Regional genomic heritability mapping for agronomic traits in sugarcane

Pedro Marcus Pereira Vidigal[1] (0000-0002-5116-9856), Mehdi Momen[2,3] (0000-0002-2562-2741), Paulo Mafra de Almeida Costa[4](0000-0001-5190-2678), Márcio Henrique Pereira Barbosa[5] (0000-0002-5838-9579), Gota Morota[2] (0000-0002-3567-6911), Luiz Alexandre Peternelli[6*] (0000-0002-7683-0583).

[1] Universidade Federal de Viçosa (UFV), Núcleo de Análise de Biomoléculas (NuBioMol), Viçosa, Minas Gerais, Brazil;
[2] Virginia Polytechnic Institute and State University (Virginia Tech), Department of Animal and Poultry Sciences, Blacksburg, Virginia, USA;
[3] University of Wisconsin–Madison (UW-Madison), Department of Surgical Sciences, School of Veterinary Medicine, Madison, Wisconsin, USA;
[4] Instituto Federal de Educação, Ciência e Tecnologia Catarinense, Concórdia, Santa Catarina, Brazil;
[5] Universidade Federal de Viçosa (UFV), Department of Agronomy, Viçosa, Minas Gerais, Brazil;
[6] Universidade Federal de Viçosa (UFV), Department of Statistics, Viçosa, Minas Gerais, Brazil;

*Corresponding author: Luiz Alexandre Peternelli (0000-0002-7683-0583). E-mail: peternelli@ufv.br.

Keywords: genotyping; GWAS; regional heritability mapping; single nucleotide polymorphism; sugarcane.

Abstract

Background. The identification of genomic regions involved in agronomic traits is the primary concern for sugarcane breeders. Genome-wide association studies (GWAS) leverage the sequence variations to bridge phenotypes and genotypes. However, their effectiveness is limited in species with high ploidy and large genomes, such as sugarcane. As an alternative, a regional heritability mapping (RHM) method can be used to capture genetic signals that may be missed by GWAS by combining genetic variance from neighboring regions. We used RHM to screen the sugarcane genome aiming to identify regions with higher heritability associated with agronomic traits. We considered percentage of fiber in sugarcane bagasse (FB), apparent percentage of sugarcane sucrose (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare (TSH).
Methods. Sequence-capture data of 508 sugarcane (*Saccharum* spp.) clones from a breeding population under selection were processed for variant calling analysis using the sugarcane genome cultivar R570 as a reference. A set of 375,195 single nucleotide polymorphisms were selected after quality control. RHM was conducted by splitting the sugarcane genome into windows of 2 Mb length.
Results. We selected the windows explaining > 20% of the total genomic heritability for TPH (64 windows - 5,654 genes) and TSH (72 windows - 6,050 genes), and > 15% for PC (16 windows - 1,517 genes) and FB (17 windows - 1,615 genes). The top five windows that explained the highest genomic heritability ranged from 20.8 to 24.6% for FB (629 genes), 18.0 to 22.0% for PC (452 genes), 53.8 to 66.0% for TPH (705 genes), and 59.5 to 67.4% for TSH (413 genes). The functional annotation of genes included in those top five windows

revealed a set of genes that encode enzymes that integrate carbon metabolism, starch and sucrose metabolism, and phenylpropanoid biosynthesis pathways.

Conclusions. The selection of windows that explained the large proportions of genomic heritability allowed us to identify genomic regions containing a set of genes that are related to the agronomic traits in sugarcane. These windows spanned a region of 58.38Mb, which corresponds to 14.28% of the reference assembly in the sugarcane genome. We contend that RHM can be used as an alternative method for sugarcane breeders to reduce the complexity of the sugarcane genome.

1. Introduction

Sugarcane (*Saccharum* spp.) is a perennial C_4 grass of the Poaceae family, which is an economically important crop for the sugar and biofuels industries. Its cultivars are multiplied through vegetative propagation and are primarily grown in tropical and subtropical regions (Barbosa et al., 2012).

Improvement of sugarcane is a challenging process guided by conventional methods, which demand time and is hindered by its genomic complexity (Barbosa et al., 2012). Due to its long production cycle, sugarcane breeding programs take at least 11 years to release a new cultivar (Barbosa et al., 2012; Peternelli et al., 2018). Sugarcane breeders also must deal with the complexity of the sugarcane genome that exceeds other crops, which is the product of interspecific hybridization between *Saccharum officinarum* and *S. spontaneum* originated from modern cultivars. These cultivars are highly heterozygous, aneuploid, and have large genomes with a variable number of chromosomes (Lu et al., 1994; D'Hont et al., 2001; Piperidis, Piperidis & D'Hont, 2010; Garsmeur et al., 2018).

High throughput sequencing technologies have been enabling significant advances in uncovering and understanding the genomic complexity of sugarcane. A comparison of the sugarcane genetic maps with other Poaceae species maps revealed microsynteny with the *Sorghum bicolor* genome (Figueira et al., 2012; Aitken et al., 2014; Yang et al., 2017). Analysis of transcriptomes further confirmed that sugarcane and sorghum transcripts share more than 90% of sequence identity (Nishiyama et al., 2014; Yang et al., 2017). Thus, the sorghum genome became the most suitable reference for sequence variation analysis in sugarcane and is of frequent use in linkage and quantitative trait loci (QTL) mapping. However, the recently released reference sequence of the sugarcane cultivar R570 monoploid genome (Garsmeur et al., 2018) opens a new opportunity for sugarcane geneticists. This reference is a high-quality sequence that represents the gene space of the sugarcane monoploid genome, which contains genes annotated in their genomic context with the respective regulatory elements.

Identification of genomic regions involved in agronomic traits related to yield and disease resistance by performing molecular marker-assisted selection is the primary concern for sugarcane breeders. To increase the resolution of single nucleotide polymorphism (SNP) genotyping, sequence-capture methods have been used as a powerful tool to assess sequence variation in target regions across the genomes and to reduce costs in comparison with whole-genome sequencing. Genome-wide association studies (GWAS) leverage these sequence variations to bridge phenotypes and genotypes. GWAS also exploit the high amount of linkage disequilibrium (LD) in sugarcane as a more suitable alternative for bi-parental QTL mapping (Raboin et al., 2008). However, many instances of linked markers still will not be

recognized due to the confounding effect of polyploidy of the sugarcane genome (Raboin et al., 2008).

Additionally, GWAS in high ploidy species, such as sugarcane, is also limited by high sequencing depth needed to call variants from the large genome and the difficulty in determining the dosage of markers (Meirmans, Liu & Van Tienderen, 2018). For instance, only a few significantly associated SNPs have been detected in agronomic traits (Yang et al., 2018; Fickett et al., 2019). As an alternative, a regional heritability mapping (RHM) method can be used to capture genetic signals that may be missed by GWAS by combining genetic variance from neighboring regions (Nagamine et al., 2012; Shirali et al., 2016; Resende et al., 2018). Here, we analyzed sequence-capture data of sugarcane clones from a breeding population under selection, which were phenotypically evaluated. We used the RHM method to identify genomic regions that explain significant additive genetic variance in agronomic traits by screening the whole sugarcane genome.

2. Materials and Methods

2.1. Plant material

The sugarcane clones analyzed in this study were selected from a breeding population of the Sugarcane Breeding Program (Programa de Melhoramento da Cana-de-Açúcar, PMGCA) at the Federal University of Viçosa (Universidade Federal de Viçosa, UFV). This population consisted of 508 clones from 100 half-sib families, originated from crossings between elite sugarcane clones and commercial cultivars.

These crossings were performed in 2010 at the Serra do Ouro's experimental station located in Murici (Alagoas, Brazil) (09°13' S, 35°50' W, 450 m altitude). The seedlings were produced at PMGCA's experimental station located in Oratórios (Minas Gerais, Brazil) (20°25' S, 42°48' W, 494 m altitude). The seedlings were conducted in the first phase trials (T1) performed in 2011 (plant cane) and 2012 (first ratoon). Sugarcane clones analyzed were originated from plants selected in T1 and advanced to the second phase trial (T2) in July 2012, which was conducted in an augmented block design experiment (Federer & Raghavarao, 1975). In the T2 trial, the experimental plots consisted of one 4 m row with a spacing of 1 m between plots. The clones (unreplicated) and two common reference cultivars (checks replicated once each) were arranged in 49 augmented blocks. The reference cultivars were RB867515 (Barbosa et al., 2001) and SP80-1842, which are cultivated in large areas and commonly used as checks in breeding experiments in Brazil.

2.2. Phenotypic data

Sugarcane clones were phenotypically evaluated in plant cane in July 2013 after 12 months of growth. Ten stalks randomly taken from each plot were used for estimating the tonnes of stalks per hectare (TSH), percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), and tonnes of pol per hectare (TPH).

TSH was obtained from the total number of stalks per row and the wet weight of 10 stalks determined with a dynamometer (Castro et al., 2016). FB was determined on a wet basis from a 500 g sample of the shredded stalk (Tanimoto, 1964; Legendre, 1992) as $[FB(\%) = ((100 \times DM) - (WM \times Brix))/(5 \times (100 - Brix))]$, where DM and WM are dry and wet mass of

the sample removed from a hydraulic press and Brix is the juice Brix measured by refractometer. The apparent percentage of sucrose in juice (polarization, POL) was measured by polarimetric determination after juice extraction from 500 g samples crushed in a hydraulic press (Schneider, 1979) and used to derive the apparent percentage of sucrose in sugarcane (PC), according to the following expression (Baffa et al., 2014): $[PC\ (\%) = POL \times (1 - 0.01 \times FB) \times (0.9961 - 0.0041 \times FB)]$. The trait TPH, expressed as a percentage of apparent sucrose on a fresh weight basis, was estimated as $[TPH\ (\%) = (TSH \times PC) / 100]$.

2.3. Statistical analysis of phenotypic data

We analyzed the phenotypic data through the mixed model methodology using the software Selegen-REML/BLUP (Resende, 2016). Variance components were estimated by restricted maximum likelihood (REML), and the genotypic effects of the clones were predicted by BLUP. We used the following linear mixed model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zu} + \mathbf{Wb} + \mathbf{e},$$

where **y** is the vector of phenotypic observations for each trait; **1** is a vector of 1s; μ is the overall mean; $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_u)$ is the vector of random genotypic effects; $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_b)$ is the vector of random block effects; and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_e)$ is the vector of residuals. **Z** and **W** are incidence matrices relating the observations to the respective model effects.

The statistical significance of the effects was tested using the likelihood ratio test under the analysis of deviance theory (Resende, 2016). The predicted genotypic values ($\hat{\mu} + \hat{u}_i$) of the clones (i = 1 to n) are listed in Table S1 and were considered as response variables in the RHM analysis.

2.4. Genotypic data: Genomic DNA sequencing

The total genomic DNA of sugarcane clones was extracted using the DNeasy Plant Mini Kit of Qiagen® following the manufacturer's guidelines. The genomic libraries were produced and sequenced by RAPiD Genomics (Florida, USA). In this sequencing, the single-end libraries were built using a capture-seq methodology (Neves et al., 2013), which includes a set of probes to capture non-repetitive and evenly distributed sequences in the sugarcane genome.

Briefly, a set of 50,000 unique sequences was identified from: i) existing expressed sequence tags (ESTs) from public sugarcane cDNA libraries and ii) whole-shotgun genome sequences available publicly, consistently distributed in the genome and assuming synteny to the sorghum genome. Biotinylated 120-mer probes that complement a segment of each of the 50,000 target regions of the sugarcane genome were synthesized and were utilized to capture sequences at each target locus. The sequencing yielded a dataset of 4.77 billion reads containing sequences with 100nt or 150nt in length.

2.5. Bioinformatic and genetic analysis of genotypic data

2.5.1. Reference genome

To evaluate the sequence variation, we selected the genome of sugarcane cultivar R570 (Grasmeur et al. 2018) as a reference, which is available at Sugarcane Genome Hub (http://sugarcane-genome.cirad.fr/organism/R570-Sugarcane/cultivar). This genome is a

Single Tilling Path (STP) assembly of 408.94Mb containing ten chromosomes and 24,341 annotated genes. The unplaced contigs were not considered for the mapping analysis.

2.5.2. Mapping analysis

The raw reads of each capture-seq library were first trimmed to remove sequence adapters and poorly sequenced regions using Trimmomatic version 0.38 (Bolger, Lohse & Usadel, 2014). Trimmed reads were mapped to the reference genome using the BWA-MEM algorithm of BWA version 0.7.17 (http://bio-bwa.sourceforge.net/) (Li & Durbin, 2009). A flag identifying the several sugarcane clones was added to each mapping file. Then, the Sequence Alignment Map files were processed using SortSam, MarkDuplicates, and BuildBamIndex tools in Picard version 2.18.27 (https://github.com/broadinstitute/picard/). As a result, we produced ordered and deduplicated Binary Alignment Map files, containing ordered and deduplicated data. A schematic overview of the analysis conducted in the current study is shown in Figure S1.

2.5.3. Variant calling

Variants were called using FreeBayes version 1.2.0 (https://github.com/ekg/freebayes) (Garrison & Marth, 2012) with a minimum mapping quality of 20, minimum base quality of 20, and minimum coverage of 20 reads at every position in the reference genome. After variant calling, SNPs were filtered using vcftools version 0.16.15 (https://vcftools.github.io/index.html), Bcftools version 1.9 (https://samtools.github.io/bcftools/), and in-house AWK shell scripts. Among the polymorphic loci detected, we selected those with biallelic SNPs with less than 25% of missing data. Missing genotypes were imputed by Beagle version 5.1 (Browning, Zhou & Browning, 2018), using a flexible localized haplotype-cluster model to group locally similar haplotypes into clusters.

2.5.4. Regional heritability mapping (RHM) analysis

We performed RHM based on the variance component method described in Nagamine et al. (2012) using REACTA (Regional Heritability Advanced Complex Trait Analysis) version 0.97 (Canela-Xandri et al., 2015). We concatenated the ten chromosomes and split the whole sugarcane genome into 409 overlapping windows with an average length of 2 Mb to estimate the proportion of phenotypic variance explained by all genome-wide SNPs or a subset of SNPs using REML. The following mixed model is considered:

$$\mathbf{y}^* = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{Q}\mathbf{r} + \mathbf{e},$$

where $\mathbf{y}^*$ is the vector of adjusted phenotypic observations (genotypic values), $\mathbf{1}$ is a vector of 1s, μ is the overall mean, $\mathbf{Z}$ and $\mathbf{Q}$ are the design matrices for the whole (without the window) and the regional random effects, respectively. The distributions and covariance structures of $\mathbf{g}$ and $\mathbf{r}$ were $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma^2_g)$ and $\mathbf{r} \sim N(\mathbf{0}, \mathbf{G}_r\sigma^2_r)$, respectively. The residual term followed $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_e)$. We also run a model using all genome-wide SNPs without any window to estimate the total genomic heritability for each trait (h^2_G).

A Bonferroni correction based on the number of independent windows was used to obtain a genome-wide significance threshold for the RHM analysis, [$\alpha_{critical} = 0.05/(0.5 \times Num.$

Windows)], as previously proposed (Nagamine et al., 2012) and implemented (Shirali et al., 2016) in the literature.

2.5.5. Functional analysis of genomic windows

Among the analyzed genomic regions, we selected the windows explaining above 20% of the total genomic heritability for TPH and TSH ($h^2_r / h^2_G \geq 20\%$) and above 15% for PC and FB ($h^2_r / h^2_G \geq 15\%$), where $h^2_r = \sigma^2_r / (\sigma^2_g + \sigma^2_r + \sigma^2_e)$ and $h^2_G = \sigma^2_g / (\sigma^2_g + \sigma^2_r + \sigma^2_e)$. The gene content of these windows was identified using bedtools version 2.28.0 and the gff3 genome annotation file of the sugarcane cultivar R570 genome. The protein sequences encoded by the genes included in those windows were functionally characterized through similarity searches using BLAST version 2.6.0 (Altschul et al., 1990), Blast2GO (Gotz et al., 2008), and KAAS (KEGG Automatic Annotation Server) (Moriya et al., 2007). The lists of genes that explained high heritability for each trait were compared through Venn diagrams using the jvenn package (Bardou et al., 2014). Additionally, the gene ontology (GO) terms of these genes were further summarized using REViGO (Supek et al., 2011).

3. Results

3.1. Phenotypic variation of analyzed traits

Phenotypic analysis of the sugarcane clones showed a considerable phenotypic variation for the traits analyzed suggesting that there is genotypic variability to be exploited (Figure 1; Tables S1 and S2). The genotypic values obtained with the adjusted model, described above, ranged from 8.39 to 16.43 (average of 11.09) for a percentage of FB, from 6.50 to 17.00 (av. 13.81) for PC, from 6.81 to 20.66 (av. 14.83) for TPH, and from 51.52 to 152.23 (av. 107.96) for TSH (Table S1). Genotypic variance (σ^2_u) and broad-sense heritability (H^2) were 2.21 and 0.56 for FB; 2.02 and 0.71 for PC; 7.48 and 0.58 for TPH; 389.98 and 0.57 for TSH, respectively (Table S2).

Pearson's pairwise correlations between adjusted genetic values of FB, PC, TPH, and TSH are shown in Figure 1 and evaluated for their significance (P-value ≤ 0.01). We observed FB and PC were negatively correlated (-0.18). TPH was positively correlated with PC (0.46) and highly correlated with TSH (0.86). FB was not significantly correlated with TSH and TPH. We observed a weak and negligible correlation between TSH and PC.

3.2. Genotyping analysis and variant selection

Among all 4.77 billion reads sequenced from the 508 sugarcane clones, 4.58 billion (96.01%) were trimmed and selected for being mapped on the reference genome (~ 9.02 million of reads per sugarcane clone). These reads were mapped with an average rate of 75.73% (ranging from 71.39 to 77.94% among the sugarcane clones) and genome coverage of 1.46X (ranging from 1.14 to 1.88 among the sugarcane chromosomes) (Table 1).

From those reads, 15.41 million polymorphic loci were identified among the sugarcane clones, and 814,987 SNPs were subsequently filtered by selecting biallelic loci with less than 25% of missing data. After SNP imputation and quality control, 375,195 SNPs were selected for the RHM analysis, which corresponds to a frequency of 917.48 SNPs/Mb in the sugarcane

genome (Table 1). Among sugarcane chromosomes (Sh), Sh05 was the least polymorphic (772.81 SNPs/Mb), and Sh04 was the most polymorphic (1,025.88 SNPs/Mb).

3.3. Regional heritability mapping analysis

The calculated total genomic heritability (h^2_G) was 0.798 ± 0.233 for FB, 0.932±0.133 for PC, 0.369 ± 0.256 for TPH, and 0.383 ± 0.210 for TSH. A suggestive significance threshold of 3.61 for $-\log_{10}$ P-value (i.e., P-value = 0.000244) was calculated for window selection in the RHM analysis. None of the analyzed windows were associated with the traits with P-values above this threshold (Figure S2). Therefore, the selection of windows was proceeded based on window heritability (h^2_r), and the proportion of h^2_G explained (h^2_r / h^2_G) aiming to identify regions that could be further exploited for prospecting genes (Figure 2).

Among the analyzed windows, we selected those that explained 15% of h^2_G for FB (17 windows - 1,615 genes) and PC (16 windows - 1,517 genes), and those that explained 20% of h^2_G for TPH (64 windows - 5,654 genes) and TSH (72 windows - 6,050 genes) (Figures 2 and 3; Table S3). Even though distributed across all the chromosomes, none of these selected windows was related to all traits jointly. One window was related to FB, PC, and TPH (w146 - Sh03: 31,787,519 to 34,081,415 - 137 genes), and two windows were related to PC, TPH, and TSH (w90 - Sh02: 25,073,847 to 27,836,987 - 153 genes; w393 - Sh10: 19,693,048 to 22,127,732 - 85 genes). TPH and TSH were the traits that shared the highest number of windows (32 windows - 3,043 genes).

The top five windows with high heritability explained 20.8 to 24.6% of h^2_G for FB (629 genes), 18.0 to 22.0% for PC (452 genes), 53.8 to 66.0% for TPH (705 genes), and 59.5 to 67.4% for TSH (413) (Figure 3). Among these windows, w297 (Sh07: 14688970 to 17462449 - 121 genes) was the only one related to two traits, and it explained 63.5% of the total heritability for TPH and 66.1% for TSH.

Functional annotations of all the 2,078 genes included in those top five windows generated a non-redundant list of the biological process containing 68 GOs for FB, PC and TPH, and 48 GOs for TSH (Figure 4). Also, 712 genes (34.26%) were classified into 244 KEGG ortholog groups, which were mapped to 104 KEGG pathways, including carbon metabolism, starch, and sucrose metabolism, and phenylpropanoid biosynthesis pathways (Table 2).

4. Discussion

Genotyping by sequencing (GBS) coupled with capture-seq methods have been a practical approach to survey polymorphisms in the genomes of *Saccharum* species. It reduces the complexity of genomes and allows genome-wide analysis aiming to develop molecular markers (Song et al., 2016; Balsalobre et al., 2017; Yang et al., 2017, 2018, 2019; Fickett et al., 2019). However, the complexity of the sugarcane genome still poses some challenges for the widespread use of GBS and the adoption of molecular marker-assisted selection in breeding programs. The high and variable ploidy of the allopolyploid genome (Thirugnanasambandam, Hoang & Henry, 2018) makes the identification of SNPs significantly associated with agronomic traits of interest in sugarcane crop a challenging task. The identification of significant SNPs would demand a high coverage when using the GBS approach, which increases the costs involved in population studies (Meirmans, Liu & Van Tienderen, 2018; Yang et al., 2018). Studies reporting SNPs associated with agronomic traits

in the sugarcane genome are still scarce. For instance, only 229 SNPs (that explained > 5% phenotypic variation at the unadjusted P-value cutoff of 0.05) have been reported for 97 clones in a breeding population (Fickett et al., 2019) and 191 SNPs (Bonferroni corrected P-value cutoff of 0.05) for 308 accessions in a germplasm collection (Yang et al., 2018).

Based on the low number of SNPs detected at 5% of significance threshold in previous work, and also because we used a low coverage capture-seq data from our breeding population (containing 508 sugarcane genotypes), we performed RHM analysis to investigate regions in the sugarcane genome related to agronomic traits. These sugarcane clones showed comparable phenotypic data, as observed in other studies (Racedo et al., 2016; Yang et al., 2018; Fickett et al., 2019). Their average values of FB, PC, TPH, and TSH were 11.09, 13.81, 14.83, and 107.96, respectively. Comparable TPH of 17.12 t/ha was observed in Reunion Island (Gouy et al., 2014), 9.22 t/ha in Tucumán (Argentina) (Racedo et al., 2016), and 8.84 t/ha in Louisiana (United States) (Fickett et al., 2019).

The high and positive pairwise correlation observed for TPH with TSH and PC is expected because TPH is estimated as a product of these two variables. On the other hand, the negative pairwise correlation between FB and PC is in agreement with the knowledge about carbon partition and metabolism in sugarcane (Hoang et al., 2017). The estimates of broad-sense heritability ranged from 0.56 to 0.71, similar to the range observed for the same traits in other studies with sugarcane (Gouy et al., 2014; Racedo et al., 2016; Fickett et al., 2019).

The RHM analysis appears to be an interesting approach to screening the sugarcane genome aiming to identify regions related to the agronomic traits. We argued that the regions explaining a higher portion of the genomic heritability could be further explored for molecular marker investigation. Unfortunately, in the present study, none of the analyzed genome windows were significantly associated with the traits considered at the suggestive significance threshold of 3.94 (Figure S2), which could be due to the lower sequencing coverage, heterogeneity of regions covered by the sequencing, and the polyploidy effect. To overcome these limitations, we considered a new strategy to infer about the regions of the genome related to the traits under study: the selection of windows showing high heritability (h^2_r) and proceeded with the analysis of their genic content to identify candidate regions to develop molecular markers. The analyzed traits showed regions with higher heritability, explaining 15% or 20% of h^2_G in almost all sugarcane chromosomes. The number of windows above these thresholds was lower for FB (17 windows > 15%; 1,615 genes) and PC (16 windows > 15%; 1,517 genes) when compared to TPH (65 windows > 20%; 5,654 genes) and TSH (73 windows > 20%; 6,050 genes) (Figures 2 and 3). These differences could be due to the lower complexity of FB and PC, which might be controlled by a smaller number of genes. The selection of top-5 windows with higher heritability was enough to reduce the complexity of the sugarcane genome and to provide insights about regions containing a set of genes possibly related to FB (629 genes), PC (452 genes), TPH (705 genes) and TSH (413 genes) traits.

Analysis of biological processes attributed to a function of genes located in the windows with higher heritability for FB (629 genes) indicated GO terms such as “cell wall organization” (GO: 0071555), “flavonoid biosynthetic process” (GO: 0009813), and “lignin catabolic process” (GO: 0046274) (Figure 4). Some of these genes encode enzymes that catalyze the biosynthesis of secondary metabolites such as flavonoids, phenylpropanoids, and lignin. Among these enzymes are shikimate O-hydroxycinnamoyltransferase (window w268; gene Sh06_g013010), scopoletin glucosyltransferase (w359/w360; Sh09_g011980), and flavonoid

3'-monooxygenase (w360; Sh09_g012740). Genes related to "carbohydrate metabolic process" (GO:0005975), "carbon utilization" (GO:0015976), and "sucrose metabolic process" (GO:0005985) are also included in genomic regions with higher heritability for FB. These genes encode enzymes such as hexokinase (w268; Sh06_g013660) and malate dehydrogenase (w353; Sh09_g008010).

Genes related to "carbohydrate metabolic process" (GO:0005975) and "photosynthesis" (GO:0015979) are among those included in the windows with higher heritability for PC (452 genes). These genes encode enzymes of glycolysis/gluconeogenesis, pentose phosphate, oxidative phosphorylation pathways, such as fructose-bisphosphate aldolase (w317; Sh08_g004080), glucose-6-phosphate 1-epimerase (2 copies in w268), phosphoglycerate mutase (w203; Sh04_g016390), succinyl-CoA synthetase (2 copies in w202), V-type proton ATPase subunit E (w144; Sh03_g016860) and vacuolar ATP synthase 16kDa proteolipid subunit (w317; Sh08_g003930). Genes involved in phenylpropanoid biosynthesis pathways, such as phenylalanine ammonia-lyase (PAL) (3 copies in w202), and phenylalanine/tyrosine ammonia-lyase (PTAL) (w202; Sh04_g015380) are also included in these windows.

Windows with higher heritability for TPH (705 genes) contain genes which are related to "carbohydrate metabolic process", "sucrose metabolic process" (GO:0005985), "cell wall organization" (GO:0071555) and "cell redox homeostasis" (GO:0045454). Pyruvate kinase (w020; Sh01_g011250), aconitate hydratase (w063; Sh01_g042590), vacuolar ATPase B subunit (w394/w395; Sh10_g012040) and beta-amylase (w063; Sh01_g042110) are among the enzymes encoded by these genes, which are part of the carbon, starch, and sucrose metabolism pathways. Also, two genes that encode the enzymes 4-coumarate-CoA ligase (w063; Sh01_g041030) and coniferyl-aldehyde dehydrogenase (w395; Sh10_g012300), which are part of phenylpropanoid biosynthesis pathways, are also included in these high heritability windows.

Among the genes located in windows with higher heritability for TSH (413 genes), some are related to "carbohydrate metabolic process" (GO:0005975), "cell wall organization" (GO:0071555), "sucrose metabolic process" (GO:0005985), "photorespiration" (GO:0009853), and "ethylene-activated signaling pathway" (GO:0009873). The genes encode the enzymes ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (w036; Sh01_g020380), alanine transaminase (w036; Sh01_g020340), trehalose 6-phosphate phosphatase (w212; Sh04_g021630), alcohol dehydrogenase (w036; Sh01_g020520), aldose 1-epimerase (w036; Sh01_g020600), and ethylene receptor (EIN4) (w244; Sh06_g001310). In contrast to the other traits, none of the windows selected for TSH contain genes related to phenylpropanoid pathways.

Taken together, 14 windows among those top-higher heritabilities for each trait analyzed contain 35 genes that encode enzymes that catalyze biochemical reactions of carbon metabolism, starch and sucrose metabolism, and phenylpropanoid biosynthesis pathways (Table 2). Carbon partitioning is a critical process by which plants distribute the energy of photosynthesis and convert the assimilated carbon into sugar or its derivatives (Wang et al., 2013). Most plants store carbon as starch or cellulose (insoluble) with a lower concentration of sucrose (soluble), while sugarcane can store high concentrations of sucrose on its stems (Wang et al., 2013). Sucrose is cleaved into fructose and UDP-Glu, which is a nucleotide sugar precursor for most cell wall polysaccharides (Verbančič et al., 2018). Sugarcane maintains a dynamic balance of degradation of sucrose for respiration or its re-synthesis for storage. During this cycle, the carbon can be partitioned into other metabolites or fixed in

polymers that can either be remobilized (such as starch in plastids) or added to structural biomass (such as cellulose, hemicelluloses, and lignin) (Wang et al., 2013). In this balance, the enzymes sucrose synthase (SuSy), sucrose phosphate synthase (SPS), sucrose phosphate phosphatase (SPP), and invertase play a central role in sucrose metabolism. At the same time, cellulose synthesis is catalyzed by enzyme complexes of cellulose synthase (CesA) (Stein & Granot, 2019). Cell wall biosynthesis can reduce sucrose accumulation since carbon fluxes directed to plant growth, and cell wall expansion may alter carbon partitioning into sucrose (Papini-Terzi et al., 2009). It is also possible that sucrose accumulation may trigger increased lignification (Papini-Terzi et al., 2009).

The STP assembly of the sugarcane genome has multiple copies of SuSy (11 copies), SPS (6 copies), SPP (2 copies), invertase (10 copies), and CesA (37 copies) annotated on its sequence. None of them is located in the top windows with higher heritability for the analyzed traits. However, window w194 (which explains 15% of h^2_G for FB and 32.2% for TSH) contains a gene that encodes a copy of invertase (Sh04_g011120). Among the enzymes encoded by genes located in the top windows with higher heritability, phenylalanine ammonia-lyase (PAL) (Sh04_g015390, Sh04_g015400, and Sh04_g015410; w202 which explains 18% for PC) stands out as a critical enzyme involved in the phenylpropanoid pathway and biosynthesis of lignin (Zhang & Liu, 2015), which is also related to sucrose content (Papini-Terzi et al., 2009).

5. Conclusions

Throughout the analyses performed here, RHM has shown to be a useful approach to identify regions in the sugarcane genome related to agronomic traits. Even with the complexity of the sugarcane genome and its polyploidy impacting the identification of regions containing SNPs significantly associated with the phenotypes analyzed, the selection of windows that explained higher proportions of genomic heritability allows us to identify genomic regions containing a set of genes that are related to them. Among the selected windows, we identified a set of genes that encode enzymes that integrate metabolic pathways directly related to the traits analyzed. The selection of windows with higher heritability, therefore, represents an alternative for sugarcane breeders to reduce the complexity of the sugarcane genome since the selected windows span a region of 58.38Mb, which corresponds to 14,28% of the STP assembly of sugarcane genome. These windows correspond to promising genomic regions for the development of gene panels aiming the practice of marker-assisted selection of traits such as percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH) and tonnes of stalks per hectare (TSH). The findings obtained in this study will contribute to the progress of the genetic improvement of sugarcane.

6. Declarations

Acknowledgments: The authors would like to thank the Núcleo de Análise de Biomoléculas (NuBioMol) and Sugarcane Breeding Program, Universidade Federal de Viçosa, MG, Brazil for providing the facilities necessary for the execution of the experiments.

Funding: This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e

Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Author's contributions: PMPV, MM, GM, and LAP analyzed data and wrote the manuscript; MHPB and LAP planned and designed the research. PMAC conducted the experiments. MHPB and LAP coordinated the research. All authors reviewed and approved the manuscript.

Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest.

7. References

Aitken KSK, McNeil MD, Berkman PJ, Hermann S, Kilian A, Bundock PCP, Li JJ. 2014. Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. *BMC Plant Biology* 14:190. DOI: 10.1186/s12870-014-0190-x.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Baffa DCF, de A. Costa PM, da Silveira G, Lopes FJF, Barbosa MHP, Loureiro ME, Cruz CD, Peternelli LA. 2014. Path Analysis for Selection of Saccharification-Efficient Sugarcane Genotypes through Agronomic Traits. *Agronomy Journal* 106:1643–1650. DOI: 10.2134/agronj13.0576.

Balsalobre TWA, da Silva Pereira G, Margarido GRA, Gazaffi R, Barreto FZ, Anoni CO, Cardoso-Silva CB, Costa EA, Mancini MC, Hoffmann HP, de Souza AP, Garcia AAF, Carneiro MS. 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18:72. DOI: 10.1186/s12864-016-3383-x.

Barbosa MHP, Resende MDV, Dias LA dos S, Barbosa GV de S, de Oliveira RA, Peternelli LA, Daros E. 2012. Genetic improvement of sugar cane for bioenergy: the brazilian experience in network research with RIDESA. *Crop Breeding and Applied Biotechnology* 12:87–98. DOI: 10.1590/S1984-70332012000500010.

Barbosa MHP, de Silveira LCI, de Oliveira MW, de Souza V de FM, Ribeiro SNN. 2001. RB867515 Sugarcane cultivar. *Crop Breeding and Applied Biotechnology* 1:437–438. DOI: 10.13082/1984-7033.v01n04a10.

Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. 2014. Jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics* 15:293. DOI: 10.1186/1471-2105-15-293.

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

Browning BL, Zhou Y, Browning SR. 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics* 103:338–348. DOI: 10.1016/j.ajhg.2018.07.015.

Canela-Xandri O, Law A, Gray A, Woolliams JA, Tenesa A. 2015. A new tool called DISSECT for analysing large genomic data sets using a Big Data approach. *Nature Communications* 6:10162. DOI: 10.1038/ncomms10162.

Castro RD, Peternelli LA, Resende MDV, Marinho CD, Costa PMA, Barbosa MHP, Moreira EFA. 2016. Selection between and within full-sib sugarcane families using the modified BLUPIS method (BLUPISM). *Genetics and Molecular Research* 15:gmr.15017334.

548 DOI: 10.4238/gmr.15017334.
549 D'Hont A, Glaszmann JC, Hogarth DM, Int Soc Sugar Cane T. 2001. Sugarcane genome
550 analysis with molecular markers: a first decade of research. In: Hogarth DM ed.
551 *Proceedings of the International Society of Sugar Cane Technologists*. Brisbane,
552 Australia.: Australian Soc Sugar Cane Technologists, 556–559.
553 Federer WT, Raghavarao D. 1975. On Augmented Designs. *Biometrics* 31:29–35. DOI:
554 10.2307/2529707.
555 Fickett N, Gutierrez A, Verma M, Pontif M, Hale A, Kimbeng C, Baisakh N. 2019. Genome-
556 wide association mapping identifies markers associated with cane yield components and
557 sucrose traits in the Louisiana sugarcane core collection. *Genomics*. DOI:
558 10.1016/j.ygeno.2018.12.002.
559 Figueira TR e S, Okura V, Rodrigues da Silva F, Jose da Silva M, Kudrna D, Ammiraju JSS,
560 Talag J, Wing R, Arruda P. 2012. A BAC library of the SP80-3280 sugarcane variety
561 (saccharum sp.) and its inferred microsynteny with the sorghum genome. *BMC research*
562 *notes* 5:185. DOI: 10.1186/1756-0500-5-185.
563 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.
564 *arXiv*:arXiv:1207.3907.
565 Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G,
566 Charron C, Hervouet C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y,
567 Sreedasyam A, Kilian A, Chan A, Van Sluys MA, Swaminathan K, Town C, Bergès H,
568 Simmons B, Glaszmann JC, Van Der Vossen E, Henry R, Schmutz J, D'Hont A. 2018. A
569 mosaic monoploid reference sequence for the highly complex genome of sugarcane.
570 *Nature Communications* 9:2638. DOI: 10.1038/s41467-018-05051-5.
571 Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon
572 M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining
573 with the Blast2GO suite. *Nucleic Acids Research* 36:3420–3435. DOI:
574 10.1093/nar/gkn176.
575 Gouy M, Rousselle Y, Thong Chane A, Anglade A, Royaert S, Nibouche S, Costet L. 2014.
576 Genome wide association mapping of agro-morphological and disease resistance traits in
577 sugarcane. *Euphytica* 202:269–284. DOI: 10.1007/s10681-014-1294-y.
578 Hoang N V., Furtado A, Donnan L, Keeffe EC, Botha FC, Henry RJ. 2017. High-Throughput
579 Profiling of the Fiber and Sugar Composition of Sugarcane Biomass. *Bioenergy*
580 *Research* 10:400–416. DOI: 10.1007/s12155-016-9801-8.
581 Legendre BL. 1992. The core/press method for predicting the sugar yield from cane for use in
582 cane payment. *Sugar Journal* 54:2–7.
583 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
584 transform. *Bioinformatics* 25:1754–1760. DOI: 10.1093/bioinformatics/btp324.
585 Lu YH, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC. 1994. Molecular diversity
586 and genome structure in modern sugarcane varieties. *Euphytica* 78:217–226. DOI:
587 10.1007/BF00027520.
588 Meirmans PG, Liu S, Van Tienderen PH. 2018. The Analysis of Polyploid Genetic Data.
589 *Journal of Heredity* 109:283–296. DOI: 10.1093/jhered/esy006.
590 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: An automatic
591 genome annotation and pathway reconstruction server. *Nucleic Acids Research*
592 35:W182-5. DOI: 10.1093/nar/gkm321.
593 Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, Campbell H, Wilson
594 J, Wild S, Hicks AA, Pramstaller PP, Hastie N, Wright AF, Haley CS. 2012. Localising
595 Loci underlying Complex Trait Variation Using Regional Genomic Relationship
596 Mapping. *PLoS ONE* 7:e46501. DOI: 10.1371/journal.pone.0046501.
597 Neves LG, Davis JM, Barbazuk WB, Kirst M. 2013. Whole-exome targeted sequencing of the

598 uncharacterized pine genome. *The Plant Journal* 75:146–156. DOI: 10.1111/tpj.12193.
599 Nishiyama MY, Ferreira SS, Tang P-Z, Becker S, Pörtner-Taliana A, Souza GM, Souza GM.
600 2014. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid
601 and ancestor genotypes. *PloS one* 9:e107351. DOI: 10.1371/journal.pone.0107351.
602 Papini-Terzi FS, Rocha FR, Vêncio RZN, Felix JM, Branco DS, Waclawovsky AJ, Del Bem
603 LEV, Lembke CG, Costa MDL, Nishiyama MY, Vicentini R, Vincentz MGA, Ulian EC,
604 Menossi M, Souza GM. 2009. Sugarcane genes associated with sucrose content. *BMC*
605 *Genomics*. DOI: 10.1186/1471-2164-10-120.
606 Peternelli LA, Bernardes DP, Brasileiro BP, Barbosa MHP, Silva RHT. 2018. Decision Trees
607 as a Tool to Select Sugarcane Families. *American Journal of Plant Sciences*. DOI:
608 10.4236/ajps.2018.92018.
609 Piperidis G, Piperidis N, D'Hont A. 2010. Molecular cytogenetic investigation of
610 chromosome composition and transmission in sugarcane. *Molecular Genetics and*
611 *Genomics* 284:65–73. DOI: 10.1007/s00438-010-0546-3.
612 Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC. 2008. Analysis of genome-
613 wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical and Applied*
614 *Genetics* 116. DOI: 10.1007/s00122-007-0703-1.
615 Racedo J, Gutiérrez L, Perera MF, Ostengo S, Pardo EM, Cuenya MI, Welin B, Castagnaro
616 AP. 2016. Genome-wide association mapping of quantitative traits in a breeding
617 population of sugarcane. *BMC Plant Biology* 16:142. DOI: 10.1186/s12870-016-0829-x.
618 Resende MDV de. 2016. Software Selegen-REML/BLUP: a useful tool for plant breeding.
619 *Crop Breeding and Applied Biotechnology* 16:330–339. DOI: 10.1590/1984-
620 70332016v16n4a49.
621 Resende RT, de Resende MD V., Azevedo CF, Silva FF e., Melo LC, Pereira HS, Souza
622 TLPO, Valdisser PAMR, Brondani C, Vianello RP. 2018. Genome-wide association and
623 Regional Heritability Mapping of plant architecture, lodging and productivity in
624 phaseolus vulgaris. *G3: Genes, Genomes, Genetics* 8:2841–2854. DOI:
625 10.1534/g3.118.200493.
626 Schneider F. 1979. *Sugar Analysis: Official and Tentative Methods Recommended by the*
627 *International Commission for Uniform Methods of Sugar Analysis (ICUMSA)*. Hyperion
628 Books.
629 Shirali M, Pong-Wong R, Navarro P, Knott S, Hayward C, Vitart V, Rudan I, Campbell H,
630 Hastie ND, Wright AF, Haley CS. 2016. Regional heritability mapping method helps
631 explain missing heritability of blood lipid traits in isolated populations. *Heredity*
632 116:333–338. DOI: 10.1038/hdy.2015.107.
633 Song J, Yang X, Resende MFR, Neves LG, Todd J, Zhang J, Comstock JC, Wang J. 2016.
634 Natural allelic variations in highly polyploidy Saccharum complex. *Frontiers in Plant*
635 *Science* 7:804. DOI: 10.3389/fpls.2016.00804.
636 Stein O, Granot D. 2019. An overview of sucrose synthases in plants. *Frontiers in Plant*
637 *Science* 10:95. DOI: 10.3389/fpls.2019.00095.
638 Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. Revigo summarizes and visualizes long lists
639 of gene ontology terms. *PLoS ONE* 6:e21800. DOI: 10.1371/journal.pone.0021800.
640 Tanimoto T. 1964. The press method of cane analysis. *Hawaiian Planters' Record* 57:133–
641 150.
642 Thirugnanasambandam PP, Hoang N V., Henry RJ. 2018. The Challenge of Analyzing the
643 Sugarcane Genome. *Frontiers in Plant Science*. DOI: 10.3389/fpls.2018.00616.
644 Verbančič J, Lunn JE, Stitt M, Persson S. 2018. Carbon Supply and the Regulation of Cell
645 Wall Synthesis. *Molecular Plant* 11:75–94. DOI: 10.1016/j.molp.2017.10.004.
646 Wang J, Nayak S, Koch K, Ming R. 2013. Carbon partitioning in sugarcane (Saccharum
647 species). *Frontiers in Plant Science* 4:201. DOI: 10.3389/fpls.2013.00201.

Yang X, Luo Z, Todd J, Sood S, Wang J. 2018. Genome-wide association study of multiple yield components in a diversity panel of polyploid sugarcane (Saccharum spp.). *bioRxiv*:387001. DOI: 10.1101/387001.

Yang X, Song J, You Q, Paudel DR, Zhang J, Wang J. 2017. Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics* 18:594. DOI: 10.1186/s12864-017-3980-3.

Yang X, Sood S, Luo Z, Todd J, Wang J. 2019. Genome-wide association studies identified resistance loci to orange rust and yellow leaf virus diseases in sugarcane (Saccharum spp.). *Phytopathology* 109:623–631. DOI: 10.1094/PHYTO-08-18-0282-R.

Zhang X, Liu CJ. 2015. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Molecular Plant* 8:17–27. DOI: 10.1016/j.molp.2014.11.001.

Figures

Figure 1. Overview of genotypic values (overall mean + BLUP) of traits evaluated in the sugarcane breeding population. Five hundred eight sugarcane clones were evaluated for percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare (TSH). σ^2_u: genotypic variance. H^2: broad-sense heritability. ***: P-value < 0.01.

Figure 2. Regional heritability mapping (RHM) analysis of the sugarcane genome. The tilemap of RHM distribution along the genome of 508 sugarcane clones highlights the genomic regions with higher heritability for the percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare (TSH). The ten chromosomes sugarcane genome (Sh01 to Sh10) were concatenated and split into windows with 2 Mb length. The color scale on the top-right side of the plot shows the magnitude of h^2_r for each window.

Figure 3. Gene content of sugarcane genomic regions with higher heritability for FB, PC, TSH, and TPH traits in regional heritability mapping analysis. A) Gene counting of all regions with higher heritabilities. All genes are listed in Table S3. **B)** Gene counting of top-5 windows with higher heritability. All genes are listed in Table S4. **C)** Detailed information about the top-5 windows. Window 297 is the only one shared by two traits and is marked in bold and underlined. Window 310 spans two chromosomes, comprising a region of 2 Mb, which begins in the last window of Sh07 and ends in the first window Sh08, and is marked with an asterisk.

Figure 4. Functional analysis of the genes identified in of top-5 windows of sugarcane genome with higher heritability for FB, PC, TSH, and TPH traits in regional heritability mapping analysis.

Tables

Table 1. Summary of genotyping of analyzed sugarcane genotypes. Capture sequencing data of 508 sugarcane genotypes were mapped to the reference genome, and variant calling analysis was performed with the SNP quality control process.

Table 2. Genes located at the top-5 windows with higher heritability and that are related to carbon metabolism, starch, and sucrose metabolism, and phenylpropanoid biosynthesis pathways. Genes that were assigned to the KEGG ortholog groups (KOs) and mapped to the KEGG pathways of carbon metabolism (CM), glycolysis and gluconeogenesis (GG), sucrose, and starch metabolism (SSM), phenylpropanoid biosynthesis (PB), and oxidative phosphorylation (OP).

Supplementary Materials

Supplementary Material 1. Tables:

Table S1. Adjusted phenotypic observations (genotypic values) of evaluated sugarcane clones. This population consisted of 508 clones from 100 half-sib families, originated from crossings between elite sugarcane clones and commercial cultivars. Sugarcane clones were evaluated for the percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare (TSH).

Table S2. Estimates of variance components and genetic parameters for percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare (TSH). σ^2_u: genotypic variance. σ^2_b: variance between blocks. σ^2_e: residual variance. σ^2_f: individual phenotypic variance. H^2: broad-sense heritability.

Table S3. Windows with higher heritability in the sugarcane genome for analyzed traits. h^2_r: window heritability. h^2_G: genomic heritability of the analyzed trait. h^2_r / h^2_G: the proportion of genomic heritability explained by a window. Analyzed traits: percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of stalks per hectare (TSH), and tonnes of pol per hectare (TPH).

Table S4. Functional annotation of genes included in top-5 windows with higher heritability in sugarcane genome for percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH) and tonnes of stalks per hectare (TSH).

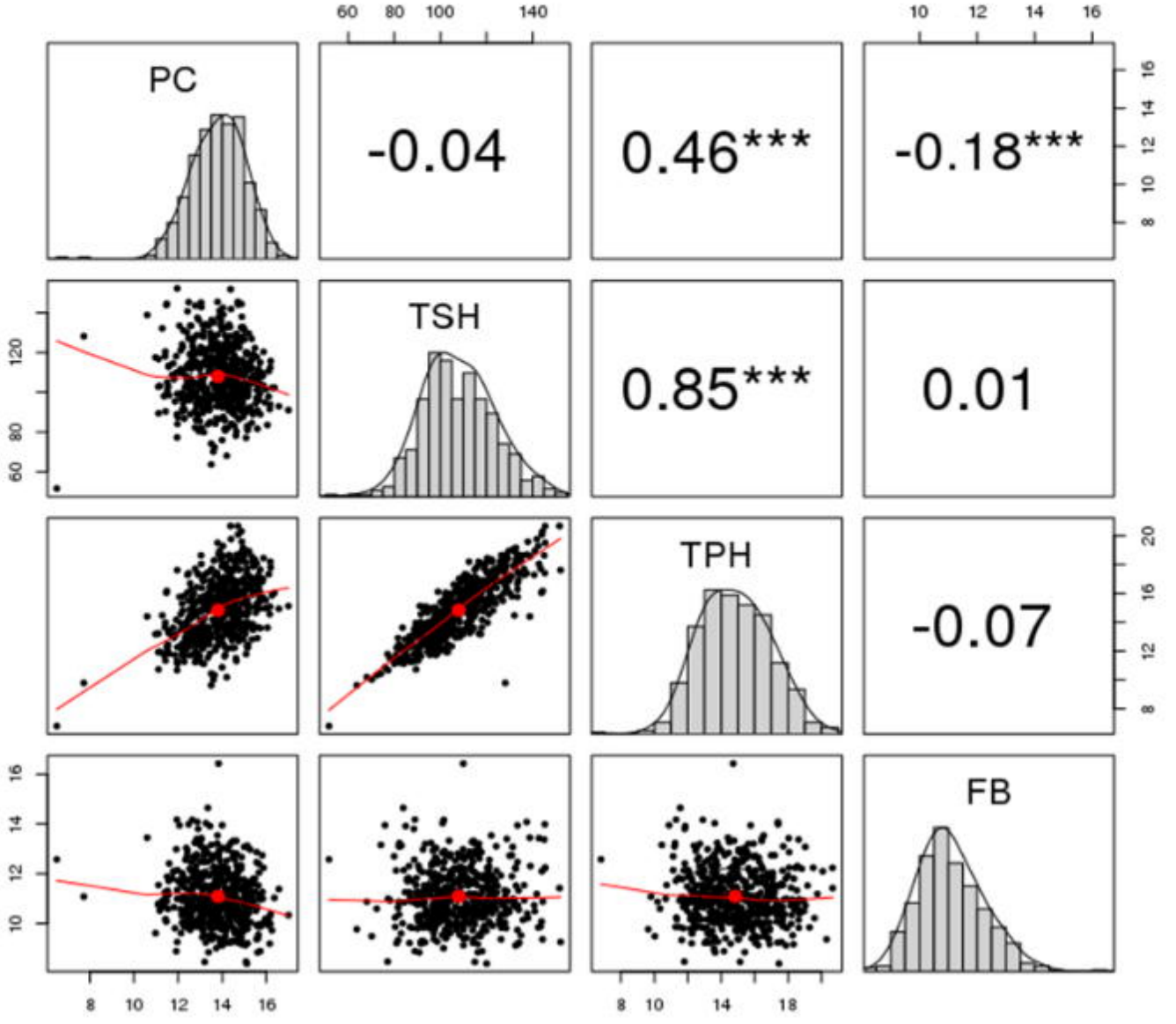
Table S5. Gene Ontology (GO) terms lists summarized by REVIGO. Non-redundant lists of GO terms assigned to the genes included in top-5 windows with higher heritability for percentage of fiber in sugarcane bagasse (FB), apparent percentage of sucrose in sugarcane (PC), tonnes of pol per hectare (TPH) and tonnes of stalks per hectare (TSH).

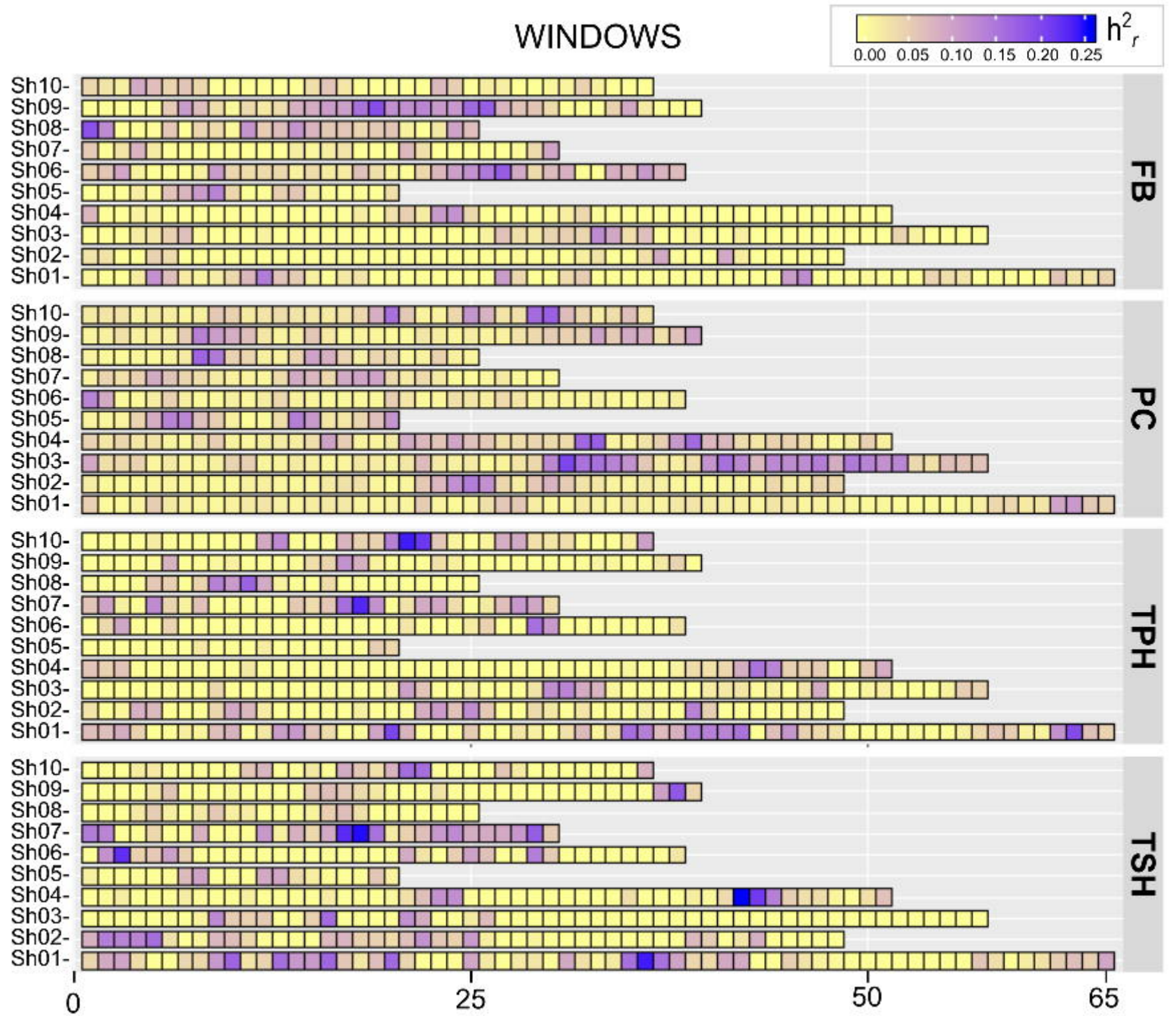
Supplementary Material 2. Figures:

Figure S1. Schematic overview of the genotyping of sugarcane clones performed in this study. The raw reads were processed, mapped to the sugarcane reference genome, and a variant calling for SNPs was performed. The software that was used for each step and their respective versions are indicated on the boxes. The command-lines used and the parameters which were considered are also shown.

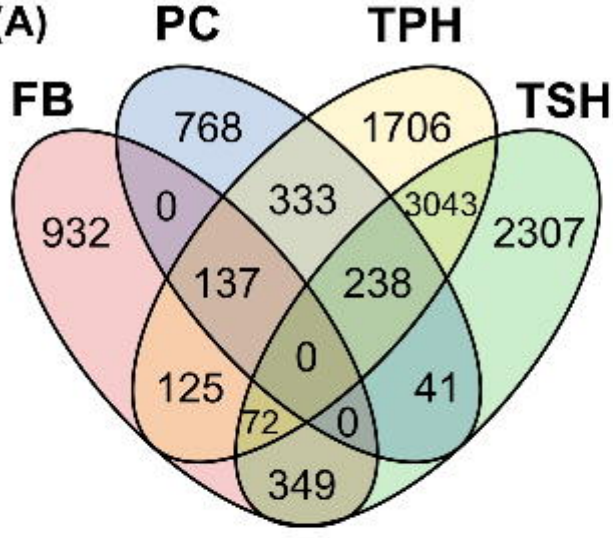
748 **Figure S2. Significance analysis of RHM window-trait associations along the genome of**
749 **analyzed sugarcane clones. (A).** Circular Manhattan plots ($-\log_{10} P$) of RHM window-trait
750 association for the percentage of fiber in sugarcane bagasse (FB), apparent percentage of
751 sucrose in sugarcane (PC), tonnes of pol per hectare (TPH), and tonnes of stalks per hectare
752 (TSH). **(B).** Quantile-quantile (QQ) plot of the data shown in the circular Manhattan plots.

Trait	Abv.	σ^2_u	H^2
Percentage of fiber in sugarcane bagasse	FB	2.21	0.56
Apparent percentage of sucrose in sugarcane	PC	2.02	0.71
Tonnes of pol per hectare	TPH	7.48	0.58
Tonnes of stalks per hectare	TSH	389.98	0.57

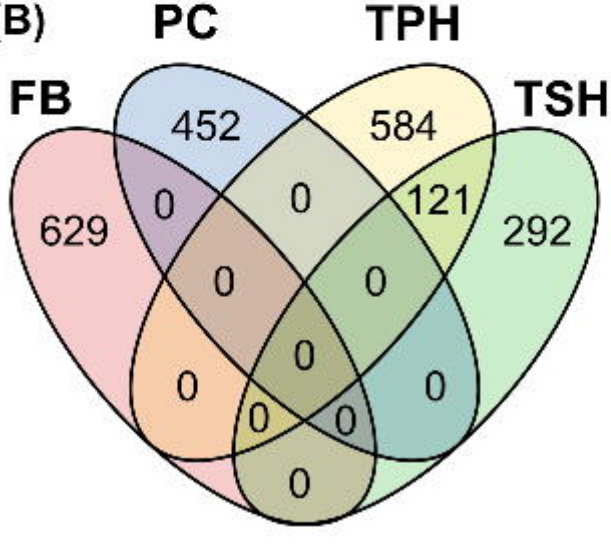



WINDOWS
h^2_r
0.00 0.05 0.10 0.15 0.20 0.25
Sh10-
Sh09-
Sh08-
Sh07-
Sh06-
Sh05-
Sh04-
Sh03-
Sh02-
Sh01-
FB
PC
TPH
TSH
0
25
50
65

(A)



(B)



(C)

Window	Chr	Start	End	h^2_r	h^2_r/h^2_G	Genes	
w268	Sh06	23227889	26181300	0.180	0.225	193	FB
w353	Sh09	15928694	18340888	0.196	0.246	110	
w359	Sh09	22026041	24621652	0.166	0.208	149	
w360	Sh09	23221276	25294245	0.178	0.223	117	
w310*	Sh07	29293193	30096520	0.192	0.241	51	
w310*	Sh08	1	1012425	0.192	0.241	78	

Window	Chr	Start	End	h^2_r	h^2_r/h^2_G	Genes	
w144	Sh03	29786348	31787511	0.205	0.220	102	PC
w202	Sh04	26497412	27954338	0.168	0.180	75	
w203	Sh04	27214734	28914864	0.177	0.190	97	
w317	Sh08	5810517	7836129	0.170	0.183	109	
w403	Sh10	29066364	30862476	0.176	0.189	103	

Window	Chr	Start	End	h^2_r	h^2_r/h^2_G	Genes	
w36	Sh01	33316393	35318896	0.244	0.638	65	TPH
w212	Sh04	35574173	37971629	0.258	0.674	137	
w244	Sh06	2033903	3698542	0.228	0.595	59	
w296	Sh07	13845897	15811781	0.231	0.603	79	
w297	Sh07	14688970	17462449	0.253	0.661	121	

Window	Chr	Start	End	h^2_r	h^2_r/h^2_G	Genes	
w20	Sh01	16169530	18591486	0.206	0.556	116	TSH
w63	Sh01	66561582	70914509	0.199	0.538	347	
w297	Sh07	14688970	17462449	0.235	0.635	121	
w394	Sh10	20759174	22907829	0.244	0.660	75	
w395	Sh10	22127735	23707822	0.220	0.595	77	

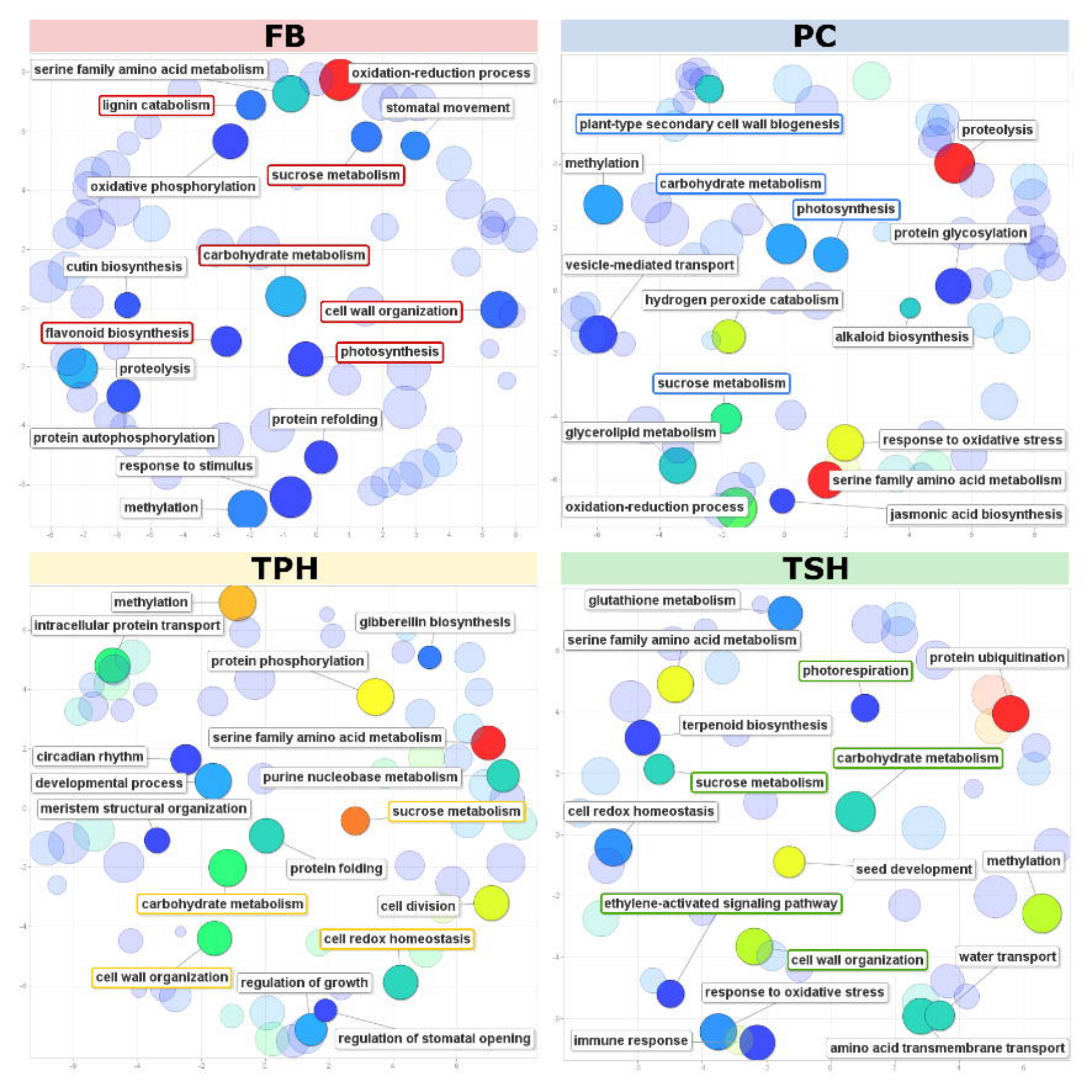

FB
serine family amino acid metabolism
oxidation-reduction process
lignin catabolism
stomatal movement
sucrose metabolism
oxidative phosphorylation
carbohydrate metabolism
cutin biosynthesis
cell wall organization
flavonoid biosynthesis
photosynthesis
proteolysis
protein refolding
protein autophosphorylation
response to stimulus
methylation
PC
plant-type secondary cell wall biogenesis
proteolysis
methylation
carbohydrate metabolism
photosynthesis
protein glycosylation
vesicle-mediated transport
hydrogen peroxide catabolism
alkaloid biosynthesis
sucrose metabolism
glycerolipid metabolism
response to oxidative stress
serine family amino acid metabolism
oxidation-reduction process
jasmonic acid biosynthesis
TPH
methylation
intracellular protein transport
gibberellin biosynthesis
protein phosphorylation
serine family amino acid metabolism
circadian rhythm
purine nucleobase metabolism
developmental process
meristem structural organization
sucrose metabolism
protein folding
carbohydrate metabolism
cell division
cell redox homeostasis
cell wall organization
regulation of growth
regulation of stomatal opening
TSH
glutathione metabolism
serine family amino acid metabolism
protein ubiquitination
photorespiration
terpenoid biosynthesis
carbohydrate metabolism
sucrose metabolism
cell redox homeostasis
seed development
methylation
ethylene-activated signaling pathway
cell wall organization
water transport
response to oxidative stress
immune response
amino acid transmembrane transport

Table 1. Summary of genotyping of analyzed sugarcane genotypes. Capture sequencing data of 508 sugarcane genotypes were mapped to the reference genome and a variant calling analysis was performed with a SNP quality control (QC) process.

Sugarcane chromosome	Length (Mb)	Number of Genes	Average of mapping coverage	Number of SNPs before QC	Number of SNPs after QC
Sh01	72.29	4,476	1.88	127,817	59,783
Sh02	52.64	3,178	1.55	95,999	44,429
Sh03	54.89	3,399	1.46	115,485	52,545
Sh04	45.58	2,791	1.54	101,102	46,756
Sh05	23.28	1,233	1.14	39,330	17,993
Sh06	35.41	2,108	1.79	74,477	34,853
Sh07	30.10	1,824	1.38	63,170	27,829
Sh08	24.32	1,307	1.25	49,150	23,001
Sh09	36.58	2,082	1.32	76,803	35,419
Sh10	34.65	1,943	1.33	71,654	32,587
Total	408.94	24,341	1.46	814,987	375,195

Table 2. Genes located at the top-5 windows with higher heritability and that are related to carbon metabolism, starch and sucrose metabolism, and phenylpropanoid biosynthesis pathways. Genes which were assigned to the KEGG ortholog groups (KOs) and mapped to the KEGG pathways of carbon metabolism (CM), glycolysis and gluconeogenesis (GG), sucrose and starch metabolism (SSM), phenylpropanoid biosynthesis (PB), and oxidative phosphorylation (OP).

Trait	Gene(window)	KO	Enzyme [EC number]	Pathway
FB	Sh09_g008010(w353)	K00029	malate dehydrogenase [1.1.1.40]	CM
FB	Sh06_g013660(w268)	K00844	hexokinase [2.7.1.1]	CM, SSM
FB	Sh06_g013010(w268)	K13065	shikimate O-hydroxycinnamoyltransferase [2.3.1.133]	PB
FB	Sh09_g011980(w359/w360)	K23260	scopoletin glucosyltransferase [2.4.1.128]	PB
PC	Sh08_g004080(w317)	K01623	fructose-bisphosphate aldolase [4.1.2.13]	CM
PC	Sh04_g016390(w203)	K01834	phosphoglycerate mutase [5.4.2.11]	CM
PC	Sh04_g015250(w202), Sh04_g015260(w202)	K01900	succinyl-CoA synthetase [6.2.1.4 6.2.1.5]	CM
PC	Sh08_g004580(w317)	K02437	glycine cleavage system H protein	CM
PC	Sh06_g014550(w268), Sh06_g014560(w268)	K01792	glucose-6-phosphate 1-epimerase [5.1.3.15]	GG
PC	Sh03_g016860(w144)	K02150	V-type H+-transporting ATPase subunit E	OP
PC	Sh08_g003930(w317)	K02155	V-type H+-transporting ATPase 16kDa proteolipid subunit	OP
PC	Sh04_g016300(w203)	K01087	trehalose 6-phosphate phosphatase [3.1.3.12]	SSM
PC	Sh10_g015790(w403), Sh10_g015810(w403)	K01187	alpha-glucosidase [3.2.1.20]	SSM
PC	Sh04_g015390(w202), Sh04_g015400 (w202), Sh04_g015410 (w202)	K10775	phenylalanine ammonia-lyase [4.3.1.24]	PB
PC	Sh04_g015380(w202)	K13064	phenylalanine/tyrosine ammonia-lyase [4.3.1.25]	PB
TPH	Sh01_g042660(w063), Sh01_g042670(w063)	K00640	serine O-acetyltransferase [2.3.1.30]	CM
TPH	Sh01_g011250(w020)	K00873	pyruvate kinase [2.7.1.40]	CM
TPH	Sh01_g042590(w063)	K01681	aconitate hydratase [4.2.1.3]	CM
TPH	Sh10_g011930(w394/w395)	K01738	cysteine synthase [2.5.1.47]	CM
TPH	Sh01_g042950(w063)	K03781	catalase [1.11.1.6]	CM
TPH	Sh10_g012040(w394/w395)	K02147	V-type H+-transporting ATPase subunit B	OP
TPH	Sh01_g042110(w063)	K01177	beta-amylase [3.2.1.2]	SSM
TPH	Sh01_g041030(w063)	K01904	4-coumarate--CoA ligase [6.2.1.12]	PB
TPH	Sh10_g012300(w395)	K12355	coniferyl-aldehyde dehydrogenase [1.2.1.68]	PB
TSH	Sh01_g020340(w036)	K00814	alanine transaminase [2.6.1.2]	CM
TSH	Sh01_g020380(w036)	K01601	ribulose-bisphosphate carboxylase large chain [4.1.1.39]	CM
TSH	Sh01_g020520(w036)	K00001	alcohol dehydrogenase [1.1.1.1]	GG
TSH	Sh01_g020600(w036)	K01785	aldose 1-epimerase [5.1.3.3]	GG
TSH	Sh04_g021630(w212)	K01087	trehalose 6-phosphate phosphatase [3.1.3.12]	SSM