4P3106357	Internode length	Diplo-additive	Full	7.46
4P6585299	Internode length	Diplo-additive	Full	7.15
4P60825826	Internode length	Diplo-additive	Full	7.16
4ind62293349	Internode length	Diplo-additive	Full	6.51
5P53571213	Internode length	Diplo-additive	Full	6.47
7P2868380	Internode length	Diplo-additive	Full	7.90
7P9127781	Internode length	Diplo-additive	Full	6.55
7P59358345	Internode length	Diplo-additive	Full	11.53
9P9107915	Internode length	Diplo-additive	Full	6.14
10P52292222	Internode length	Diplo-additive	Full	7.12
1ind5376369	Internode length	1-dom-alt	Full	6.07
1P18164757	Internode length	1-dom-alt	Full	7.90
1P27173085	Internode length	1-dom-alt	Full	7.05
1ind80028141	Internode length	1-dom-alt	Full	6.49
2P23105630	Internode length	1-dom-alt	Full	7.05
2P54560011	Internode length	1-dom-alt	Full	6.46
2P63843655	Internode length	1-dom-alt	Full	7.51
2P67022177	Internode length	1-dom-alt	Full	6.52
2P69011550	Internode length	1-dom-alt	Full	6.75
3P3400683	Internode length	1-dom-alt	Full	10.83
3P4233123	Internode length	1-dom-alt	Full	6.29
3P4235169	Internode length	1-dom-alt	Full	7.86
3P4235202	Internode length	1-dom-alt	Full	7.52
3P58486015	Internode length	1-dom-alt	Full	7.38
4P3106357	Internode length	1-dom-alt	Full	7.64
4P6585299	Internode length	1-dom-alt	Full	7.05
4P58275825	Internode length	1-dom-alt	Full	6.79
4P60825826	Internode length	1-dom-alt	Full	7.09
4ind62293349	Internode length	1-dom-alt	Full	6.50
7P2868380	Internode length	1-dom-alt	Full	7.90

7P9127781	Internode length	1-dom-alt	Full	7.00
7P59358345	Internode length	1-dom-alt	Full	11.49
9P9107915	Internode length	1-dom-alt	Full	6.95
10P49024161	Internode length	1-dom-alt	Full	6.59
1P5406884	Internode length	1-dom-ref	Full	7.05
1P9820249	Internode length	1-dom-ref	Full	6.36
2P49257930	Internode length	1-dom-ref	Full	6.92
2P69863652	Internode length	1-dom-ref	Full	7.82
3P6323064	Internode length	1-dom-ref	Full	6.62
3P46952220	Internode length	1-dom-ref	Full	7.05
5P53571213	Internode length	1-dom-ref	Full	6.42
6P46131193	Internode length	1-dom-ref	Full	6.77
10P52292222	Internode length	1-dom-ref	Full	7.09
1P57661367	Internode length	General	Sacc	8.57
1P59991637	Internode length	General	Sacc	6.89
1ind69782978	Internode length	General	Sacc	8.82
1P77825524	Internode length	General	Sacc	5.70
1P79325131	Internode length	General	Sacc	6.13
2P5443742	Internode length	General	Sacc	5.90
2P10400743	Internode length	General	Sacc	6.17
3P12136508	Internode length	General	Sacc	5.82
3P50031763	Internode length	General	Sacc	5.90
3P73713218	Internode length	General	Sacc	5.84
4P5670529	Internode length	General	Sacc	6.03
4ind67150492	Internode length	General	Sacc	9.83
8P4222225	Internode length	General	Sacc	8.57
8P4222233	Internode length	General	Sacc	9.83
8ind7147109	Internode length	General	Sacc	5.38
9P54269438	Internode length	General	Sacc	5.36
1P59991637	Internode length	General	Spon	7.25
2P72986213	Internode length	General	Spon	6.93
3P1274812	Internode length	General	Spon	5.40
3P4650393	Internode length	General	Spon	5.51
4P4992036	Internode length	General	Spon	5.60
4ind67150479	Internode length	General	Spon	5.48
4ind67150492	Internode length	General	Spon	5.88
6P10569626	Internode length	General	Spon	5.43
7P5526692	Internode length	General	Spon	6.69
8P4222233	Internode length	General	Spon	6.34
10P57560591	Internode length	General	Spon	5.64
1P12592602	Internode length	Additive	Spon	5.42

6P42692072	Leaf length	1-dom-ref	P8	5.19
1P63650261	Leaf width	General	Full	6.77
1P66407411	Leaf width	General	Full	6.29
2P8062355	Leaf width	General	Full	6.10
2P8663239	Leaf width	General	Full	10.58
2P45971305	Leaf width	General	Full	6.98
2ind65539620	Leaf width	General	Full	6.67
2P66158603	Leaf width	General	Full	6.45
2P75203341	Leaf width	General	Full	6.30
3P917667	Leaf width	General	Full	6.64
3P53284899	Leaf width	General	Full	10.16
3P56030744	Leaf width	General	Full	8.32
3P57179114	Leaf width	General	Full	6.41
3P58845520	Leaf width	General	Full	6.90
3P67095919	Leaf width	General	Full	6.12
4P2535140	Leaf width	General	Full	8.15
4P16089454	Leaf width	General	Full	7.92
4P53646791	Leaf width	General	Full	6.96
5ind60876916	Leaf width	General	Full	8.95
6P48326143	Leaf width	General	Full	6.58
6P53031390	Leaf width	General	Full	6.76
7P54423302	Leaf width	General	Full	6.33
8P62190177	Leaf width	General	Full	6.24
9P3981443	Leaf width	General	Full	7.08
9P3981469	Leaf width	General	Full	6.91
9P3981476	Leaf width	General	Full	6.94
9P56695054	Leaf width	General	Full	8.60
9P58067302	Leaf width	General	Full	6.20
9P58067306	Leaf width	General	Full	6.88
10P952532	Leaf width	General	Full	6.54
10P952546	Leaf width	General	Full	6.73
10ind56691142	Leaf width	General	Full	6.59
3P60794079	Leaf width	Additive	Full	7.24
1P55069349	Leaf width	Diplo-general	Full	6.05
3P68383201	Leaf width	Diplo-general	Full	6.13
4P57110289	Leaf width	Diplo-general	Full	6.09
10P7940593	Leaf width	Diplo-general	Full	6.25
10P55668582	Leaf width	Diplo-general	Full	6.43
10P55668582	Leaf width	1-dom-ref	Full	7.12
1P66407411	Leaf width	General	Sacc	5.89
1P71934598	Leaf width	General	Sacc	6.28

2P8663239	Leaf width	General	Sacc	9.87
2ind12470564	Leaf width	General	Sacc	6.01
2ind65539620	Leaf width	General	Sacc	5.81
2P66158603	Leaf width	General	Sacc	6.15
2P75203341	Leaf width	General	Sacc	5.38
3P917667	Leaf width	General	Sacc	5.81
3P53284899	Leaf width	General	Sacc	8.86
3P56030744	Leaf width	General	Sacc	7.28
3P57179114	Leaf width	General	Sacc	6.41
3P58845520	Leaf width	General	Sacc	6.41
3P67095919	Leaf width	General	Sacc	5.62
4P2535140	Leaf width	General	Sacc	6.37
4P16089454	Leaf width	General	Sacc	7.09
4P51358883	Leaf width	General	Sacc	6.49
4P53646791	Leaf width	General	Sacc	6.51
5ind60876916	Leaf width	General	Sacc	6.55
6P2803959	Leaf width	General	Sacc	6.40
6P48326143	Leaf width	General	Sacc	5.74
6P53031390	Leaf width	General	Sacc	5.49
7ind6235143	Leaf width	General	Sacc	6.08
7P61801701	Leaf width	General	Sacc	5.38
8P2787878	Leaf width	General	Sacc	5.36
8P62190177	Leaf width	General	Sacc	5.85
9P3981443	Leaf width	General	Sacc	6.21
9P3981469	Leaf width	General	Sacc	6.25
9P3981476	Leaf width	General	Sacc	6.31
9P3986537	Leaf width	General	Sacc	5.88
9P56695054	Leaf width	General	Sacc	7.40
9P58067302	Leaf width	General	Sacc	6.18
9P58067306	Leaf width	General	Sacc	6.48
10P952546	Leaf width	General	Sacc	6.11
10ind2865773	Leaf width	General	Sacc	5.42
10ind56691142	Leaf width	General	Sacc	6.04
10P7940593	Leaf width	Diplo-general	Sacc	5.33
10P55668582	Leaf width	1-dom-ref	Sacc	5.59
1P70492806	Leaf width	General	Spon	6.94
2P54564794	Leaf width	General	Spon	7.76
2P59550746	Leaf width	General	Spon	8.02
3P31596498	Leaf width	General	Spon	5.86
3P55473119	Leaf width	General	Spon	5.91
3ind55473132	Leaf width	General	Spon	6.72

4P12574117	Leaf width	General	Spon	5.40
4P22574534	Leaf width	General	Spon	8.18
7P787036	Leaf width	General	Spon	6.48
7P64752471	Leaf width	General	Spon	5.96
9P180773	Leaf width	General	Spon	8.19
9P2630254	Leaf width	General	Spon	6.86
9P6750858	Leaf width	General	Spon	5.61
1P64954222	Leaf width	General	Hybrid	6.62
2P8663239	Leaf width	General	Hybrid	7.45
2ind12470564	Leaf width	General	Hybrid	5.53
2P18690374	Leaf width	General	Hybrid	5.71
3P5244300	Leaf width	General	Hybrid	5.64
3P18402714	Leaf width	General	Hybrid	5.42
3P49105822	Leaf width	General	Hybrid	5.48
3P53284899	Leaf width	General	Hybrid	6.30
5P11920928	Leaf width	General	Hybrid	5.38
7P8264992	Leaf width	General	Hybrid	5.77
9P56695054	Leaf width	General	Hybrid	5.53
9P58067306	Leaf width	General	Hybrid	5.60
1P9326642	Leaf width	General	P8	5.99
1P64954222	Leaf width	General	P8	6.85
1P75002839	Leaf width	General	P8	5.35
2P8062355	Leaf width	General	P8	5.33
2P8663239	Leaf width	General	P8	7.96
2ind12470564	Leaf width	General	P8	8.08
2P45971305	Leaf width	General	P8	6.13
2P59201187	Leaf width	General	P8	5.87
2ind65916673	Leaf width	General	P8	5.61
2P66946732	Leaf width	General	P8	5.77
3ind5418130	Leaf width	General	P8	5.48
3P53284899	Leaf width	General	P8	7.55
3P56030744	Leaf width	General	P8	5.92
3P58845520	Leaf width	General	P8	7.61
3P66737730	Leaf width	General	P8	6.12
4P2535140	Leaf width	General	P8	6.37
4P16089454	Leaf width	General	P8	7.14
4P51358883	Leaf width	General	P8	5.77
4P58413532	Leaf width	General	P8	5.35
5P9496728	Leaf width	General	P8	5.39
5ind60876916	Leaf width	General	P8	7.55
6P19319761	Leaf width	General	P8	6.15

6P19319765	Leaf width	General	P8	5.65
6P48326143	Leaf width	General	P8	5.73
9P3981443	Leaf width	General	P8	5.89
9P3981469	Leaf width	General	P8	6.11
9P3981476	Leaf width	General	P8	6.16
9P3986537	Leaf width	General	P8	5.35
9P7561214	Leaf width	General	P8	5.77
9P56695054	Leaf width	General	P8	7.62
9P58067306	Leaf width	General	P8	5.28
10P55831327	Leaf width	General	P8	5.36
1P63650261	Leaf width	Additive	P8	5.74
1P79853549	Leaf width	Additive	P8	6.76
3P6606911	Leaf width	Diplo-general	P8	5.37
3P66737730	Leaf width	Diplo-general	P8	8.49
3P66744355	Leaf width	Diplo-general	P8	5.42
3P66942928	Leaf width	Diplo-general	P8	5.34
1P79853549	Leaf width	Diplo-additive	P8	5.76
3P66737730	Leaf width	Diplo-additive	P8	8.49
3P66744355	Leaf width	Diplo-additive	P8	6.03
3P6606911	Leaf width	1-dom-alt	P8	5.37
3P66737730	Leaf width	1-dom-alt	P8	8.49
3P66744355	Leaf width	1-dom-alt	P8	6.22
3P66942928	Leaf width	1-dom-alt	P8	5.34
1P79853549	Leaf width	1-dom-ref	P8	5.76
4P57110289	Leaf width	1-dom-ref	P8	5.83
4P57110290	Leaf width	1-dom-ref	P8	5.27
1ind17427442	Stalk diameter	General	Full	9.96
1P65993901	Stalk diameter	General	Full	7.75
1P70496492	Stalk diameter	General	Full	7.78
2P8663239	Stalk diameter	General	Full	6.15
2P16274208	Stalk diameter	General	Full	7.14
3ind11618215	Stalk diameter	General	Full	6.73
3P50031909	Stalk diameter	General	Full	6.38
4P2535140	Stalk diameter	General	Full	7.70
5ind60876916	Stalk diameter	General	Full	7.75
6ind46223964	Stalk diameter	General	Full	6.72
8P5859661	Stalk diameter	General	Full	6.47
9P46359323	Stalk diameter	General	Full	6.04
9P58076687	Stalk diameter	General	Full	6.10
10P57767995	Stalk diameter	General	Full	6.07
1ind17427442	Stalk diameter	Additive	Full	9.45

9P58076687	Stalk diameter	Additive	Full	6.39
1P70512341	Stalk diameter	Diplo-general	Full	6.22
3P65407586	Stalk diameter	Diplo-general	Full	8.97
9P46359323	Stalk diameter	Diplo-general	Full	9.15
10P55668582	Stalk diameter	Diplo-general	Full	6.73
1P70512341	Stalk diameter	Diplo-additive	Full	6.22
4ind2492462	Stalk diameter	Diplo-additive	Full	6.40
1P70512341	Stalk diameter	1-dom-alt	Full	6.22
1ind17427442	Stalk diameter	General	Sacc	7.96
1P65993901	Stalk diameter	General	Sacc	5.93
1P70496492	Stalk diameter	General	Sacc	5.91
2P16274208	Stalk diameter	General	Sacc	5.57
4P2535140	Stalk diameter	General	Sacc	6.82
5ind60876916	Stalk diameter	General	Sacc	6.16
7P64921958	Stalk diameter	General	Sacc	5.49
1ind17427442	Stalk diameter	Additive	Sacc	7.83
6P54543379	Stalk diameter	Additive	Sacc	5.81
3P60871782	Stalk diameter	Diplo-general	Sacc	5.99
3P65407586	Stalk diameter	Diplo-general	Sacc	7.37
9P46359323	Stalk diameter	Diplo-general	Sacc	7.58
10P55668582	Stalk diameter	Diplo-general	Sacc	6.20
3P60871782	Stalk diameter	Diplo-additive	Sacc	5.99
3P60871782	Stalk diameter	1-dom-alt	Sacc	5.99
10P55668582	Stalk diameter	1-dom-ref	Sacc	5.24
9P58085131	Stalk diameter	General	Hybrid	6.49
1ind17427442	Stalk diameter	General	P8	5.84
1P66407437	Stalk diameter	General	P8	5.50
2P8663239	Stalk diameter	General	P8	6.15
4P2535140	Stalk diameter	General	P8	5.63
4P2535154	Stalk diameter	General	P8	5.28
1ind17427442	Stalk diameter	Additive	P8	5.82
1P66845769	Stalk diameter	Additive	P8	5.61
3P65407586	Stalk diameter	Diplo-general	P8	6.03
3P66737730	Stalk diameter	Diplo-general	P8	6.05
9P46359323	Stalk diameter	Diplo-general	P8	6.03
3P66737730	Stalk diameter	Diplo-additive	P8	6.05
3P66744355	Stalk diameter	Diplo-additive	P8	5.52
3P66737730	Stalk diameter	1-dom-alt	P8	6.05
3P66744355	Stalk diameter	1-dom-alt	P8	5.48
1P60062022	Stalk diameter	1-dom-ref	P8	5.06
2P65627224	Stalk number	General	Full	6.03

2P65627224	Stalk number	General	Sacc	5.45
10P60783572	Stalk number	General	Sacc	6.07
4P65103065	Stalk number	General	Spon	6.47
10P60783572	Stalk number	General	Spon	6.00
3P66955169	Stalk number	General	P8	5.86
3P63471814	Total weight	General	Spon	6.42
7P2769299	Total weight	General	P8	6.04
3P61470337	Total weight	Diplo-general	P8	6.06
3P61470337	Total weight	Diplo-additive	P8	6.06
3P61470337	Total weight	1-dom-alt	P8	6.06
6P37710980	Water content	Diplo-general	Sacc	5.81
6P37710980	Water content	Diplo-additive	Sacc	5.81
6P37710980	Water content	1-dom-alt	Sacc	5.81
3P63471814	Water content	General	Spon	5.53
1P3184652	Water content	General	P8	5.65
7P2769299	Water content	General	P8	5.97
1P3184652	Water content	Additive	P8	6.17
3P61470337	Water content	Diplo-general	P8	5.35
3P61470337	Water content	1-dom-alt	P8	5.35

Note: ^a GWAS runs in Full set and several subests. Full = all accessions in the diversity panel; Sacc = 273 accessions from *S. spontaneum*, *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; Spon = 100 accessions from *S. spontaneum*; Hybrid = 173 accessions from *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; and P8 = 162 accessions with the same ploidy level (ploidy =8).

817 **Table S3** Non-redundant associated markers and candidate genes with genome-wide associations study (GWAS)
818 hits.

Marker	Trait	GWAS runs* and models	REF	ALT	Mutation type	Effect	Sorghum gene	Annotation	Domains
2P5849859	Brix	Full (1-dom-ref)	A	G	3_prime_UTR_variant	MODIFIER	<i>Sobic.002G060900</i>	similar to Transducin / WD-40 repeat protein-like	NA
2P6525024	Brix	Full (1-dom-ref)	T	G	intron_variant	MODIFIER	<i>Sobic.002G268400</i>	similar to Putative uncharacterized protein	NA
3P8744861	Brix	Spon (General)	A	C	splice_region_variant&intron_variant	LOW	<i>Sobic.003G099300</i>	similar to Putative uncharacterized protein P0439B06.21	FDF domain, Sed6-like Sm domain
10P5776795	Stalk diameter	Full (General)	C	G	missense_variant	MODERATE	<i>Sobic.010G235200</i>	similar to Putative uncharacterized protein	Essential protein Yae1, N terminal
1ind17427442	Stalk diameter	Full (General), Full (Additive), Sacc (General), Sacc (Additive), P8 (General), P8 (Additive)	CT	C	intron_variant	MODIFIER	<i>Sobic.001G194600</i>	similar to Putative uncharacterized protein	Myb/SANT-like DNA-binding domain
1P60062022	Stalk diameter	P8 (1-dom-ref)	G	A	synonymous_variant	LOW	<i>Sobic.001G313400</i>	PTHR10887//PTHR10887:S F353 - DNA2/NAM7 HELICASE FAMILY // SUBFAMILY NOT NAMED (1 of 1)	AAA domain
1P65993901	Stalk diameter	Full (General), Sacc (General)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G371700</i>	similar to Hydroxyproline-rich glycoprotein-like	NA
1P66407437	Stalk diameter	P8 (General)	G	A	intron_variant	MODIFIER	<i>Sobic.001G375700</i>	similar to ASF/SF2-like pre-mRNA splicing factor SRP32	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
1P66845769	Stalk diameter	P8 (Additive)	A	G	missense_variant	MODERATE	<i>Sobic.001G380400</i>	similar to Reticulon family protein, putative, expressed	Reticulon
1P70496492	Stalk diameter	Full (General), Sacc (General)	C	T	missense_variant	MODERATE	<i>Sobic.001G425000</i>	similar to Serine/threonine-protein phosphatase PP1	Calcineurin-like phosphoesterase
1P70512341	Stalk diameter	Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	C	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G425150</i>	KOG0752 - Mitochondrial solute carrier protein (1 of 15)	Mitochondrial carrier protein
2P16274208	Stalk diameter	Full (General), Sacc (General)	C	A	5_prime_UTR_variant	MODIFIER	<i>Sobic.002G122500</i>	similar to Putative uncharacterized protein P0030H06.108	P21-Rho-binding domain
2P8663239	Stalk diameter/ Leaf width	Full (General), Sacc (General), Hybrid (General), P8 (General)/ Full (General), P8 (General)	G	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.002G081600</i>	similar to 40S ribosomal protein S12	Ribosomal protein L7Ae/L30e/S12e/ Gadd45 family
3ind11618215	Stalk diameter	Full (General)	G	GA	intron_variant	MODIFIER	<i>Sobic.003G126800</i>	similar to Putative cystatin	Aspartic acid proteinase inhibitor
3P50031909	Stalk diameter	Full (General)	C	T	intron_variant	MODIFIER	<i>Sobic.003G185500</i>	similar to Putative uncharacterized protein	NA
3P60871782	Stalk diameter	Sacc (1-dom-alt), Sacc (Diplo-general), Sacc (Diplo-additive)	G	A	intron_variant	MODIFIER	<i>Sobic.003G272400</i>	PTHR12161:SF13 - REGULATOR OF VPS4 ACTIVITY PROTEIN-RELATED (1 of 1)	Regulator of Vps4 activity in the MVB pathway
3P65407586	Stalk diameter	Full (Diplo-general), Sacc (Diplo-general)	C	T	synonymous_variant	LOW	<i>Sobic.003G328700</i>	similar to Putative DNA-directed RNA polymerase II 23K chain	RNA polymerase Rpb5, C-terminal domain, RNA polymerase Rpb5, N-terminal domain

		P8 (Diplo-general)							
4ind2492462	Stalk diameter	Full (Diplo-additive)	CT	C	intergenic_region	MODIFIER	<i>Sobic.004G031001-Sobic.004G031100</i>	NA/similar to Os02g0132500 protein	NA/bZIP transcription factor
4P2535154	Stalk diameter	P8 (General)	G	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.004G031400</i>	PTHR27006:SF36 - LRR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASE RKF3-RELATED (1 of 1)	Protein kinase domain
6ind46223964	Stalk diameter	Full (General)	TGA CC	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.006G092100</i>	similar to Putative uncharacterized protein	No apical meristem (NAM) protein
6P54543379	Stalk diameter	Sacc (Additive)	A	C	stop_gained	HIGH	<i>Sobic.006G190950</i>	PTHR10887//PTHR10887:S F373 - DNA2/NAM7 HELICASE FAMILY // SUBFAMILY NOT NAMED (1 of 6)	AAA domain
7P64921958	Stalk diameter	Sacc (General)	C	T	intron_variant	MODIFIER	<i>Sobic.007G220900</i>	similar to Putative uncharacterized protein	Ribosomal protein L9, N-terminal domain
8P5859661	Stalk diameter	Full (General)	A	G	intergenic_region	MODIFIER	<i>Sobic.008G056200-Sobic.008G056350</i>	PTHR32133:SF82 - F-BOX ASSOCIATED UBIQUITINATION EFFECTOR FAMILY PROTEIN-RELATED (1 of 5)/PTHR32133:SF82 - F-BOX ASSOCIATED UBIQUITINATION EFFECTOR FAMILY PROTEIN-RELATED (1 of 5)	F-box domain/F-box domain
9P46359323	Stalk diameter	Full (General), Full (Diplo-general), Sacc (Diplo-general), P8 (Diplo-general)	C	T	intron_variant	MODIFIER	<i>Sobic.009G118500</i>	similar to Os05g0373500 protein	NA
9P58076687	Stalk diameter	Full (General), Full (Additive)	A	T	intron_variant	MODIFIER	<i>Sobic.009G245000</i>	PTHR22572:SF110 - GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE LARGE SUBUNIT 2, CHLOROPLASTIC (1 of 2)	Nucleotidyl transferase
9P58085131	Stalk diameter	Hybrid (General)	G	A	intron_variant	MODIFIER	<i>Sobic.009G245101</i>	NA	NA
10P55668582	Stalk diameter/ Leaf width	Full (Diplo-general), Sacc (1-dom-ref), Sacc (Diplo-general)/Full (1-dom-ref), Full (Diplo-general), Sacc (1-dom-ref)	T	C	intergenic_region	MODIFIER	<i>Sobic.010G213600-Sobic.010G213700</i>	similar to Os06g0654300 protein/similar to Putative uncharacterized protein P0709F06.15	His Kinase A (phospho-acceptor) domain, Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, Response regulator receiver domain/NA
3P66737730	Stalk diameter/ Leaf width	P8 (1-dom-alt), P8 (Diplo-general), P8 (Diplo-additive)/P8 (General), P8 (1-dom-alt), P8 (Diplo-general), P8 (Diplo-additive)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G346600</i>	similar to 40S ribosomal protein S23-1	Ribosomal protein S12/S23
3P66744355	Stalk diameter/ Leaf width	P8 (1-dom-alt), P8 (Diplo-additive)/P8 (1-dom-alt), P8 (Diplo-general), P8 (Diplo-additive)	G	A	intron_variant	MODIFIER	<i>Sobic.003G346700</i>	similar to Os01g0834700 protein	DRG Family Regulatory Proteins, Tma46
4P2535140	Stalk diameter/	Full (General)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.004G031400</i>	PTHR27006:SF36 - LRR RECEPTOR-LIKE	Protein kinase domain

	Leaf width	Sacc (General), P8 (General)/ Full (General), Sacc (General), P8 (General)						SERINE/THREONINE-PROTEIN KINASE RKF3-RELATED (1 of 1)	
5ind60876916	Stalk diameter/ Leaf width	Full (General), Sacc (General), P8 (General)/ Full (General), Sacc (General), P8 (General)	A	AT	intron_variant	MODIFIER	<i>Sobic.005G144201</i>	PTHR21229:SF28 - LUNG SEVEN TRANSMEMBRANE RECEPTOR-LIKE PROTEIN (1 of 1)	Lung seven transmembrane receptor
10ind4226205	Dry weight	Sacc (General)	TTTC	T	intron_variant	MODIFIER	<i>Sobic.010G053800</i>	K12391 - AP-1 complex subunit gamma-1 (AP1G1) (1 of 2)	Adaptin C-terminal domain, Adaptin N terminal region
10P54834195	Dry weight	Sacc (General)	C	T	missense_variant	MODERATE	<i>Sobic.010G205500</i>	similar to Vacuolar H ⁺ -pyrophosphatase	Inorganic H ⁺ -pyrophosphatase
1ind8950292	Dry weight	Sacc (General)	TGG A	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G114200</i>	PTHR13317:SF4 - TRANSMEMBRANE ANTERIOR POSTERIOR TRANSFORMATION PROTEIN 1 HOMOLOG (1 of 2)	Eukaryotic membrane protein family
1P75446476	Dry weight	Spon (General)	A	G	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G483200</i>	PF07145 - Ataxin-2 C-terminal region (PAM2) (1 of 11)	NA
2P65567954	Dry weight	P8 (General)	A	T	intron_variant	MODIFIER	<i>Sobic.002G272600</i>	similar to Putative GTPase activating protein	Rab-GTPase-TBC domain
3P74306004	Dry weight	Full (General), Sacc (General), P8 (General)	A	G	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G445000</i>	similar to Putative pectinacetyltransferase	Pectinacetyltransferase
4P42706832	Dry weight	Full (General), Full (Additive), Full (1-dom-ref), Full (Diplo-general), Full (Diplo-additive)	C	T	intron_variant	MODIFIER	<i>Sobic.004G143475</i>	PTHR16166:SF93 - PROTEIN T08G11.1, ISOFORM A (1 of 1)	Vacuolar protein sorting-associated protein
4P42731578	Dry weight	Full (General), Full (Additive), Full (1-dom-ref), Full (Diplo-general), Full (Diplo-additive)	G	A	intron_variant	MODIFIER	<i>Sobic.004G143475</i>	PTHR16166:SF93 - PROTEIN T08G11.1, ISOFORM A (1 of 1)	Vacuolar protein sorting-associated protein
4P5522303	Dry weight	Full (General), Sacc (General)	A	T	intergenic_region	MODIFIER	<i>Sobic.004G067700- Sobic.004G067800</i>	similar to Chromosome undetermined scaffold_79, whole genome shotgun sequence/similar to Putative zinc-binding protein	BRCA1-associated protein 2, Ring finger domain, Zn-finger in ubiquitin-hydrolases and other protein/PLATZ transcription factor
4P6158926	Dry weight	P8 (1-dom-ref)	A	T	splice_region_variant&intron_variant	LOW	<i>Sobic.004G075700</i>	similar to Putative uncharacterized protein	PHD-finger, WSTF, HB1, Ict1p, MBD9 motif 1
4P64402962	Dry weight	P8 (General)	A	G	synonymous_variant	LOW	<i>Sobic.004G305600</i>	similar to Chromosome chr11 scaffold_13, whole genome shotgun sequence	SKIP/SNW domain
4P65854015	Dry weight	Full (General), Sacc (General)	A	T	missense_variant	MODERATE	<i>Sobic.004G323900</i>	K09553 - stress-induced-phosphoprotein 1 (STIP1) (1 of 3)	NA
6ind1921629	Dry weight	Full (General),	GAA	G	intergenic_region	MODIFIER	<i>Sobic.006G013500-</i>	similar to Cytidine/deoxycytidylate	NA/NA

		Sacc (General)					<i>Sobic.006G013600</i>	deaminase-like protein/NA	
6P2929777	Dry weight	Sacc (General)	T	A	missense_variant	MODERATE	<i>Sobic.006G017800</i>	similar to OSIGBa0136008-OSIGBa0153H12.2 protein	NA
6P59188088	Dry weight	Full (Diplo-general)	T	A	synonymous_variant	LOW	<i>Sobic.006G254300</i>	similar to Os04g0662800 protein	Regulator of chromosome condensation (RCC1) repeat
6P60911181	Dry weight	P8 (General)	G	T	intron_variant	MODIFIER	<i>Sobic.006G278500</i>	similar to Putative uncharacterized protein	Glutamine amidotransferase class-I
7P51411398	Dry weight	Full (General), P8 (Additive), P8 (1-dom-alt), P8 (Diplo-general), P8 (Diplo-additive)	T	C	intron_variant	MODIFIER	<i>Sobic.007G119500</i>	PF05664 - Plant family of unknown function (DUF810) (DUF810) (1 of 7)	Plant family of unknown function (DUF810)
7P6235402	Dry weight	Full (General)	T	C	intron_variant	MODIFIER	<i>Sobic.007G059600</i>	similar to Benzothiadiazole-induced somatic embryogenesis receptor kinase 1	Leucine rich repeat N-terminal domain, Protein kinase domain
9ind14840	Dry weight	Full (General), Full (Diplo-general), Sacc (General), Sacc (Diplo-general)	G	GTTA	intron_variant	MODIFIER	<i>Sobic.009G000100</i>	similar to Putative uncharacterized protein	C-terminal domain of Sin3a protein, Paired amphipathic helix repeat, Sin3 family co-repressor
9P14902	Dry weight	Full (General), Sacc (General), Sacc (Diplo-general)	C	G	missense_variant	MODERATE	<i>Sobic.009G000100</i>	similar to Putative uncharacterized protein	C-terminal domain of Sin3a protein, Paired amphipathic helix repeat, Sin3 family co-repressor
9P14909	Dry weight	Full (Diplo-general), Sacc (Diplo-general)	T	C	synonymous_variant	LOW	<i>Sobic.009G000100</i>	similar to Putative uncharacterized protein	C-terminal domain of Sin3a protein, Paired amphipathic helix repeat, Sin3 family co-repressor
9P14943	Dry weight	Full (General), Full (Diplo-general), Sacc (General), Sacc (Diplo-general)	G	A	missense_variant	MODERATE	<i>Sobic.009G000100</i>	similar to Putative uncharacterized protein	C-terminal domain of Sin3a protein, Paired amphipathic helix repeat, Sin3 family co-repressor
3P6442205	Dry weight	P8 (1-dom-ref)	C	T	intergenic_region	MODIFIER	<i>Sobic.003G075632- Sobic.003G075700</i>	NA/NA	NA/NA
3P63471814	Dry weight/Water content/Total weight	Spon (General)/Spon (General)/Spon (General)	G	A	synonymous_variant	LOW	<i>Sobic.003G304400</i>	similar to Os05g0513200 protein	NLI interacting factor-like phosphatase
10P49024161	Internode length	Full (General), Full (1-dom-alt)	T	C	intron_variant	MODIFIER	<i>Sobic.010G166000</i>	similar to Os05g0513300 protein	Hinge domain of cleavage stimulation factor subunit 2, Transcription termination and cleavage factor C-terminal
10P52292222	Internode length	Full (General), Full (1-dom-ref), Full (Diplo-general), Full (Diplo-additive)	A	T	missense_variant	MODERATE	<i>Sobic.010G183100</i>	similar to Agenet domain-containing protein-like	Agenet domain, ENT domain
10P57560591	Internode length	Spon (General)	C	A	synonymous_variant	LOW	<i>Sobic.010G232600</i>	PTHR23051//PTHR23051:S F1 - SOLUTE CARRIER FAMILY 35, MEMBER F5 // SUBFAMILY NOT	Solute carrier family 35

								NAMED (1 of 1)	
10P9053698	Internode length	Full (General)	G	T	intron_variant	MODIFIER	<i>Sobic.010G099900</i>	similar to Os06g0239700 protein	NA
1ind10881500	Internode length	Full (General), Full (Diplo-general)	T	TGT A	5_prime_UTR_variant	MODIFIER	<i>Sobic.001G137200</i>	similar to Knotted1-interacting protein	Homeobox KN domain
1ind5376369	Internode length	Full (1-dom-alt)	TA	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G070500</i>	similar to Photosystem I reaction center subunit III, chloroplast precursor	Photosystem I reaction centre subunit III
1ind69782978	Internode length	Full (General), Sacc (General)	TC	T	intron_variant	MODIFIER	<i>Sobic.001G416300</i>	similar to Expressed protein	NA
1ind80028141	Internode length	Full (1-dom-alt), Full (Diplo-additive)	TGT TA	T	intron_variant	MODIFIER	<i>Sobic.001G537300</i>	similar to Rolled leaf 2	Homeobox domain, MEKHLA domain, START domain
1P12592602	Internode length	Spon (Additive)	T	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G155500</i>	similar to Pyruvate dehydrogenase E1 component beta subunit, putative, expressed	Transketolase, C-terminal domain, pyrimidine binding domain
1P18164757	Internode length	Full (Additive), Full (1-dom-alt), Full (Diplo-additive)	A	C	missense_variant	MODERATE	<i>Sobic.001G200500</i>	similar to Heavy-metal-associated domain-containing protein, putative, expressed	Heavy-metal-associated domain
1P27173085	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	G	A	synonymous_variant	LOW	<i>Sobic.001G251600</i>	similar to SET domain protein 123	SET domain
1P5406884	Internode length	Full (1-dom-ref), Full (Diplo-general), Full (Diplo-additive)	A	G	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G070900</i>	weakly similar to Putative uncharacterized protein	NA
1P57661367	Internode length	Full (General), Sacc (General)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.001G296700</i>	similar to Probable histone H2AXb	Core histone H2A/H2B/H3/H4
1P59991637	Internode length	Full (General), Sacc (General), Spon (General)	G	A	synonymous_variant	LOW	<i>Sobic.001G312900</i>	similar to Hexose transporter	Sugar (and other) transporter
1P74010368	Internode length	Full (General), Full (Additive)	A	G	intron_variant	MODIFIER	<i>Sobic.001G466900</i>	similar to Os12g0406100 protein	Glycosyl hydrolases family 43
1P77825524	Internode length	Full (General), Sacc (General)	G	A	missense_variant	MODERATE	<i>Sobic.001G510500</i>	similar to CCT motif family protein, expressed	CCT motif
1P79325131	Internode length	Full (General), Sacc (General)	C	T	synonymous_variant	LOW	<i>Sobic.001G529600</i>	PTHR12741:SF12 - CALLOSE SYNTHASE 9 (1 of 1)	1,3-beta-glucan synthase component
1P9820249	Internode length	Full (1-dom-ref)	T	C	intergenic_region	MODIFIER	<i>Sobic.001G124600- Sobic.001G124800</i>	NA/NA	NA/NA
2P10400743	Internode length	Full (General), Sacc (General)	A	G	synonymous_variant	LOW	<i>Sobic.002G097900</i>	weakly similar to Putative uncharacterized protein	Domain of unknown function (DUF1719)
2P23105630	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	A	T	intron_variant	MODIFIER	<i>Sobic.002G141900</i>	similar to Similarity to Na+/H+ antiporter	Citrate transporter
2P49257930	Internode length	Full (1-dom-ref),	G	A	missense_variant	MODERATE	<i>Sobic.002G159200</i>	K17985 - activating molecule in BECN1-	NA

		Full (Diplo-general), Full (Diplo-additive)						regulated autophagy protein 1 (AMBRA1) (1 of 2)	
2P5443742	Internode length	Sacc (General)	G	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.002G056200</i>	similar to 40S ribosomal protein S15	Ribosomal protein S19
2P54560011	Internode length	Full (1-dom-alt)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.002G172401</i>	similar to GTP-binding protein, ras-like	Ras family
2P63085415	Internode length	Full (General)	T	C	intron_variant	MODIFIER	<i>Sobic.002G241100</i>	PF01501 - Glycosyl transferase family 8 (Glyco_transf_8) (1 of 34)	Glycosyl transferase family 8
2P63159953	Internode length	Full (General)	G	A	intron_variant	MODIFIER	<i>Sobic.002G242500</i>	similar to Os09g0481700 protein	WRKY DNA-binding domain
2P63843655	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	G	A	intron_variant	MODIFIER	<i>Sobic.002G251800</i>	PF01535//PF13041//PF13812 - PPR repeat (PPR) // PPR repeat family (PPR_2) // Pentatricopeptide repeat domain (PPR_3) (1 of 59)	PPR repeat family
2P67022177	Internode length	Full (1-dom-alt)	C	T	missense_variant	MODERATE	<i>Sobic.002G292300</i>	similar to Os09g0555400 protein	PPR repeat family
2P69011550	Internode length	Full (1-dom-alt), Full (Diplo-general)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.002G317400</i>	similar to Putative nuclear protein	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
2P69863652	Internode length	Full (General), Full (1-dom-ref), Full (Diplo-general), Full (Diplo-additive)	A	T	intron_variant	MODIFIER	<i>Sobic.002G329700</i>	similar to Putative uncharacterized protein	Ribosomal L28 family
2P72986213	Internode length	Spon (General)	T	C	synonymous_variant	LOW	<i>Sobic.002G371000</i>	similar to Putative nucleic acid binding protein	Alfin, PHD-finger
3P12136508	Internode length	Full (General), Sacc (General)	G	A	synonymous_variant	LOW	<i>Sobic.003G129600</i>	similar to Os01g0275900 protein	NA
3P1274812	Internode length	Spon (General)	G	A	missense_variant	MODERATE	<i>Sobic.003G014300</i>	KOG4457 - Uncharacterized conserved protein (1 of 2)	Uncharacterized conserved protein (DUF2358)
3P3400683	Internode length	Full (Additive), Full (1-dom-alt), Full (Diplo-additive)	A	G	splice_region_variant&synonymous_variant	LOW	<i>Sobic.003G036900</i>	similar to Putative uncharacterized protein	NA
3P4233123	Internode length	Full (1-dom-alt)	T	G	intron_variant	MODIFIER	<i>Sobic.003G046300</i>	similar to Ubiquitin carboxyl-terminal hydrolase	UBA/TS-N domain, Ubiquitin carboxyl-terminal hydrolase, Zn-finger in ubiquitin-hydrolases and other protein
3P4235169	Internode length	Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	A	T	synonymous_variant	LOW	<i>Sobic.003G046300</i>	similar to Ubiquitin carboxyl-terminal hydrolase	UBA/TS-N domain, Ubiquitin carboxyl-terminal hydrolase, Zn-finger in ubiquitin-hydrolases and other protein
3P4235202	Internode length	Full (1-dom-alt), Full (1-dom-alt), Full (Diplo-general)	T	C	synonymous_variant	LOW	<i>Sobic.003G046300</i>	similar to Ubiquitin carboxyl-terminal hydrolase	UBA/TS-N domain, Ubiquitin carboxyl-terminal hydrolase, Zn-finger in ubiquitin-hydrolases and other protein
3P4650393	Internode length	Spon (General)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G051100</i>	similar to Putative brassinosteroid insensitive 1-associated receptor kinase 1	Leucine rich repeat, Protein kinase domain
3P46952220	Internode length	Full (General), Full (1-dom-ref), Full	T	C	intron_variant	MODIFIER	<i>Sobic.003G180100</i>	PTHR12741//PTHR12741:S F13 - LYST-INTERACTING PROTEIN LIP5 DOPAMINE RESPONSIVE PROTEIN	1,3-beta-glucan synthase component

		(Diplo-general), Full (Diplo-additive)						DRG-1 // SUBFAMILY NOT NAMED (1 of 1)	
3P50031763	Internode length	Full (General), Sacc (General)	G	A	intron_variant	MODIFIER	<i>Sobic.003G185500</i>	similar to Putative uncharacterized protein	NA
3P58486015	Internode length	Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G245900</i>	similar to Os01g0661000 protein	Domain of unknown function (DUF588)
3P62738654	Internode length	Full (General)	G	A	intron_variant	MODIFIER	<i>Sobic.003G295200</i>	similar to Putative uncharacterized protein	Ankyrin repeats (3 copies), Protein tyrosine kinase
3P6323064	Internode length	Full (General), Full (1-dom-ref)	C	G	intron_variant	MODIFIER	<i>Sobic.003G074200</i>	similar to Putative uncharacterized protein	Actin
3P67226374	Internode length	Full (General)	G	A	synonymous_variant	LOW	<i>Sobic.003G353400</i>	similar to Zinc finger DHHC domain containing protein 2-like	DHHC palmitoyltransferase
3P67957169	Internode length	Full (General)	G	A	synonymous_variant	LOW	<i>Sobic.003G362000</i>	similar to Chromosome chr8 scaffold_23, whole genome shotgun sequence	3-Deoxy-D-manno-octulosonic-acid transferase (kdtransferase)
3P73713218	Internode length	Full (General), Sacc (General)	C	A	missense_variant	MODERATE	<i>Sobic.003G436000</i>	Predicted protein	Protein of unknown function (DUF3615)
4ind62293349	Internode length	Full (1-dom-alt), Full (Diplo-additive)	CTT CG	C	intergenic_region	MODIFIER	<i>Sobic.004G280500- Sobic.004G280600</i>	similar to RAV-like B3 domain DNA binding protein-like/NA	B3 DNA binding domain/NA
4ind67150479	Internode length	Spon (General)	A	AAA TC	intergenic_region	MODIFIER	<i>Sobic.004G339600- Sobic.004G339700</i>	similar to Ankyrin repeat protein-like/similar to Peroxisomal membrane protein-like	GPCR-chaperone/NA
4ind67150492	Internode length	Full (General), Sacc (General), Spon (General)	G	GAC CA	intergenic_region	MODIFIER	<i>Sobic.004G339600- Sobic.004G339700</i>	similar to Ankyrin repeat protein-like/similar to Peroxisomal membrane protein-like	GPCR-chaperone/NA
4P3106357	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	A	G	synonymous_variant	LOW	<i>Sobic.004G038100</i>	K14303 - nuclear pore complex protein Nup160 (NUP160, NUP120) (1 of 1)	Family of unknown function (DUF5311), Nucleoporin Nup120/160
4P4992036	Internode length	Spon (General)	T	C	missense_variant	MODERATE	<i>Sobic.004G061900</i>	similar to Putative uncharacterized protein	Sad1 / UNC-like C-terminal
4P5670529	Internode length	Sacc (General)	A	G	synonymous_variant	LOW	<i>Sobic.004G070100</i>	similar to Putative uncharacterized protein	Exostosin family
4P58275825	Internode length	Full (1-dom-alt), Full (Diplo-general)	T	C	intron_variant	MODIFIER	<i>Sobic.004G234500</i>	similar to Aspartyl protease-like	Xylanase inhibitor C-terminal, Xylanase inhibitor N-terminal
4P59750250	Internode length	Full (General)	A	G	intron_variant	MODIFIER	<i>Sobic.004G251200</i>	similar to Probable methionyl-tRNA synthetase	Putative tRNA binding domain, tRNA synthetases class I (M)
4P60209923	Internode length	Full (General)	G	A	intron_variant	MODIFIER	<i>Sobic.004G256200</i>	similar to Zinc finger protein	CCT motif
4P60825826	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	T	C	synonymous_variant	LOW	<i>Sobic.004G263200</i>	similar to Chromosome chr1 scaffold_5, whole genome shotgun sequence	Anion-transporting ATPase
4P6585299	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-	C	A	intron_variant	MODIFIER	<i>Sobic.004G080200</i>	similar to Putative uncharacterized protein	Beta-ketoacyl synthase, C-terminal domain, Beta-ketoacyl synthase, N-terminal domain

		general), Full (Diplo- additive)							
5P412571 74	Internode length	Full (General)	C	T	3_prime_UTR_variant	MODIFI ER	<i>Sobic.005G11 3300</i>	similar to Serine hydroxymethyltransferase	Serine hydroxymethyltra nsferase
5P535712 13	Internode length	Full (1- dom-ref), Full (Diplo- additive)	G	C	intergenic_region	MODIFI ER	<i>Sobic.005G12 1700- Sobic.005G12 1800</i>	similar to Expressed protein/similar to Expressed protein	NA/NA
5P561103 5	Internode length	Full (General)	T	G	intron_variant	MODIFI ER	<i>Sobic.005G05 5600</i>	similar to Coiled-coil protein	Filament-like plant protein, long coiled-coil
6ind49957 559	Internode length	Full (General)	GTC A	G	splice_region_variant&intron_varian t	LOW	<i>Sobic.006G13 7300</i>	similar to H0311C03.4 protein	Methyladenine glycosylase
6P105696 26	Internode length	Spon (General)	G	A	intron_variant	MODIFI ER	<i>Sobic.006G03 5300</i>	similar to Os04g0298600 protein	Uncharacterised conserved protein
6P461311 93	Internode length	Full (General), Full (1- dom-ref)	T	C	5_prime_UTR_variant	MODIFI ER	<i>Sobic.006G09 1500</i>	similar to Putative uncharacterized protein	K-box region, SRF-type transcription factor (DNA- binding and dimerisation domain)
7P286838 0	Internode length	Full (Additive), Full (1- dom-alt), Full (Diplo- additive)	G	A	intergenic_region	MODIFI ER	<i>Sobic.007G03 2300- Sobic.007G03 2400</i>	similar to Putative uncharacterized protein/K11462 - polycomb protein EED (EED) (1 of 2)	NA/NA
7P552669 2	Internode length	Spon (General)	G	A	intergenic_region	MODIFI ER	<i>Sobic.007G05 4400- Sobic.007G05 4500</i>	PF03101/PF04434/PF1055 1 - FAR1 DNA-binding domain (FAR1) // SWIM zinc finger (SWIM) // MULE transposase domain (MULE) (1 of 27)/similar to Putative uncharacterized protein	FAR1 DNA- binding domain, MULE transposase domain/ABC-2 type transporter
7P593583 45	Internode length	Full (General), Full (1- dom-alt), Full (Diplo- general), Full (Diplo- additive)	C	T	synonymous_variant	LOW	<i>Sobic.007G15 8900</i>	PTHR32297:SF1 - SODIUM CHANNEL MODIFIER 1 (1 of 1)	Acidic C-terminal region of sodium channel modifier 1 SCN1M1, Zinc- finger of sodium channel modifier 1
7P912778 1	Internode length	Full (1- dom-alt), Full (Diplo- general), Full (Diplo- additive)	C	T	5_prime_UTR_premature_start_codo n_gain_variant	LOW	<i>Sobic.007G07 8000</i>	similar to H0702G05.9 protein	Protein phosphatase 2C
8ind71471 09	Internode length	Sacc (General)	CCG AG	C	intron_variant	MODIFI ER	<i>Sobic.008G06 4900</i>	similar to Plant protein family protein, expressed	GDP-fucose protein O- fucosyltransferase
8P422222 5	Internode length	Sacc (General)	A	C	3_prime_UTR_variant	MODIFI ER	<i>Sobic.008G04 2900</i>	similar to Harpin-induced protein 1 containing protein, expressed	Late embryogenesis abundant protein
8P422223 3	Internode length	Full (General), Sacc (General), Spon (General)	A	T	3_prime_UTR_variant	MODIFI ER	<i>Sobic.008G04 2900</i>	similar to Harpin-induced protein 1 containing protein, expressed	Late embryogenesis abundant protein
9P471536 94	Internode length	Full (General)	C	A	intron_variant	MODIFI ER	<i>Sobic.009G12 1000</i>	similar to Putative uncharacterized protein	NA
9P495010 84	Internode length	Full (General)	A	G	synonymous_variant	LOW	<i>Sobic.009G13 8000</i>	PTHR23155/PTHR23155:S F538 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 6)	NB-ARC domain
9P542694 38	Internode length	Sacc (General)	A	G	intron_variant	MODIFI ER	<i>Sobic.009G19 0600</i>	similar to Putative uncharacterized protein OSJNBb0108E17.2	Cathepsin propeptide inhibitor domain (I29), Granulin, Papain family cysteine protease
9P572887 33	Internode length	Full (General)	G	A	synonymous_variant	LOW	<i>Sobic.009G23 3200</i>	similar to Sucrose phosphate synthase	Glycosyl transferases group 1, Sucrose synthase, Sucrose- 6F-phosphate phosphohydrolase

9P8930087	Internode length	Full (General)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.009G072900</i>	similar to Putative uncharacterized protein	NA
9P9107915	Internode length	Full (General), Full (1-dom-alt), Full (Diplo-general), Full (Diplo-additive)	G	C	intron_variant	MODIFIER	<i>Sobic.009G073400</i>	similar to Putative uncharacterized protein OSJNBa0007C23.15	Galactosyltransferase
6P42692072	Leaf length	P8 (1-dom-ref)	C	T	intergenic_region	MODIFIER	<i>Sobic.006G066500- Sobic.006G066600</i>	similar to OSJNBa0042115.15 protein/similar to DNA-binding protein S1FA2	Transcriptional activator/DNA binding protein S1FA
10P60783572	Stalk number	Sacc (General), Spon (General)	C	T	intron_variant	MODIFIER	<i>Sobic.010G274500</i>	similar to Catalase isozyme 1	Catalase, Catalase-related immune-responsive
2P65627224	Stalk number	Full (General), Sacc (General)	T	A	intergenic_region	MODIFIER	<i>Sobic.002G273000- Sobic.002G273100</i>	KOG1947 - Leucine rich repeat proteins, some proteins contain F-box (1 of 33)/similar to 30S ribosomal protein S31, mitochondrial precursor	F-box domain/Ribosomal protein S31e
3P66955169	Stalk number	P8 (General)	C	A	intergenic_region	MODIFIER	<i>Sobic.003G349600- Sobic.003G349700</i>	similar to Os01g0838600 protein/similar to C2H2 zinc finger protein	C2H2-type zinc finger/C2H2-type zinc finger
4P65103065	Stalk number	Spon (General)	T	C	intergenic_region	MODIFIER	<i>Sobic.004G314300- Sobic.004G314400</i>	similar to Putative laccase/similar to Universal stress protein / early nodulin ENOD18-like	Multicopper oxidase/Universal stress protein family
1P3184652	Water content	P8 (General), P8 (Additive)	A	G	intergenic_region	MODIFIER	<i>Sobic.001G042700- Sobic.001G042800</i>	PF04195 - Putative gypsy type transposon (Transposase_28) (1 of 16)/similar to Putative NUDIX hydrolase	NA/NUDIX domain
6P37710980	Water content	Sacc (1-dom-alt), Sacc (Diplo-general), Sacc (Diplo-additive)	C	A	intergenic_region	MODIFIER	<i>Sobic.006G050300- Sobic.006G050400</i>	similar to OSIGBa0148P16.3 protein/PF04359 - Protein of unknown function (DUF493) (DUF493) (1 of 2)	NA/Protein of unknown function (DUF493)
7P2769299	Total weight/Water content	P8 (General)/P8 (General)	G	A	synonymous_variant	LOW	<i>Sobic.007G030700</i>	similar to Putative phosphate/phosphoenolpyruvate translocator	Triose-phosphate Transporter family
3P61470337	Total weight/Water content	P8 (1-dom-alt), P8 (Diplo-general), P8 (Diplo-additive)/P8 (1-dom-alt), P8 (Diplo-general)	A	G	intron_variant	MODIFIER	<i>Sobic.003G278633</i>	NA	NA
10ind2865773	Leaf width	Sacc (General)	TAA AG	T	intron_variant	MODIFIER	<i>Sobic.010G035400</i>	similar to Putative uncharacterized protein	Domain of unknown function (DUF1995)
10ind56691142	Leaf width	Full (General), Sacc (General)	TAA C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.010G224900</i>	similar to Tubulin beta-2/beta-3 chain	Tubulin C-terminal domain, Tubulin/FtsZ family, GTPase domain
10P55831327	Leaf width	P8 (General)	C	A	synonymous_variant	LOW	<i>Sobic.010G215400</i>	NA	NA
10P7940593	Leaf width	Full (Diplo-general), Sacc (Diplo-general)	C	T	intron_variant	MODIFIER	<i>Sobic.010G091000</i>	similar to Glutaredoxin-related-like protein	Domain found in Dishevelled, Egl-10, and Pleckstrin (DEP), Glutaredoxin, Protein of unknown function, DUF547
10P952532	Leaf width	Full (General)	T	C	intergenic_region	MODIFIER	<i>Sobic.010G010700- Sobic.010G010900</i>	2.7.10.2 - Non-specific protein-tyrosine kinase / Cytoplasmic protein tyrosine kinase (1 of 46)/similar to Putative uncharacterized protein	NA/NA
10P952546	Leaf width	Full (General), Sacc (General)	T	C	intergenic_region	MODIFIER	<i>Sobic.010G010700- Sobic.010G010900</i>	2.7.10.2 - Non-specific protein-tyrosine kinase / Cytoplasmic protein tyrosine kinase (1 of 46)/similar to Putative uncharacterized protein	NA/NA

1P550693 49	Leaf width	Full (Diplo- general)	A	G	missense_variant&splice_region_var iant	MODER ATE	<i>Sobic.001G28 1100</i>	similar to Pleckstrin domain-containing protein, putative, expressed	START domain, Protein of unknown function (DUF1336)
1P636502 61	Leaf width	Full (General), P8 (Additive)	G	T	3_prime_UTR_variant	MODIFI ER	<i>Sobic.001G34 7100</i>	PF01535/PF13041/PF1443 2 - PPR repeat (PPR) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase) (1 of 48)	DYW family of nucleic acid deaminases, PPR repeat family
1P649542 22	Leaf width	Hybrid (General), P8 (General)	G	A	3_prime_UTR_variant	MODIFI ER	<i>Sobic.001G36 0000</i>	similar to Pyruvate dehydrogenase kinase isoform 1	Mitochondrial branched-chain alpha-ketoacid dehydrogenase kinase
1P664074 11	Leaf width	Full (General), Sacc (General)	A	T	intron_variant	MODIFI ER	<i>Sobic.001G37 5700</i>	similar to ASF/SF2-like pre- mRNA splicing factor SRP32	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
1P704928 06	Leaf width	Spon (General)	A	G	3_prime_UTR_variant	MODIFI ER	<i>Sobic.001G42 4800</i>	similar to Expressed protein	NA
1P719345 98	Leaf width	Sacc (General)	G	T	stop_gained	HIGH	<i>Sobic.001G44 1600</i>	similar to Tropomyosin, putative, expressed	Coiled-coil region of Oberon, PHD - plant homeodomain finger protein
1P750028 39	Leaf width	P8 (General)	A	G	3_prime_UTR_variant	MODIFI ER	<i>Sobic.001G47 7700</i>	similar to Expressed protein	WD domain, G- beta repeat
1P798535 49	Leaf width	P8 (Additive), P8 (1- dom-ref), P8 (Diplo- additive)	G	C	missense_variant	MODER ATE	<i>Sobic.001G53 5100</i>	similar to Expressed protein	NA
1P932664 2	Leaf width	P8 (General)	T	C	3_prime_UTR_variant	MODIFI ER	<i>Sobic.001G11 9900</i>	similar to DEAD-box ATP- dependent RNA helicase 21	DEAD/DEAH box helicase
2ind12470 564	Leaf width	Sacc (General), Hybrid (General), P8 (General)	A	ATT	frameshift_variant	HIGH	<i>Sobic.002G10 5200</i>	similar to Putative uncharacterized protein	Protein phosphatase 2A regulatory B subunit (B56 family)
2ind65539 620	Leaf width	Full (General), Sacc (General)	GTC A	G	3_prime_UTR_variant	MODIFI ER	<i>Sobic.002G27 2400</i>	similar to Putative uncharacterized protein	NA
2ind65916 673	Leaf width	P8 (General)	CAC CA	C	3_prime_UTR_variant	MODIFI ER	<i>Sobic.002G27 6600</i>	similar to Endoglucanase 24 precursor	Glycosyl hydrolase family 9
2P186903 74	Leaf width	Hybrid (General)	T	C	intron_variant	MODIFI ER	<i>Sobic.002G13 1100</i>	Predicted protein	Development and cell death domain
2P459713 05	Leaf width	Full (General), P8 (General)	T	G	3_prime_UTR_variant	MODIFI ER	<i>Sobic.002G15 3700</i>	similar to Thyroid hormone receptor-associated protein complex component TRAP170-like protein	Mediator complex subunit MED14
2P545647 94	Leaf width	Spon (General)	A	T	intergenic_region	MODIFI ER	<i>Sobic.002G17 2401- Sobic.002G17 2500</i>	similar to GTP-binding protein, ras-like/similar to Putative uncharacterized protein	Ras family/DYW family of nucleic acid deaminases, PPR repeat family
2P592011 87	Leaf width	P8 (General)	G	A	synonymous_variant	LOW	<i>Sobic.002G20 2300</i>	similar to Putative golgi SNAP receptor complex member 1	Snare region anchored in the vesicle membrane C-terminus
2P595507 46	Leaf width	Spon (General)	C	T	intron_variant	MODIFI ER	<i>Sobic.002G20 4000</i>	similar to 26S proteasome regulatory subunit-like protein	NA
2P661586 03	Leaf width	Full (General), Sacc (General)	T	C	synonymous_variant	LOW	<i>Sobic.002G27 9900</i>	similar to Beta- galactosidase 12 precursor	Galactose binding lectin domain
2P669467 32	Leaf width	P8 (General)	C	T	intron_variant	MODIFI ER	<i>Sobic.002G29 1300</i>	similar to Putative uncharacterized protein	NUDIX domain
2P752033 41	Leaf width	Full (General), Sacc (General)	G	A	synonymous_variant	LOW	<i>Sobic.002G40 1500</i>	similar to Putative uncharacterized protein	Sec23-binding domain of Sec16
2P806235 5	Leaf width	Full (General), P8 (General)	A	C	missense_variant	MODER ATE	<i>Sobic.002G07 7400</i>	PF00069/PF00560/PF0826 3/PF13855 - Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine rich repeat N-terminal domain (LRRNT_2) // Leucine rich repeat (LRR_8) (1 of 55)	Leucine rich repeat, Protein kinase domain
3ind54181 30	Leaf width	P8 (General)	GCT C	G	intron_variant	MODIFI ER	<i>Sobic.003G06 1900</i>	similar to Os01g0159300 protein	Ring finger domain
3ind55473 132	Leaf width	Spon (General)	AT	A	intron_variant	MODIFI ER	<i>Sobic.003G21 9400</i>	similar to Heterogeneous nuclear ribonucleoprotein A2/B1-like	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)

3P18402714	Leaf width	Hybrid (General)	C	T	intron_variant	MODIFIER	<i>Sobic.003G159500</i>	similar to Os05g0154600 protein	Ring finger domain
3P31596498	Leaf width	Spon (General)	G	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G170700</i>	KOG1893 - Uncharacterized conserved protein (1 of 1)	Piezo non-specific cation channel, R-Ras-binding domain
3P49105822	Leaf width	Hybrid (General)	T	A	intergenic_region	MODIFIER	<i>Sobic.003G184400- Sobic.003G184500</i>	PTHR12480:SF6 - JMJC DOMAIN-CONTAINING PROTEIN 4 (1 of 1)/similar to Putative uncharacterized protein	Cupin-like domain/Neprosin activation peptide
3P5244300	Leaf width	Hybrid (General)	G	A	intron_variant	MODIFIER	<i>Sobic.003G059400</i>	similar to OSJNBa0065H10.9 protein	BP28CT (NUC211) domain, U3 small nucleolar RNA-associated protein 10
3P53284899	Leaf width	Full (General), Sacc (General), Hybrid (General), P8 (General)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G202900</i>	similar to Putative uncharacterized protein	NA
3P55473119	Leaf width	Spon (General)	A	G	intron_variant	MODIFIER	<i>Sobic.003G219400</i>	similar to Heterogeneous nuclear ribonucleoprotein A2/B1-like	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
3P56030744	Leaf width	Full (General), Sacc (General), P8 (General)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G224300</i>	similar to Putative uncharacterized protein	Ribosomal protein L6
3P57179114	Leaf width	Full (General), Sacc (General)	A	C	synonymous_variant	LOW	<i>Sobic.003G232700</i>	similar to Putative polygalacturonase	Polypeptide deformylase
3P58845520	Leaf width	Full (General), Sacc (General), P8 (General)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G250100</i>	PF00046 - Homeobox domain (Homeobox) (1 of 64)	Homeobox domain
3P60794079	Leaf width	Full (Additive)	T	C	synonymous_variant	LOW	<i>Sobic.003G271400</i>	PTHR11689:SF14 - CHLORIDE CHANNEL PROTEIN 2 (1 of 1)	Voltage gated chloride channel
3P6606911	Leaf width	P8 (1-dom-alt), P8 (Diplo-general)	G	T	synonymous_variant	LOW	<i>Sobic.003G077500</i>	similar to Os01g0141900 protein	NA
3P66942928	Leaf width	P8 (1-dom-alt), P8 (Diplo-general)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G349600</i>	similar to Os01g0838600 protein	C2H2-type zinc finger
3P67095919	Leaf width	Full (General), Sacc (General)	G	A	3_prime_UTR_variant	MODIFIER	<i>Sobic.003G351900</i>	similar to Endosperm specific protein-like	Fasciclin domain
3P68383201	Leaf width	Full (Diplo-general)	G	C	5_prime_UTR_variant	MODIFIER	<i>Sobic.003G368000</i>	similar to Putative uncharacterized protein	F-box domain, Tub family
3P917667	Leaf width	Full (General), Sacc (General)	C	G	intergenic_region	MODIFIER	<i>Sobic.003G010650- Sobic.003G010700</i>	NA/similar to Putative uncharacterized protein	NA/Transcriptional repressor, ovate
4P12574117	Leaf width	Spon (General)	C	T	5_prime_UTR_variant	MODIFIER	<i>Sobic.004G118800</i>	similar to Os02g0459600 protein	Legume lectin domain, Protein kinase domain
4P16089454	Leaf width	Full (General), Sacc (General), P8 (General)	C	A	intron_variant	MODIFIER	<i>Sobic.004G128200</i>	similar to Putative uncharacterized protein	Leucine rich repeat N-terminal domain, Protein kinase domain
4P22574534	Leaf width	Spon (General)	T	G	intron_variant	MODIFIER	<i>Sobic.004G134200</i>	PTHR10751//PTHR10751:SF44 - GUANYLATE BINDING PROTEIN // SUBFAMILY NOT NAMED (1 of 2)	Guanylate-binding protein, C-terminal domain, N-terminal domain
4P51358883	Leaf width	Sacc (General), P8 (General)	T	C	missense_variant	MODERATE	<i>Sobic.004G164200</i>	similar to Copper chaperone homolog CCH	Heavy-metal-associated domain
4P53646791	Leaf width	Full (General), Sacc (General)	C	T	3_prime_UTR_variant	MODIFIER	<i>Sobic.004G182600</i>	weakly similar to Os06g0500300 protein	NA
4P57110289	Leaf width	Full (Diplo-general),	T	A	intergenic_region	MODIFIER	<i>Sobic.004G220700- Sobic.004G22</i>	similar to Phenylalanine ammonia-lyase/NA	Aromatic amino acid lyase/NA

		P8 (1-dom-ref)					0900	
4P57110290	Leaf width	P8 (1-dom-ref)	A	T	intergenic_region	MODIFIER	<i>Sobic.004G220700- Sobic.004G220900</i>	similar to Phenylalanine ammonia-lyase/NA Aromatic amino acid lyase/NA
4P58413532	Leaf width	P8 (General)	T	C	intron_variant	MODIFIER	<i>Sobic.004G236300</i>	similar to Shikimate kinase 1 Shikimate kinase
5P11920928	Leaf width	Hybrid (General)	G	A	intron_variant	MODIFIER	<i>Sobic.005G086200</i>	similar to PHD finger protein, putative, expressed Alfin, PHD-finger
5P9496728	Leaf width	P8 (General)	G	T	intergenic_region	MODIFIER	<i>Sobic.005G076800- Sobic.005G077001</i>	NA/PF00076//PF03478 - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) (RRM_1) // Protein of unknown function (DUF295) (DUF295) (1 of 2) NA/RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
6P19319761	Leaf width	P8 (General)	C	T	intergenic_region	MODIFIER	<i>Sobic.006G040101- Sobic.006G040201</i>	PF03004 - Plant transposase (PttA/En/Spm family) (Transposase_24) (1 of 16)/PTHR23002//PTHR23002:SF63 - ZINC FINGER CCHC DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 162) Plant transposase (PttA/En/Spm family)/NA
6P19319765	Leaf width	P8 (General)	T	C	intergenic_region	MODIFIER	<i>Sobic.006G040101- Sobic.006G040201</i>	PF03004 - Plant transposase (PttA/En/Spm family) (Transposase_24) (1 of 16)/PTHR23002//PTHR23002:SF63 - ZINC FINGER CCHC DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 162) Plant transposase (PttA/En/Spm family)/NA
6P2803959	Leaf width	Sacc (General)	T	C	intergenic_region	MODIFIER	<i>Sobic.006G017500- Sobic.006G017600</i>	weakly similar to Os04g0166000 protein/K03017 - DNA-directed RNA polymerase II subunit RPB9 (RPB9, POLR2I) (1 of 5) NA/RNA polymerases M/15 Kd subunit, Transcription factor S-II (TFIIS)
6P48326143	Leaf width	Full (General), Sacc (General), P8 (General)	G	T	missense_variant	MODERATE	<i>Sobic.006G115466</i>	NA NA
6P53031390	Leaf width	Full (General), Sacc (General)	C	T	missense_variant	MODERATE	<i>Sobic.006G174850</i>	K15193 - protein SPT2 (SPTY2D1, SPT2) (1 of 2) SPT2 chromatin protein
7ind6235143	Leaf width	Sacc (General)	G	GAC	intron_variant	MODIFIER	<i>Sobic.007G059600</i>	similar to Benzothiadiazole-induced somatic embryogenesis receptor kinase 1 Leucine rich repeat N-terminal domain, Protein kinase domain
7P54423302	Leaf width	Full (General)	T	C	missense_variant	MODERATE	<i>Sobic.007G128500</i>	Predicted protein NA
7P61801701	Leaf width	Sacc (General)	T	C	synonymous_variant	LOW	<i>Sobic.007G184700</i>	similar to Putative uncharacterized protein OJ1211_G06.22 Plant protein of unknown function (DUF936)
7P64752471	Leaf width	Spon (General)	C	T	synonymous_variant	LOW	<i>Sobic.007G219200</i>	similar to Putative uncharacterized protein Endomembrane protein 70
7P787036	Leaf width	Spon (General)	G	A	intergenic_region	MODIFIER	<i>Sobic.007G008600- Sobic.007G008700</i>	similar to Os08g0108700 protein/weakly similar to Putative BLE2 protein Glycolipid transfer protein (GLTP)/NA
7P8264992	Leaf width	Hybrid (General)	C	T	intergenic_region	MODIFIER	<i>Sobic.007G073100- Sobic.007G073300</i>	similar to Polyamine oxidase precursor/similar to Putative serine carboxylase II-3 Flavin containing amine oxidoreductase/serine carboxypeptidase
8P2787878	Leaf width	Sacc (General)	T	C	synonymous_variant	LOW	<i>Sobic.008G031200</i>	similar to Pentatricopeptide, putative, expressed PPR repeat family
8P62190177	Leaf width	Full (General), Sacc (General)	A	G	intergenic_region	MODIFIER	<i>Sobic.008G187200- Sobic.008G187300</i>	similar to Putative uncharacterized protein/NA short chain dehydrogenase/NA
9P180773	Leaf width	Spon (General)	T	C	3_prime_UTR_variant	MODIFIER	<i>Sobic.009G001900</i>	similar to Os05g0103500 protein NA
9P2630254	Leaf width	Spon (General)	G	A	intron_variant	MODIFIER	<i>Sobic.009G029500</i>	similar to Putative Zn transporter Cation efflux family
9P3981443	Leaf width	Full (General), Sacc (General), P8 (General)	A	G	synonymous_variant	LOW	<i>Sobic.009G041400</i>	PTHR23155//PTHR23155:S F525 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 1) NB-ARC domain
9P3981469	Leaf width	Full (General), Sacc	C	T	missense_variant	MODERATE	<i>Sobic.009G041400</i>	PTHR23155//PTHR23155:S F525 - LEUCINE-RICH REPEAT-CONTAINING NB-ARC domain

		(General), P8 (General)						PROTEIN // SUBFAMILY NOT NAMED (1 of 1)	
9P398147 6	Leaf width	Full (General), Sacc (General), P8 (General)	A	T	synonymous_variant	LOW	<i>Sobic.009G04 1400</i>	PTHR23155//PTHR23155:S F525 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 1)	NB-ARC domain
9P398653 7	Leaf width	Sacc (General), P8 (General)	G	A	intron_variant	MODIFI ER	<i>Sobic.009G04 1600</i>	similar to Unknow protein	CRS1 / YhbY (CRM) domain
9P566950 54	Leaf width	Full (General), Sacc (General), Hybrid (General), P8 (General)	A	G	synonymous_variant	LOW	<i>Sobic.009G22 5000</i>	similar to AT5g49540/K6M13_9	Rab5-interacting protein (Rab5ip)
9P580673 02	Leaf width	Full (General), Sacc (General)	A	T	splice_region_variant&intron_varian t	LOW	<i>Sobic.009G24 4900</i>	similar to (S)-adenosyl-L- methionine:delta 24-sterol methyltransferase	Sterol methyltransferase C-terminal
9P580673 06	Leaf width	Full (General), Sacc (General), Hybrid (General), P8 (General)	C	A	missense_variant&splice_region_var iant	MODER ATE	<i>Sobic.009G24 4900</i>	similar to (S)-adenosyl-L- methionine:delta 24-sterol methyltransferase	Sterol methyltransferase C-terminal
9P675085 8	Leaf width	Spon (General)	C	G	intron_variant	MODIFI ER	<i>Sobic.009G06 3400</i>	similar to Cellulose synthase-1	Cellulose synthase, Zinc- binding RING- finger
9P756121 4	Leaf width	P8 (General)	T	A	3_prime_UTR_variant	MODIFI ER	<i>Sobic.009G06 8800</i>	similar to Os05g0182900 protein	NA

Note: ^a GWAS runs in Full set and several subests. Full = all accessions in the diversity panel; Sacc = 273 accessions from *S. spontaneum*, *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; Spon = 100 accessions from *S. spontaneum*; Hybrid = 173 accessions from *S. officinarum*, modern *S. hybrids* and ancient *S. hybrids*; and P8 = 162 accessions with the same ploidy level (ploidy =8).

Figure S1 Distribution of the nine yield components in the sugarcane diversity panel, (A) Stalk number (Stalk); (B) Internode length (cm) (Internode); (C) Stalk diameter (mm) (Diameter); (D) Leaf width (mm) (Width); (E) Leaf length (cm) (length); (F) Brix; (G) Total weight (kg) (Weight); (H) Dry weight (kg) (Dry); (I) Water content (kg) (Water). For each trait, Z-score for each accession was calculated in each replicate, and Z scores of the diversity panel in three replicates were plotted.

856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876

Figure S2 Significant marker-trait associations identified in full set of the sugarcane diversity panel with 300 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

Figure S3 Significant marker-trait associations identified in *Saccharum* set of the sugarcane diversity panel with 273 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

921

922

923

924

925

926

927

Figure S4 Significant marker-trait associations identified in *Saccharum spontaneum* set of the sugarcane diversity panel with 100 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

Figure S5 Significant marker-trait associations identified in sugarcane hybrid set of the sugarcane diversity panel with 173 accessions. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

979

980

981

982

983

984

985

Figure S6 Significant marker-trait associations identified in the sugarcane diversity panel with 162 accessions with ploidy = 8. Stalk = Stalk number; Diameter = Stalk diameter; Width = Leaf width; Length = Leaf length; Weight = Total weight; Dry = Dry weight; Water = Water content; Internode = Internode length.

1007

1008

1009

1010

1011

1012

1013

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034