**Genome-wide association studies unveils the genetic basis of cell wall composition and saccharification of cassava pulp**

Pongsakorn Sunvittayakula,b, Passorn Wonnapinija, Pitchaporn Wannitikulb, Phongnapha Phanthanongc, Kanokpoo Changwitchukarnc, Anongpat Suttangkakula, Supanut Utthiyaa, Apimon Phraemuanga,Pasajee Kongsilc,d , แพรวa,

Hernan Ceballose, Leonardo D. Gomezf, Piya Kittipadakulc,d, Supachai Vuttipongchaikija,g,h,\*

a Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand

b Department of Agriculture, Ministry of Agriculture and Cooperatives, Bangkok, Thailand

c Department of Agronomy, Faculty of Agriculture, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand

d Center for Advanced Studies of Agriculture and Food (CASAF), Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand

e The Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT), Cali, Colombia

f Centre of Novel Agricultural Products (CNAP), Department of Biology, University of York, York,

United Kingdom

g Center of Advanced Studies for Tropical Natural Resources, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand

h Omics Center for Agriculture, Bioresources, Food and Health, Kasetsart University (OmiKU), Bangkok, Thailand

**Corresponding Author: Supachai Vuttipongchaikij**

Department of Genetics, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Road, Chatuchak, Bangkok 10900, Thailand;

Email: supachai.v@ku.th

**Abstract**

Cassava (*Manihot esculenta* Crantz) is a key crop for starch and biofuels production. This study focuses on the polysaccharide composition and saccharification efficiency in cassava pulp through genome-wide association studies (GWAS), targeting the improvement of root characteristics for industrial use. We analyzed 135 partially inbred lines population, performing monosaccharide composition and saccharification analyses to reveal substantial variability in storage root biomass. Among 33 traits examined, 128 significant SNPs were associated with 23 biomass traits, highlighting a complex genetic architecture. Saccharification potential varied from 39 to 95 nmol Glu mg-1 h-1, with high broad-sense heritability for saccharification and several monosaccharide traits, indicating a strong genetic control. Our findings revealed that cassava pulp comprises similar proportions of pectin, hemicellulose, and cellulose in all genotypes. Correlation analysis showed significant associations between cellulose content and saccharification, suggesting that enhancing these traits can improve bioconversion efficiency. Negative correlations with glucose and glucuronic acid in hemicellulose and pectin fractions imply these components may inhibit saccharification. We identified 118 candidate genes associated with 21 traits, with many involved in stress responses affecting cell wall composition. This study verified 12 key candidate genes through sequence and expression analysis, including MANES\_07G081200, a YTH domain-containing protein associated with saccharification. Several stress-response genes, such as MANES\_04G118600 and MANES\_09G174600, were linked to monosaccharide traits, suggesting that adaptive stress pathways influence biomass characteristics. This study provides insights into the genetic determinants of cassava pulp's saccharification and polysaccharide composition, aiding breeding efforts to develop cassava varieties optimized for industrial applications.

Keywords: Cassava pulp; Saccharification; Genome-wide association studies (GWAS); Cell wall composition

**Introduction**

Cassava (*Manihot esculenta* Crantz) is a crucial tropical crop, ranked as the fourth most important for carbohydrate supply both as staple food and an industrial crop (Parmar et al., 2017). Its resilience in poor soil conditions, low fertilizer input, and resistance to climate change make it a key crop for food security and adaptation strategies (Jarvis et al., 2012). Industrially, cassava is a feedstock for starch production and contributes to various products, including animal feed, bioethanol, and ingredients for the pharmaceutical and cosmetic industries (Li et al., 2017). Cassava processing leads to a substantial amount of waste known as cassava pulp, which represents about 15-20% of the total biomass (Djuma’Ali et al., 2011). This pulp, which is rich in starch, cellulose, and hemicellulose (Patle and Lal 2008) poses environmental challenges as a waste, but also offers opportunities for use in bio-refinery processes. While cassava pulp holds potential for various value-added products, its utilization remains largely underexplored (Sriroth et al., 2000). Utilizing cassava pulp offers sustainability and economic benefits, especially as it is readily available from the processing mills, minimizing transportation costs.

The composition of cassava pulp harbors a significant potential for the production of value-added products through advanced bio-refinery techniques, but at present it remains underutilized (Li et al., 2017). The potential applications of cassava pulp are diverse, encompassing biofuel (Poonsrisawat et al., 2017), production of bio-chemicals (Zhang et al., 2016), biogas generation (Lerdlattaporn et al., 2021), animal feed (Dagaew et al., 2022), and bio-fertilizers (Iwai et al., 2013). This highlights the necessity of studying cassava pulp to effectively valorize this biomass for different applications. Understanding the composition of its cell walls and the saccharification efficiency, which is the process of breaking down polysaccharides into fermentable sugars, can assist in the selection of traits that unlock its use. Despite recent research efforts aimed at enhancing ethanol production from cassava pulp (Kosugi et al., 2009; Apiwatanapiwat et al., 2011; Virunanon et al., 2013; Wattanagonniyom et al., 2017), the relationship between cassava genotype, saccharification efficiency, and cell wall composition remains poorly understood. Enabling the use of cassava pulp is a critical step for maximizing the efficiency of cassava root utilization and guiding future breeding programs.

Genome-wide association studies (GWAS) offer a powerful tool for unraveling the genetic basis of complex traits in crops. By leveraging GWAS data, genetic markers associated with important agronomic traits can be identified, facilitating targeted breeding efforts for crop improvement. GWAS has played a pivotal role in identifying genes related to traits such as yield (Zhang et al., 2018; Rabbi et al., 2022; Phumichai et al., 2022; Sunvittayakul et al., 2024), resistance to diseases like cassava mosaic virus (Wolfe et al., 2016; Rabbi et al., 2020) and root rot (Brito et al., 2017), starch content in the roots, cassava starch paste properties (Santos et al., 2022; Phumichai et al., 2022), levels of carotenoids (Zargar et al., 2015), as well as the concentration of cyanide, a toxic substance, in cassava (Ogbonna et al., 2021). While GWAS has been successfully applied to various aspects of cassava research, its application to the study of the saccharification efficiency and cell wall composition in cassava pulp has not been explored.

In this study, we conducted an analysis of cassava pulp composition and saccharification potential. We utilized an inbred population derived from Thai commercial varieties and performed a monosaccharide composition analysis to characterize different cell wall fractions, including hemicellulose, pectin, and cellulose, as well as saccharification analysis. Our primary objective was to identify genomic regions significantly associated with saccharification efficiency and cell wall composition. Through this research we contribute valuable insights into the genetic determinants of the processing and utilization of cassava pulp, ultimately supporting efforts to enhance the efficiency of cassava root utilization, advancing breeding programs for improved industrial varieties.

**Materials and Methods**

**Germplasm, experimental design and cultivation conditions**

Details regarding the cassava germplasm and genotypes utilized in this study can be found in Sunvittayakul et al., (2024). In brief, the germplasm comprised six S0 progenitors, namely R1, HNT, R5, R90, HB80, and KU50. S1 and S2 partial-inbred lines were generated through controlled self-pollination, with seedlings displaying stunted growth or small size being discarded. The pre-selection of S1 and S2 lines was based on a prior single-row yield trial, covering a range of high, medium, and low root yields. The study encompassed 102 S1, 33 S2 genotypes and the six non-inbred progenitors. Planting was conducted using 20-cm cuttings in a single-row trial with spacing of 1.5 ´ 1.5 m (within and between rows). An augmented randomized complete block design (augmented RCBD) was employed, comprising nine blocks with non-replicated samples. Each block included six commercial varieties as checks. Within each row, eight stem cuttings were planted, and data collection and analysis were performed on the middle six plants. Border rows were incorporated into the experimental field layout. The experiment was conducted in central Thailand (location coordinates: 13.653699 and 99.821265), during April 2019 to March 2020 without irrigation. Fertilizer application occurred four months after planting. Roots were hand-harvested 11 months after planting, sliced (300 g per sample in fresh weight) and dried in paper bags at 65°C. The data presented in this study were collected from one year of harvest.

**Sample preparation**

The dry cassava chips were prepared by removing the cortex and then grinding through a 0.5-micron sieve using an Ultra Centrifugal Mill ZM 200 (RETSCH). The ground sample were used for starch extraction based on the modified Uppsala method (Theander et al., 1995). Briefly, 5 g of sample were treated with a-amylase solution (concentration unspecified) in 20 ml of acetate buffer overnight at 80°C with constant shaking. After cooled down to 40°C, 100 µl of Amyloglucosidase solution (280 U/ml) was added and incubated overnight at 60°C with shaking. The sample was precipitated by mixing with 40 ml of absolute ethanol and then centrifugation. The pellet was washed twice by suspension with 20 ml of 80% ethanol and centrifugation. Subsequently, the samples were freeze-dried and identified as the de-starched material.

**Saccharification analysis**

Saccharification analysis was conducted following the method by Gomez et al., (2010) using a 96-well-plate format on a robotic platform. In this assay, 10 mg of the de-starched cassava material underwent pretreatment with 350 µl of 0.5 M NaOH solution at 90°C for 30 minutes, followed by five washes with 500 µl of 25 mM sodium acetate buffer at pH 4.5. The sample was then hydrolyzed using a Celluclast cocktail from *Trichoderma reesei* and Novozyme 188 (Novozymes A/S, Bagavaerd, Denmark), mixed at a ratio of 4:1 with an enzyme loading of 22.5 Filter Paper Units per gram of sample material at 50°C for 8 hours. Hydrolysates were quantified using a reducing sugar assay based on a modified MBTH method. At least three replicates were performed for each cassava genotype, with each replicate undergoing five technical repetitions. The quantities of reducing sugars released were expressed in nanomoles based on glucose standards for calibration.

**Cell wall extraction**

De-starched materials were fractionated into hot water-CDTA extract (pectin fraction), hemicellulose fraction and cellulose residue. Ten milligrams of the de-starched material were extracted in 1 ml of distilled water at 120°C for 1 hour. The supernatant was separated by centrifugation at 14,000 rpm for 20 minutes, and the pellet was further extracted with 1 ml of 50 mM CDTA pH 6.5 at room temperature for 16 hours. The supernatant was collected by centrifugation. The pellets were then washed twice with 1.5 ml of distilled water, followed by centrifugation. The supernatants from all stages were combined and precipitated using 80% ethanol, collected by centrifugation, and subsequently dried using a speed vacuum concentrator before proceeding to monosaccharide analysis. For the hemicellulose fraction, the remaining pellets were hydrolyzed with 0.5 ml of 2M trifluoroacetic acid (TFA) at 100°C for 4 hours, and then used directly for monosaccharide analysis. The remaining pellets were kept for subsequent cellulose analysis.

**Monosaccharide composition analysis**

Hot water-CDTA extracts were subjected to hydrolysis using 2 M trifluoroacetic acid (TFA) at 100°C for 4 hours, followed by separation via high-performance anion-exchange chromatography on a Dionex Carbopac PA-20 column and pulsed amperometric detection. Monosaccharides were quantified using external calibration with an equimolar mixture of nine monosaccharide standards (arabinose, fucose, rhamnose, xylose, glucose, galactose, mannose, glucuronic acid, and galacturonic acid). The total hemicellulose or total pectin content, expressed in micrograms per milligram of biomass, were calculated by aggregating the amounts of these nine monosaccharides obtained from the NaOH and CDTA extracts, respectively. The total hemicellulose and pectin content were then combined designated as total hemicellulose and pectin. Three replicates were performed for each cassava genotype.

**Cellulose quantification**

Residual pellets from the TFA hydrolysis were subjected to cellulose analysis. The pellet was washed three times with 1.5 ml of distilled water, dried using a vacuum evaporator, and then subjected to Saeman hydrolysis (Saeman, 1945). To initiate the hydrolysis, 90 µl of 72% sulfuric acid was added and incubated for 4 hours at 25°C. The mixture was then diluted with 1890 µl of water and further incubated at 120°C for an additional 4 hours. Following incubation, the supernatant was separated by centrifugation at 12,000 rpm for 10 minutes, and the glucose content was quantified using the colorimetric Anthrone assay (Morse, 1947). Three replicates were performed for each cassava genotype.

**gDNA isolation, Genotyping-by-Sequencing and genome re-sequencing**

Cassava genomic DNA (gDNA) was extracted from young leaves using a modified CTAB method in combination with the FavorPrep™ Plant Genomic DNA Extraction Mini Kit (Favorgen Biotech Corp, Ping Tung, Taiwan). The quality of the extracted DNA was evaluated using agarose gel electrophoresis and Nanodrop spectrophotometry. Genotyping-by-sequencing (GBS) data was obtained from paired-end reads using the Illumina® HiSeq/MiSeq platform (Illumina Inc, San Diego, CA, USA), with read lengths ranging from 31 to 144 bp at each end. The sequenced reads were aligned to the reference genome (Manihot\_esculenta\_v8) using the BWA software version 0.7.17. BAM files were processed using SAMTools version 1.10 (Li et al., 2009), and variant calling was performed using Freebayes (Garrison and Marth, 2012). The average read depth across the reference genome (639.6 Mb) ranged from 2.03X to 3.06X. A total of 2,078,599 bi-allelic single nucleotide polymorphisms (SNPs) were identified, spanning the 18 nuclear chromosomes, two organelle genomes, and 25 unplaced contigs. Variant filtering was applied to retain SNPs with a minor allele frequency (MAF) of 0.01 and a SNP missing rate of 0% using bcftools (Narasimhan et al., 2016). Consequently, 29,733 bi-allelic SNP markers distributed across the 18 cassava nuclear chromosomes were selected for subsequent population structure analysis, genomic kinship estimation, and GWAS. Additionally, observed heterozygosity and polymorphic information content (PIC) were calculated for the polymorphic SNP sites. Re-sequencing was perfromed on the six progenitors using XXX and analyzed by … The sequencing data have been deposited in the NCBI database. These included re-sequencing data for S0 progenitors including KU50, R1, R90, R5, HB80 and HNT and GBS reads of partial inbred lines and the six progenitors (see File S1 for accession numbers).

**Data analysis and GWAS**

All data analysis was conducted using the R 4.2.1 software package. Pearson’s correlation analysis was performed using the PerformanceAnalytics package (Peterson et al., 2018), utilizing the correlation matrix chart function. Principal component analysis (PCA) was computed using the FactoMineR and factoextra R packages. GWAS was conducted using the mixed linear model (MLM) as well as the fixed and random effects model for circulating probability unification (FarmCPU), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) models were implemented using the GAPIT (Lipka et al., 2012). Bonferroni corrections were applied to establish the significance threshold to minimize false positives. Additionally, the genomic region linked to the statistically associated SNPs was identified using the GALLO package (Fonseca et al., 2020). The lme4 package facilitated the estimation of variance components via restricted maximum likelihood (REML) and broad-sense heritability (*H*2), calculated as *H*2  where represents genotypic variance and represents residual variance.

**Sequence analysis**

Analysis of sequence varience. Cladigram was built…

**RNA isolation and quantitative real-time PCR**

RNA was extracted from storage root samples of three cassava varieties: KU50, HNT, and HB60 (a sibling of HB80). These varieties were grown in pots for approximately three months before sample collection. For each variety, three biological replicates were used for RNA extraction. Total RNA was isolated using a CTAB method and followed up with Dnase I treatment (New England Biolabs, MA, USA). The quality and quantity of the isolated RNA were assessed using Nanodrop and 1% (w/v) agarose gel electrophoresis, respectively. For each sample, 1 µg of total RNA was used for cDNA synthesis. The cDNA was generated in a total reaction volume of 20 µl using oligo-dT primers and MMuLV Reverse transcriptase (Biotechrabbit, Berlin, Germany) following the manufacture’s protocol, and subsequently diluted to 55 µl with ultrapure water. For quantitative real-time PCR (qRT-PCR), each reaction was prepared with a total volume of 20 µl, consisting of 1 µl of cDNA, 250 nM of each primer (see Table S1), and 2X SuperReal PreMix Plus (SYBR Green) (Tiangen Biotech, Beijing, China). The qRT-PCR was performed using CFX Duet Real-Time PCR system (Bio-Rad, CA, USA) for 40 cycles, with an initial denaturation at 95°C for 15 minutes, followed by 40 cycles of 95°C for 10 seconds and 55°C for 20 seconds. Gene expression data were normalized against cassava4.1\_006884 as a reference gene based on Hu et al. (2016), and relative expression levels were calculated using the 2^-∆∆Ct method.

**Results**

**Phenotypic variability of cassava pulp saccharification and biomass traits in partially inbred populations**

We analyzed 142 cassava lines comprising 102 S1 generation lines, 33 S2 generation lines, and the six S0 progenitors (Thai elite commercial varieties). Our phenotypic analysis shows a broad range of variability in saccharification and 32 other biomass traits of cassava pulp. Details of these 33 traits are summarized in Table 1, and Figure S1 illustrates the trait distributions within the population. The saccharification potential of cassava pulp varied from 39.02 to 95.93 nmol Glu mg-1 h-1. The de-starched material showed a range from 67 to 747 mg g-1 DW, and the cellulose content ranged from 19.67 to 234.60 µg mg-1 DW. The content of nine monosaccharides in pectin or hemicellulose fractions ranged from 0.01 to 311 µg mg-1 DW, and the total cell wall matrix polysaccharides ranged from 1.08 to 178.39 µg mg-1 DW. These data suggest that the cell wall composition of cassava pulp, derived mainly from parenchyma with starch deposits and some vascular tissues, typically comprises equal amounts of cellulose, hemicellulose and pectin, although specific lines showed variability in the ratios of these components. Saccharification exhibited a wide distribution, whereas monosaccharide contents in pectin, particularly galactose, glucose, and xylose, appeared in narrower ranges. In contrast, arabinose, rhamnose, mannose, galacturonic acid, and glucuronic acid showed distributions skewed towards either higher or lower extremes (Figure S1). Monosaccharide content in hemicellulose and the cellulose content trended towards higher values with broader distributions. Notably, the de-starched material, hemicellulose, and pectin contents also trend towards higher values and greater distributions, indicating that selective breeding could effectively enhance these traits (Figure S1).

To elucidate the genetic basis of this variability, we assessed the broad-sense heritability (*H2)* of these traits (Table 1). Saccharification of cassava pulp showed high heritability (*H2* > 0.9), suggesting that it is highly stable and selectable in breeding programs. In contrast, cellulose content exhibited moderate genetic control (*H2* = 0.409). The heritability of the de-starched material amount was also high. Heritability of the monosaccharide traits varied extensively, ranging from 0.158 to 0.999, with most displaying moderate (>0.5) to remarkably high (>0.8) heritability values. However, rhamnose and mannose in hemicellulose, total sugar in pectin, and rhamnose in both hemicellulose and pectin fractions exhibited notably lower heritability values (<0.3). These findings indicate that most of these cassava pulp traits are highly heritable. These results provide a robust framework for further investigations into the genetic factors influencing cassava pulp traits.

Table 1. Biomass traits of cassava pulp analyzed in this work. Nine monosaccharides of hemicellulose and pectin fractions and combined of both fractions are presented. Saccharification is expressed in nmol Glu mg-1 h-1, and de-starched sample is expressed as mg g-1 DW. Monosaccharides traits are expressed in µg mg-1 DW. Broad sense heritability (*H2*) for each trait is shown.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Trait | Abbre | Min | Max | Mean | *H2* |
| 1 | Saccharification | Sac | 39.02 | 95.93 | 64.823 | 0.995 |
| 2 | De-starched material | DS | 67 | 747 | 181.51 | 0.632 |
| 3 | Cellulose content | Cell | 19.67 | 234.60 | 117.346 | 0.409 |
| 4 | Fucose in hemicellulose | HFuc | 0.70 | 5.50 | 2.492 | 0.929 |
| 5 | Arabinose in hemicellulose | HAra | 1.48 | 12.39 | 4.688 | 0.597 |
| 6 | Rhamnose in hemicellulose | HRha | 0.43 | 10.19 | 2.204 | 0.291 |
| 7 | Galactose in hemicellulose | HGal | 3.17 | 33.00 | 11.703 | 0.720 |
| 8 | Glucose in hemicellulose | HGlc | 5.36 | 23.41 | 11.307 | 0.796 |
| 9 | Xylose in hemicellulose | HXyl | 4.31 | 34.22 | 11.613 | 0.473 |
| 10 | Mannose in hemicellulose | HMan | 0.47 | 19.53 | 3.869 | 0.158 |
| 11 | Galacturonic acid in hemicellulose | HGalA | 1.37 | 26.20 | 5.722 | 0.816 |
| 12 | Glucuronic acid in hemicellulose | HGlcA | 0.08 | 5.22 | 0.747 | 0.374 |
| 13 | Total sugar in hemicellulose | THem | 28.50 | 159.29 | 54.841 | 0.327 |
| 14 | Fucose in Pectin | PFuc | 0.33 | 6.14 | 2.258 | 0.587 |
| 15 | Arabinose in Pectin | PAra | 1.78 | 22.31 | 5.583 | 0.726 |
| 16 | Rhamnose in Pectin | PRha | 0.56 | 9.15 | 4.115 | 0.507 |
| 17 | Galactose in Pectin | PGal | 1.13 | 76.55 | 13.721 | 0.983 |
| 18 | Glucose in Pectin | PGlc | 3.50 | 311.79 | 27.175 | 0.999 |
| 19 | Xylose in Pectin | PXyl | 0.23 | 14.61 | 1.792 | 0.367 |
| 20 | Mannose in Pectin | PMan | 0.30 | 9.08 | 3.128 | 0.502 |
| 21 | Galacturonic acid in Pectin | PGalA | 0.02 | 55.52 | 16.430 | 0.573 |
| 22 | Glucuronic acid in Pectin | PGlcA | 0.01 | 12.23 | 2.637 | 0.769 |
| 23 | Total sugar in Pectin | TPec | 18.50 | 375.40 | 77.117 | 0.245 |
| 24 | Fucose in hemicellulose and Pectin | THPFuc | 1.08 | 8.88 | 4.804 | 0.422 |
| 25 | Arabinose in hemicellulose and Pectin | THPAra | 3.50 | 20.91 | 10.299 | 0.762 |
| 26 | Rhamnose in hemicellulose and Pectin | THPRha | 1.63 | 15.96 | 6.404 | 0.291 |
| 27 | Galactose in hemicellulose and Pectin | THPGal | 8.86 | 51.68 | 25.500 | 0.720 |
| 28 | Glucose in hemicellulose and Pectin | THPGlc | 13.87 | 178.39 | 37.954 | 0.959 |
| 29 | Xylose in hemicellulose and Pectin | THPXyl | 4.80 | 34.88 | 13.417 | 0.473 |
| 30 | Mannose hemicellulose and Pectin | THPMan | 1.52 | 29.65 | 7.240 | 0.691 |
| 31 | Galacturonic acid in hemicellulose and Pectin | THPGalA | 2.97 | 83.79 | 21.861 | 0.763 |
| 32 | Glucuronic acid in hemicellulose and Pectin | THPGlcA | 0.13 | 24.55 | 3.924 | 0.506 |
| 33 | Total sugar in hemicellulose and Pectin | THP | 59.20 | 267.80 | 131.380 | 0.924 |

**Correlation analysis of biomass traits in cassava pulp**

To investigate the relationships among various biomass traits in cassava pulp, we performed a Pearson's correlation analysis (Figure 1). We found that cellulose content, saccharification, and de-starched material are positively and strongly correlated. Interestingly, cellulose and saccharification show negative correlations with glucose in the hemicellulose and pectin fractions, and with glucuronic acid in the hemicellulose fraction, suggesting that the saccharification potential in cassava largely depends on the cellulose content across different varieties. The de-starched material not only correlates positively with most monosaccharide contents in hemicellulose, excluding glucose and glucuronic acid, but also shows a strong correlation with saccharification, indicating a compositional influence of these matrix polysaccharides on the saccharification potential. Whereas the negative correlation between de-starched material and glucose reflects the proportion of glucose derived from cellulose or other cell wall fractions.

Within the hemicellulose fraction, strong positive correlations (r2 ≥ 0.6) were observed among most monosaccharides, including fucose, arabinose, rhamnose, galactose, xylose, mannose, and galacturonic acid, with the exception of glucose, which consistently exhibited negative correlations with other sugars. Glucuronic acid showed strong negative correlations with arabinose and cellulose and modest negative correlations with all other monosaccharides. In the pectin fraction, most monosaccharides displayed weak positive or negative correlations among themselves, except for arabinose, rhamnose, galactose, and mannose, which showed strong positive correlations. Additionally, cellulose content showed a strong positive correlation with arabinose and rhamnose contents in the pectin fraction but a negative correlation with glucose and glucuronic acid contents, highlighting complex interactions within the cell wall matrix. Moreover, a notable correlation was observed between the total sugar in pectin and glucose in pectin, suggesting that higher pectin content in cassava pulp is associated with higher glucose content. This correlation might indicate shared biochemical pathways or regulatory mechanisms, which could be targeted in breeding programs to enhance biomass utilization, particularly for biofuel production. These findings underscore the interplay among different cell wall components and their collective impact on cassava pulp composition.

**A diagram of a gene analysis

Description automatically generated with medium confidence**

Figure 1. Pearson’s correlation analysis of 33 biomass traits in cassava pulp. Positive and negative correlations are represented in blue and orange, respectively, at significant level (P<0.05). The size of the circle follows the level of correlation. The trait abbreviations are indicated in Table 2.

**Significant SNPs associated with saccharification and monosaccharide composition traits in cassava pulp**

We conducted a GWAS on saccharification, cellulose content, and monosaccharide composition traits in cassava pulp using SNPs obtained from GBS across the 141 genotypes and 33 traits. A total of 29,733 SNPs, based on the cassava nuclear genome sequence version 8.1, were obtained from our previous study (Sunvittayakul et al., 2024). The cassava germplasm and genotypes are presented in File S1. Linkage disequilibrium (LD) decay had the midpoint r2 value of 0.04, corresponding to approximately 60 kb. GWAS was performed using three methods including MLM, FarmCPU, and BLINK using mean datasets for each parameter. Out of the 33 traits analyzed, a total of 128 significant SNPs were identified for 23 traits, excluding cellulose and de-starched material. Manhattan and QQ plots displaying significant SNPs of the 23 traits are presented in Figure 2, Figure S2 and Table S2. The significant SNPs were distributed across all 18 chromosomes, with phenotype variance explain (PVE) values ranging from low percentages up to 90 %. Saccharification was associated with four SNPs on chromosomes 7, 8, and 12, with PVE values ranging from 3.5% to 21.5%. For example, SNP 9955 on chromosome 7 had the highest PVE at 21.5% and was associated with the G/C allele. These four SNPs were specifically associated with saccharification, indicating unique loci for this trait.

In contrast, monosaccharide composition traits for matrix polysaccharides were associated with multiple SNPs distributed across all 18 chromosomes. Pectin and hemicellulose-related traits had multiple significant SNPs, with some traits showing up to 18 and 23 significant SNPs for PGal and PXyl, respectively, indicating the multigenic nature underlying these traits. Ara was also associated with multiple significant SNPs across several chromosomes, including SNP 14962 on chromosome 9, which had a high PVE of 59.7% indicating a strong influence on this hemicellulose trait. Notable SNPs for HRha included SNP 2330 on chromosome 2, explaining 35.2% of the phenotypic variance and SNP 6474 on chromosome 4 explaining 45.1% the phenotypic variance, suggesting key genetic regions regulating rhamnose content. Glc had significant SNPs on chromosome 17 and 12 associated with glucose content in hemicellulose, with SNP 26922 on chromosome 17 explaining 48.1% of phenotypic variance. PGal was found to be associated with SNPs on chromosomes 13 and 16 including SNP 25495 on chromosome 16, which explained 23.5% of phenotypic variance for galactose content in pectins. Man, Xyl and GlcA traits also had SNP associations across various chromosomes, with notable PVE of 41%, 46% and 41%, respectively.

Certain chromosomes, such as chromosomes 1, 3, 4, 7, 9, 10, 15, 16 and 17, exhibit higher density of significant SNPs, suggesting these chromosomes evolved to harbor key genetic regions affecting monosaccharide composition of matrix polysaccharide in cassava root pulp. Highlighted regions in blue across the chromosomes denote clusters of significant SNPs, marking these regions as potential hotspots for genes regulating biomass-related traits (Figure 2). Interestingly, several SNPs are shared among different traits, showing genetic loci influencing multiple traits simultaneously suggesting the possibility of pleiotropic effects. For example, SNP 6474 on chromosome 4 was repeatedly identified across eight monosaccharides traits, with PVE values ranging 13% to 65 %, making it a key locus for multiple biomass-related traits including Rha, Xyl, Man, GalA and GluA. SNP 14962 was also repeatedly associated with high phenotypic variance in multiple traits, including Ara, Gal, Xyl, Man, and Xyl, with PVE values ranging from 15.6% to 59.6%, highlighting its role in cassava biomass traits. Furthermore, SNP 2330 on chromosome 2 was identified in five traits relating to Rha, Man, and GlcA, with PVE values ranging from 6% to 35%. Additionally, SNPs on chromosome 1 and 18 were associated with various hemicellulose traits, while regions on chromosome 1, 10, 16 and 17 were associated with pectin traits. Regions on chromosome 2, 4, 9,15 were associated with both hemicellulose and pectin traits. The shared SNPs across traits provide insights into the interconnected pathways of matrix cell wall biosynthesis, aiding in the identification of key regulatory genes and pathways.

A diagram of a dna sequence

Description automatically generated

Figure 2. Distribution of significant SNPs associated with saccharification and monosaccharide composition of hemicellulose and pectin fractions in cassava pulp across the 18 chromosomes of the cassava genome. Each chromosome is depicted with significant SNPs indicated by their positions along the chromosome. Traits with significant SNPs are listed on the right panel, with the number of significant SNPs indicated in brackets. Shared SNPs across different traits are shown along the same lines. Yellow highlights denote regions with clusters of significant SNPs, suggesting potential genetic hotspots.

**Identification of candidate genes**

To identify candidate genes associated with the traits of interest, we examined regions 60 kb upstream and downstream of the identified SNPs. Detailed information on all associated gene loci and annotations is provided in File S2. Although many gene loci were identified, establishing a direct link between these genes and the analyzed traits based on SNP positions was challenging. To refine our gene list, we used transcriptome data from cassava storage roots (Wilson et al., 2017) and single-cell transcriptome data (Zang et al., 2024). These datasets allowed us to focus on genes active during storage root development. From an initial pool of 674 candidate genes associated with 23 traits, we narrowed it down to 118 genes for 21 traits that showed differential expression or specific expression in storage roots (File S3). We cross-referenced our gene list with reports from various cassava GWAS studies (Esuma et al., 2016; Okeke et al., 2017; Rabbi et al., 2017; Zhang et al., 2018; Ezenwaka et al., 2018; Kayondo et al., 2018; Uchendu et al., 2021; dos Santos Silva et al., 2021; Santos et al., 2022; Phumichai et al., 2022; Hohenfeld et al., 2022; Rabbi et al., 2022; Ding et al., 2023; Baguma et al., 2024) and identified two matching genes out of the 118. These include THPGlc-MANES\_05G026700 (common plant regulatory factor 1-like), which was associated with lobular length of cassava roots as reported by Zhang et al. (2008), and THPGalA-MANES\_08G046400 (Zn finger (CCCH-type) family protein), which was associated to cassava green mite resistance in the study by Ezenwaka et al. (2018). The Upset plot illustrates the number of genes shared among the 21 biomass traits based on significant SNPs (Figure 3). Each trait has a distinct set of genes identified through GWAS and transcriptome data, capturing the extent of gene sharing among these traits. The pectic Glc had the highest number of genes (23), while the saccharification trait had the fewest (1). Notably, the largest intersection involves 11 genes shared among 4-9 traits, 36 genes shared among 2-3 traits, and 71 genes found uniquely in each trait. These overlaps suggest common genetic pathways or regulatory mechanisms influencing multiple biomass traits in cassava pulp.

A graph with numbers and a chart

Description automatically generated with medium confidence

Figure 3. Upset plot showing the number of genes shared among 21 biomass traits analyzed in cassava pulp samples. Each bar on the upper part of the plot represents the size of gene intersections among the traits, with the height indicating the number of shared genes. The traits involved in each intersection are marked by dots in the matrix below the bars. The rightmost bar shows that 3 genes are shared among the largest intersection of traits. The sizes of individual trait gene sets are depicted by the horizontal bars on the bottom left side of the plot, with PGlc having the largest set of 23 genes. The top panel indicates the number of traits shared in each gene group.

For the saccharification trait, 11 candidate genes were initially identified based on four significant SNPs. However, after filtering through the transcriptome data, only one gene remained: YTH domain-containing protein ECT2 (MANES\_07G081200). This gene, along with the other 10 candidates, has no direct relation to plant cell wall processes and biomass based on current annotations. For monosaccharide traits, we identified 11 candidate genes shared among 4-9 traits (Table 2). These genes should be considered for their potential involvement in functions related to the properties of matrix polysaccharides in cassava pulp. Notably, these genes are shared among monosaccharide traits strictly related to either the hemicellulose or pectin fraction, suggesting their involvement in matrix polysaccharides from either group. Among these genes, five are of particular interest: MANES\_04G118100 (Stress-response A/B barrel domain-containing protein HS1-like), MANES\_09G174600 (Protein DEHYDRATION-INDUCED 19 homolog 3-like), MANES\_17G070300 (Calcyclin-binding protein-like), MANES\_10G035900 (Protein gar2), and MANES\_16G111000 (Ethylene-responsive transcription factor TINY-like). While their exact roles in cell wall metabolism are not well-defined, these genes are involved in stress-related responses and signaling that can impact plant growth and development, potentially influencing cell wall composition and integrity. Other candidate genes are mostly involved in regulating gene expression and protein biosynthesis machinery, making their role in carbohydrate metabolism or plant cell walls unclear. This genome-wide analysis identifies crucial genetic loci associated with saccharification and monosaccharide composition traits in cassava pulp, making them candidates for further genetic studies.

Table 2. Gene loci shared among 4-9 monosaccharides traits.

|  |  |  |
| --- | --- | --- |
| Gene loci | Annotation | Shared traits |
| MANES\_04G118100 | stress-response A/B barrel domain-containing protein HS1-like | HGal, HGalA, HGlcA, HMan, HRha, HXyl, THem, THPMan, THPXyl |
| MANES\_04G118600 | probable histone H2A variant 3 | HGal, HGalA, HGlcA, HMan, HRha, HXyl, THem, THPMan, THPXyl |
| MANES\_04G119100 | plastidic ATP/ADP-transporter-like | HGal, HGalA, HGlcA, HMan, HRha, HXyl, THem, THPMan, THPXyl |
| MANES\_09G174300 | eukaryotic translation initiation factor 3 subunit A-like | HAra, HGal, HGlcA, HMan, HXyl, THem, THPMan, THPXyl |
| MANES\_09G174600 | protein DEHYDRATION-INDUCED 19 homolog 3-like | HAra, HGal, HGlcA, HMan, HXyl, THem, THPMan, THPXyl |
| MANES\_17G070300 | calcyclin-binding protein-like | PAra, PGal, PGlc, PXyl, THPGalA, THPGlc, TPec |
| MANES\_10G035900 | protein gar2 | HGal, HGlcA, HMan, THem, THPMan |
| MANES\_15G075700 | protein translation factor SUI1 homolog 2-like | PAra, PGal, PGlc, PXyl, TPec |
| MANES\_16G110600 | SNW/SKI-interacting protein-like | PAra, PGal, PGlc, PXyl, TPec |
| MANES\_16G111000 | ethylene-responsive transcription factor TINY-like | PAra, PGal, PGlc, PXyl, TPec |
| MANES\_16G030700 | MACPF domain-containing protein NSL1-like | PGal, PGlc, PXyl, TPec |

**Sequence and expression analysis of candidate genes**

To further investigate the 12 candidate genes identified, we analyzed the sequence variation among six cassava progenitors: KU50, R1, R90, R5, HB80, and HNT. The coding sequences were derived from re-sequencing data of these progenitors and mapped onto the corresponding reference sequences. We translated these sequences to amino acids to infer potential functional differences among the progenitors, with a focus on their impact on cell wall traits in the S1 and S2 progeny. The sequence analysis showed that MANES\_04G118600 had no variation among progenitors, suggesting a conserved role across the different genetic backgrounds. Cladograms were constructed to illustrate the amino acid sequence variation encoded by the remaining 11 candidate genes among the progenitors (Figure 4), with detailed sequence alignments presented in File S4. Sequence variations were observed in genes including MANES\_07G081200, MANES\_04G119100, MANES\_09G174300, MANES\_10G035900, MANES\_16G111000, and MANES\_16G030700. Notably, premature stop codons were identified at position 92 of MANES\_16G030700 in HB80, and at position 380 of MANES\_09G174300 in KU50, which could have a significant impact on gene function. Distinct grouping patterns were also observed in several genes, indicating potential functional divergence among progenitors. For example, MANES\_04G118100 grouped into two clusters, with a single amino acid difference noted for R90, while MANES\_09G174600 and MANES\_15G075700 each formed three distinct groups. MANES\_16G110600 formed two clusters, and premature stop codons were identified at positions 12 and 15 in this gene for HB80, and at position 75 in MANES\_15G075700 for KU50. These amino acid variations, particularly premature stop codons and group-specific amino acid changes, suggest functional alterations that may contribute to phenotypic diversity in cell wall composition among cassava progeny, potentially impacting biomass-related traits.

To determine whether gene expression levels contribute to variation in cell wall traits, we performed qRT-PCR analysis on the 12 candidate genes using RNA isolated from 3-month-old young storage roots from the varieties KU50, HNT, and HB60 (a sibling of HB80). Of the 12 genes analyzed, five showed no detectable expression, while the remaining seven exhibited varying expression levels (Figure 5). Among the genes showing detectable expression, MANES\_04G118600, MANES\_09G174600, MANES\_10G035900, and MANES\_16G111000 showed consistent expression across the varieties without significant changes. However, three genes including MANES\_17G070300, MANES\_15G075700, and MANES\_16G030700 demonstrated significant variation in expression levels, with KU50 consistently exhibiting higher expression compared to HNT and HB60. These findings suggest that differential expression of these three genes may also play roles in the variation of cell wall composition traits among progeny.

**A group of white paper with black text

Description automatically generated**

Figure 4. Cladograms of amino acid sequences encoded by the 11 candidate genes from six cassava progenitors (KU50, R1, R90, R5, HB80, and HNT). These sequences were obtained from re-sequencing data and aligned to identify variations among the progenitors. Each cladogram represents the phylogenetic relationships and sequence similarities among the six progenitors for a specific candidate gene. The bootstrap values indicate the reliability of the groupings.

**A black background with white text

Description automatically generated with medium confidence**

**Figure 5. qRT-PCR analysis of candidate genes in KU50, HNT and HB60 varieties**. Three-month-old young storage root samples were used for RNA isolation. Each sample had three biological replicates, with technical duplicates for each replicate. Gene expression data are presented as fold change, normalized to cassava4.1\_006884 as a reference gene. Statistical significance was determined using the Kruskal-Wallis test (P < 0.05). Bars represent mean values, and error bars indicate standard error of the mean. Non-significant differences are indicated by "ns".

**Discussion**

Cell wall composition plays a crucial role in determining the feasibility of using plant biomass for various applications that require biomass breakdown (Marriott et al., 2014; Wu et al., 2014; De Souza et al., 2015; Ostos Garrido et al., 2018). Studies on various feedstocks have demonstrated that the ease of converting cellulose and hemicellulose to fermentable sugars is a key determinant of a crop's suitability for biofuel production (Alvira et al., 2013). Cassava pulp, a byproduct of starch extraction, represents a significant untapped resource (Sriroth et al., 2000), while the residual starch in the pulp can further be utilized, adding value to the bioconversion process (Adetunji et al., 2016; Poonsrisawat et al., 2017). In this study, we show that saccharification in our cassava population ranges between 39-95 nmol Glu mg-1 h-1, indicating a potential for further development in cassava breeding to enhance the use of cassava pulp for bioconversion. High estimates of broad-sense heritability for saccharification and many monosaccharide traits also suggest that these traits can be improved through selective breeding. The recalcitrance of wood and grass feedstocks is largely due to complex interactions among cellulose, xylan, and lignin in secondary cell walls. In contrast, cassava pulp, primarily composed of the primary walls of parenchyma cells with starch deposits and limited vascular bundles, poses different challenges (Odoch et al., 2017; Sun et al., 2023).

The matrix polysaccharides and cellulose composition in cassava are crucial not only for plant structure and growth but also for their dynamic transition between structural and storage components during root development (Sun et al., 2023). Compared to other biofuel crops like corn stover, switchgrass, and sugarcane bagasse, cassava pulp has unique cell wall characteristics that may influence its bioconversion efficiency. For instance, cassava typically exhibits a higher proportion of pectin and lower lignin content than those crops (Sriroth et al., 2000; Odoch et al., 2017). This is advantageous because lignin can impede the saccharification process (Halpin, 2019; Johnston et al., 2020; López-Malvar et al., 2021; Mota et al., 2021). In this study, analysis of over 100 cassava varieties showed that pectin can account for a proportion equal to hemicellulose and cellulose in the de-starched material. These structural and compositional differences suggest that cassava may require less pretreatment, significantly reducing the cost of bioconversion compared to more recalcitrant materials.

The correlation analysis of biomass traits in cassava pulp underscores significant associations that are crucial for optimizing cassava for both growth traits and saccharification efficiency for industrial applications. The positive correlations between cellulose content, saccharification, and de-starched material suggest that cassava’s bioconversion potential can be considerably enhanced by targeting these traits through selective breeding. Negative correlations of cellulose and saccharification with glucose in the hemicellulose and pectin fractions (and with glucuronic acid in the hemicellulose fraction), suggest that saccharification efficiency may be inhibited by these components. This aligns with findings in other bioenergy crops, where hemicellulose and pectin constituents negatively impact saccharification due to their complex interaction with cellulose, affecting enzymatic hydrolysis (Xu et al., 2012; Francocci et al., 2013; Leijdekkers et al., 2013; Chen et al., 2016; De Souza et al., 2015; Deralia et al., 2021).

Strong positive correlations among non-glucose monosaccharides within the hemicellulose fraction support the integrity of hemicellulose biosynthesis involving different types of polysaccharides. Such correlations have been observed in other biomass crops, where interdependent monosaccharide levels have been manipulated to optimize growth and bioconversion efficiencies (Zhang et al., 2021). Additionally, relationships observed in the pectin fraction, particularly strong correlations between arabinose, rhamnose, galactose, and mannose, suggest similar regulatory processes in pectin biosynthesis. These relationships have been observed in crops like maize, sorghum, and switchgrass (Petti et al., 2013).

The GWAS on cassava pulp traits identified 128 significant SNPs across 23 biomass traits, including saccharification and monosaccharide compositions in hemicellulose and pectin. This underscores the distinct genetic basis of saccharification efficiency and highlights the complex genetic architecture controlling these traits in cassava. These findings are consistent with observations in other bioenergy crops like rice straw, corn, sorghum, and Andean lupin, where distinct genomic regions have been associated with traits conducive to higher biofuel yield (Wang et al., 2013; Li et al., 2016; Nguyen et al., 2020; Panahabadi et al., 2022; Gesteiro et al., 2023; Gulisano et al., 2023). This complexity is particularly evident from SNPs with high phenotypic variance explained (PVE), indicating that certain genetic loci have substantial influence over specific traits. The distribution of SNPs associated with saccharification and monosaccharide traits across all 18 chromosomes suggests a polygenic basis for these biomass traits. For instance, SNPs located on chromosomes 7, 8, and 12 specifically influence saccharification, while other SNPs exhibit broader impacts across multiple monosaccharide traits. SNPs with consistently high PVE values, such as those on chromosomes 7, 9, and 17, point to key genomic regions that can be targeted in breeding programs (Wolfe et al., 2016). Moreover, the identification of certain chromosomes, specifically chromosomes 1, 3, 4, 7, 9, 10, 15, 16, and 17, as harboring a higher density of significant SNPs provides insights into potential genomic "hotspots" for biomass improvement. These chromosomes can serve as focal points for further research and breeding strategies aimed at increasing biomass yield, optimizing saccharification, and enhancing the composition of cassava pulp for bioenergy.

Notably, SNPs associated with saccharification of cassava pulp are distinct from those associated with monosaccharide composition traits. The co-localization of significant SNPs for monosaccharide traits suggests that these regions may harbor key genes involved in the biosynthesis and regulation of cell wall polysaccharides. Importantly, SNPs repeatedly associated with multiple traits suggest the presence of pleiotropic loci or tightly linked regions affecting several traits simultaneously, for example, SNP 6474 on chromosome 4 and SNP 14962 on chromosome 9 for their widespread effects on cell wall polysaccharide composition. These findings align with previous studies demonstrating pleiotropy in cassava genetics (Ezenwaka et al., 2018; Rabbi et al., 2022; Santos et al., 2022; Baguma et al., 2024).

The GWAS has provided an extensive list of candidate genes potentially involved in the biochemistry of cell wall composition and properties, particularly those influencing saccharification and monosaccharide composition. However, we did not find genes directly related to cell wall biosynthesis such as UDP-sugar conversions, glycosyltransferase, or cell wall modification enzymes like glycohydrolases. By integrating SNP positions and expression data from transcriptome studies, 118 genes for 21 traits were identified. The distribution of these genes across traits revealed overlaps within the same wall fraction, either hemicellulose or pectin, supporting shared genetic pathways or regulatory mechanisms. Such overlaps have been observed in other studies, such as in wheat and barley, where genes influencing lignin content and cellulose composition showed pleiotropic effects or close linkage (Hao and Mohnen 2014; Muchero et al., 2015; Li et al., 2016; Yang et al., 2019; Bahri et al., 2020).

The identification and sequence and expression analysis of the 12 candidate genes have added a deeper understanding of the genetic factors governing saccharification and biomass traits in cassava. Notably, the YTH domain-containing protein ECT2 (MANES\_07G081200), identified as a key gene for saccharification, suggests a potential role in biomass conversion efficiency despite its lack of a direct function in cell wall processes. This gene has been shown to play a role in mRNA processing and is involved in trichome development, growth, and stress responses in rice (Scutenaire et al., 2018; Arribas-Hernández et al., 2021; Ma et al., 2022). This observation aligns with previous findings where genes involved in regulatory networks influenced lignin and cell wall biosynthesis (Ferreira et al., 2016; Rao et al., 2019), highlighting its potential indirect role in biomass conversion. In addition, several candidate genes identified for monosaccharide traits, such as MANES\_04G118100 and MANES\_09G174600, are primarily related to stress responses. Their association with biomass traits may indicate an adaptive mechanism where stress-response pathways indirectly modify the cell wall to cope with environmental stress, thereby influencing biomass characteristics. This is consistent with studies in Arabidopsis and rice, where stress-related genes were shown to alter cell wall integrity and composition under stress conditions (Mirdar Mansuri et al., 2020; Cao et al., 2021). Moreover, sequence variation among the identified genes, particularly in amino acid sequences, suggests a potential role in the phenotypic variation of the cell wall traits across different cassava progenitors. This sequence diversity points to functional divergence in how these genes regulate cell wall traits. Notably, the expression analysis through qRT-PCR revealed differential expression patterns in specific genes, including MANES\_16G030700 (calcyclin-binding protein-like), MANES\_15G075700 (protein translation factor SUI1 homolog 2-like), and MANES\_16G030700 (MACPF domain-containing protein NSL1-like), which may contribute to the observed phenotypic variation in cell wall composition. The identification of these candidate genes provides a promising foundation for future functional studies to confirm their roles and usability as markers for breeding cassava varieties with enhanced biomass properties.

**Conclusion**

Overall, this study highlights significant genetic effects on cassava pulp traits, particularly those related to saccharification and cell wall composition. The integration of phenotypic data with GWAS and transcriptome analysis offers insights into the genetic basis of these traits, underpinning future breeding efforts to improve cassava varieties for industrial applications. The identification of significant SNPs associated with these traits, as well as putative candidate genes, provides a valuable resource for marker-assisted breeding efforts and validation of causal genes controlling these industrially-relevant biomass characteristics. These insights pave the way for targeted genetic improvements, which could significantly impact the economic viability and sustainability of cassava as a versatile bioenergy crop.

**Supplementary Information**

Table S1. qRT-PCR primers.

Table S2. Summary of 128 SNPs significantly associated with 23 biomass traits from cassava pulp. The trait abbreviations are indicated in Table 2.

Figure S1. Distribution of 33 biomass traits based on natural Log of 141 genotypes. The trait abbreviations are indicated in Table 2.

Figure S2. Circular Manhattan plots and QQ plots indicating the single-nucleotide polymorphisms (SNPs) associated with 23 biomass traits of cassava pulp traits. The trait abbreviations are indicated in Table 2.

File S1. The cassava germplasm and genotypes used in this work. NCBI accessions of raw reads for each genotype are included.

File S2. All associated SNPs associated with 23 biomass traits and candidates genes identified based on 120 kb window.

File S3. Candidates genes filtered through transcriptome data based on cassava storage roots (Wilson et al., 2017) and single-cell transcriptome data (Zang et al., 2024) (for vascular tissue and meristem).

File S4. Amino acid sequence alignments of 12 candidate genes from six cassava progenitors.  
The amino acid sequences of 12 candidate genes from six cassava progenitors (HB80, HNT, R5, KU50, R1, and R90) obtained from genome re-sequencing data were aligned using CLUSTAL O (version 1.2.4) for multiple sequence alignment. The alignment highlights sequence variations among the progenitors.

**Author Contributions:** Conceptualization, LG and SV; methodology, PS, PWo, PWa, PP, KC, AS, PK LG and SV; formal analysis, PS, SV, PWo, PWa, AS, HC, LG and PK; resources, PK; Figure preparation, PS, SV and PWo; writing, PS, HC, LG and SV; All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the Office of the Ministry of Higher Education, Science, Research and Innovation; and the Thailand Science Research and Innovation through the Kasetsart University Reinventing University Program 2021. PS was supported by The Royal Golden Jubilee (RGJ) Ph.D. Programme (Thailand): PHD/0164/2560, the Thailand Research Fund (TRF) and Department of Agriculture, Ministry of Agriculture and Cooperative (Thailand). SV was supported by the National Research Council of Thailand (NRCT): NRCT5-RSA63002-02. PK was supported by NSTDA (P-18-52170). PWo was supported by NRCT-N42A650286. We are also grateful for the support of the Biotechnology and Biological Sciences Research Council (grant reference BB/Y51424X/1 - BBSRC International Institutional Sponsorship Awards Tranche 1 York).

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We thank for support from the Kasetsart University Research and Development Institute (KURDI) and Center for Advanced Studies of Agriculture and Food (CASAF).

**Declarations**

**Conflicts of Interest:** The authors declare no conflict of interest.

**Declaration of generative AI and AI-assisted technologies in the writing process**

During the preparation of this work the author(s) used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

**References**

Adetunji, A.I., Du Clou, H., Walford, S.N. and Taylor, J.R., 2016. Complementary effects of cell wall degrading enzymes together with lactic acid fermentation on cassava tuber cell wall breakdown. *Industrial Crops and Products*, *90*, pp.110-117.

Alvira, P., Ballesteros, M. and Negro, M.J., 2013. Progress on enzymatic saccharification technologies for biofuels production. *Biofuel technologies: recent developments*, pp.145-169.

Apiwatanapiwat, W., Murata, Y., Kosugi, A., Yamada, R., Kondo, A., Arai, T., Rugthaworn, P. and Mori, Y., 2011. Direct ethanol production from cassava pulp using a surface-engineered yeast strain co-displaying two amylases, two cellulases, and β-glucosidase. *Applied microbiology and biotechnology*, *90*, pp.377-384.

Arribas-Hernández, L., Rennie, S., Schon, M., Porcelli, C., Enugutti, B., Andersson, R., Nodine, M.D. and Brodersen, P., 2021. The YTHDF proteins ECT2 and ECT3 bind largely overlapping target sets and influence target mRNA abundance, not alternative polyadenylation. *Elife*, *10*, p.e72377.

Baguma, J.K., Mukasa, S.B., Nuwamanya, E., Alicai, T., Omongo, C.A., Ochwo-Ssemakula, M., Ozimati, A., Esuma, W., Kanaabi, M., Wembabazi, E. and Baguma, Y., 2024. Identification of genomic regions for traits associated with flowering in cassava (Manihot esculenta Crantz). *Plants*, *13*(6), p.796.

Bahri, B.A., Daverdin, G., Xu, X., Cheng, J.F., Barry, K.W., Brummer, E.C., Missaoui, A. and Devos, K.M., 2020. Natural variation in lignin and pectin biosynthesis-related genes in switchgrass (Panicum virgatum L.) and association of SNP variants with dry matter traits. *BioEnergy research*, *13*, pp.79-99.

Brito AC, Oliveira SAS, Oliveira EJ. Genome-wide association study for resistance to cassava root rot. The Journal of Agricultural Science. 2017; 155(9): 1424–1441.

Cao, Y., Zhang, Y., Chen, Y., Yu, N., Liaqat, S., Wu, W., Chen, D., Cheng, S., Wei, X., Cao, L. and Zhang, Y., 2021. OsPG1 encodes a polygalacturonase that determines cell wall architecture and affects resistance to bacterial blight pathogen in rice. *Rice*, *14*, pp.1-15.

Chen, H., Zhao, X. and Liu, D., 2016. Relative significance of the negative impacts of hemicelluloses on enzymatic cellulose hydrolysis is dependent on lignin content: evidence from substrate structural features and protein adsorption. *ACS Sustainable Chemistry & Engineering*, *4*(12), pp.6668-6679.

Cock JH. Cassava: a basic energy source in the tropics. Science. 1982 Nov 19;218(4574):755-62. doi: 10.1126/science.7134971. PMID: 7134971.

Dagaew, Gamonmas, et al. "Feed utilization efficiency and ruminal metabolites in beef cattle fed with cassava pulp fermented yeast waste replacement soybean meal." Scientific Reports 12.1 (2022): 16090.

De Souza, A.P., Kamei, C.L.A., Torres, A.F., Pattathil, S., Hahn, M.G., Trindade, L.M. and Buckeridge, M.S., 2015. How cell wall complexity influences saccharification efficiency in Miscanthus sinensis. *Journal of Experimental Botany*, *66*(14), pp.4351-4365.

Deralia, P.K., Jensen, A., Felby, C. and Thygesen, L.G., 2021. Chemistry of lignin and hemicellulose structures interacts with hydrothermal pretreatment severity and affects cellulose conversion. *Biotechnology Progress*, *37*(5), p.e3189.

Ding, Z., Fu, L., Wang, B., Ye, J., Ou, W., Yan, Y., Li, M., Zeng, L., Dong, X., Tie, W. and Ye, X., 2023. Metabolic GWAS-based dissection of genetic basis underlying nutrient quality variation and domestication of cassava storage root. *Genome Biology*, *24*(1), p.289.

Djuma’ali, D.A., Soewarno, N., Sumarno, S., Primarini, D. and Sumaryono, W., 2011. Cassava pulp as a biofuel feedstock of an enzymatic hydrolysis proces. *Makara Journal of Technology*, *15*(2), p.14.

dos Santos Silva, P.P., Sousa, M.B.E., de Oliveira, E.J., Morgante, C.V., de Oliveira, C.R.S., Vieira, S.L. and Borel, J.C., 2021. Genome-wide association study of drought tolerance in cassava. *Euphytica*, *217*, pp.1-26.

Esuma, W., Herselman, L., Labuschagne, M.T., Ramu, P., Lu, F., Baguma, Y., Buckler, E.S. and Kawuki, R.S., 2016. Genome-wide association mapping of provitamin A carotenoid content in cassava. *Euphytica*, *212*, pp.97-110.

Ezenwaka, L., Del Carpio, D.P., Jannink, J.L., Rabbi, I., Danquah, E., Asante, I., Danquah, A., Blay, E. and Egesi, C., 2018. Genome‐wide association study of resistance to cassava green mite pest and related traits in cassava. *Crop Science*, *58*(5), pp.1907-1918.

Ferreira, S.S., Hotta, C.T., Poelking, V.G.D.C., Leite, D.C.C., Buckeridge, M.S., Loureiro, M.E., Barbosa, M.H.P., Carneiro, M.S. and Souza, G.M., 2016. Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant molecular biology*, *91*, pp.15-35.

Fonseca, P.A., Suárez-Vega, A., Marras, G. and Cánovas, Á., 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *Gigascience*, *9*(12), p.giaa149.

Francocci, F., Bastianelli, E., Lionetti, V., Ferrari, S., De Lorenzo, G., Bellincampi, D. and Cervone, F., 2013. Analysis of pectin mutants and natural accessions of Arabidopsis highlights the impact of de-methyl-esterified homogalacturonan on tissue saccharification. *Biotechnology for biofuels*, *6*, pp.1-8.

Garrison, E. and Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gesteiro, N., Butrón, A., Santiago, R., Gomez, L.D., López-Malvar, A., Álvarez-Iglesias, L., Revilla, P. and Malvar, R.A., 2023. Breeding dual-purpose maize: Grain production and biofuel conversion of the stover. *Agronomy*, *13*(5), p.1352.

Gomez, L.D., Whitehead, C., Barakate, A., Halpin, C. and McQueen-Mason, S.J., 2010. Automated saccharification assay for determination of digestibility in plant materials. *Biotechnology for biofuels*, *3*, pp.1-12.

Gulisano, A., Dechesne, A., Paulo, M.J. and Trindade, L.M., 2023. Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: First genome‐wide association study on Lupinus mutabilis biomass quality. *GCB Bioenergy*, *15*(1), pp.38-57.

Halpin, C., 2019. Lignin engineering to improve saccharification and digestibility in grasses. *Current opinion in biotechnology*, *56*, pp.223-229.

Hao, Z. and Mohnen, D., 2014. A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes. *Critical reviews in biochemistry and molecular biology*, *49*(3), pp.212-241.

Hohenfeld, C.S., Passos, A.R., de Carvalho, H.W.L., de Oliveira, S.A.S. and de Oliveira, E.J., 2022. Genome-wide association study and selection for field resistance to cassava root rot disease and productive traits. *Plos one*, *17*(6), p.e0270020.

Hu, M., Hu, W., Xia, Z., Zhou, X. and Wang, W., 2016. Validation of reference genes for relative quantitative gene expression studies in cassava (Manihot esculenta Crantz) by using quantitative real-time PCR. *Frontiers in Plant Science*, *7*, p.680.

Iwai, C.B., Ta-Oun, M.O.N.G.K.O.N., Chuasavatee, T.H.A.M.M.A.R.E.D. and Boonyotha, P.R.A.W.E.E.N.A., 2013. Management of municipal sewage sludge by vermicomposting technology: Converting a waste into a bio fertilizer for agriculture. *International Journal of Environmental and Rural Development*, *4*(1), pp.169-174.

Jarvis A, Ramirez-Villegas J, Campo BVH, Navarro-Racines C. 2012. Is cassava the answer to African climate change adaptation? Tropical Plant Biology 5(1): 9-29.

Johnston, P.A., Zhou, H., Aui, A., Wright, M.M., Wen, Z. and Brown, R.C., 2020. A lignin-first strategy to recover hydroxycinnamic acids and improve cellulosic ethanol production from corn stover. *Biomass and bioenergy*, *138*, p.105579.

Kayondo, S.I., Pino Del Carpio, D., Lozano, R., Ozimati, A., Wolfe, M., Baguma, Y., Gracen, V., Offei, S., Ferguson, M., Kawuki, R. and Jannink, J.L., 2018. Genome-wide association mapping and genomic prediction for CBSD resistance in Manihot esculenta. *Scientific Reports*, *8*(1), p.1549.

Kosugi, A., Kondo, A., Ueda, M., Murata, Y., Vaithanomsat, P., Thanapase, W., Arai, T. and Mori, Y., 2009. Production of ethanol from cassava pulp via fermentation with a surface-engineered yeast strain displaying glucoamylase. *Renewable Energy*, *34*(5), pp.1354-1358.

Leijdekkers, A.G.M., Bink, J.P.M., Geutjes, S., Schols, H.A. and Gruppen, H., 2013. Enzymatic saccharification of sugar beet pulp for the production of galacturonic acid and arabinose; a study on the impact of the formation of recalcitrant oligosaccharides. *Bioresource technology*, *128*, pp.518-525.

Lerdlattaporn, R., Phalakornkule, C., Trakulvichean, S. and Songkasiri, W., 2021. Implementing circular economy concept by converting cassava pulp and wastewater to biogas for sustainable production in starch industry. *Sustainable Environment Research*, *31*(1), p.20.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *bioinformatics*, *25*(16), pp.2078-2079.

Li, K., Wang, H., Hu, X., Liu, Z., Wu, Y. and Huang, C., 2016. Genome-wide association study reveals the genetic basis of stalk cell wall components in maize. *PloS one*, *11*(8), p.e0158906.

Li, S., Cui, Y., Zhou, Y., Luo, Z., Liu, J., & Zhao, M. (2017). The industrial applications of cassava: current status, opportunities and prospects. Journal of the Science of Food and Agriculture, 97(8), 2282–2290. doi:10.1002/jsfa.8287

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z., 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics*, *28*(18), pp.2397-2399.

López-Malvar, A., Santiago, R., Malvar, R.A., Martín, D., Pereira dos Santos, I., Batista de Carvalho, L.A., Faas, L., Gómez, L.D. and da Costa, R.M., 2021. FTIR screening to elucidate compositional differences in maize recombinant inbred lines with contrasting saccharification efficiency yields. *Agronomy*, *11*(6), p.1130.

Ma, W., Cui, S., Lu, Z., Yan, X., Cai, L., Lu, Y., Cai, K., Zhou, H., Ma, R., Zhou, S. and Wang, X., 2022. YTH domain proteins play an essential role in rice growth and stress response. *Plants*, *11*(17), p.2206.

Marriott, P.E., Sibout, R., Lapierre, C., Fangel, J.U., Willats, W.G., Hofte, H., Gómez, L.D. and McQueen-Mason, S.J., 2014. Range of cell-wall alterations enhance saccharification in Brachypodium distachyon mutants. *Proceedings of the National Academy of Sciences*, *111*(40), pp.14601-14606.

Mirdar Mansuri, R., Shobbar, Z.S., Babaeian Jelodar, N., Ghaffari, M., Mohammadi, S.M. and Daryani, P., 2020. Salt tolerance involved candidate genes in rice: an integrative meta-analysis approach. *BMC Plant Biology*, *20*, pp.1-14.

Morse EE. 1947. Anthrone in estimating low concentration of sucrose. Anal Chem 19, 1012-1013.

Mota, T.R., Oliveira, D.M., Simister, R., Whitehead, C., Lanot, A., Dos Santos, W.D., Rezende, C.A., McQueen-Mason, S.J. and Gomez, L.D., 2021. Design of experiments driven optimization of alkaline pretreatment and saccharification for sugarcane bagasse. *Bioresource technology*, *321*, p.124499.

Muchero, W., Guo, J., DiFazio, S.P., Chen, J.G., Ranjan, P., Slavov, G.T., Gunter, L.E., Jawdy, S., Bryan, A.C., Sykes, R. and Ziebell, A., 2015. High-resolution genetic mapping of allelic variants associated with cell wall chemistry in Populus. *BMC genomics*, *16*, pp.1-14.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C. and Durbin, R., 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), pp.1749-1751.

Nguyen, D.T., Gomez, L.D., Harper, A., Halpin, C., Waugh, R., Simister, R., Whitehead, C., Oakey, H., Nguyen, H.T., Nguyen, T.V. and Duong, T.X., 2020. Association mapping identifies quantitative trait loci (QTL) for digestibility in rice straw. *Biotechnology for biofuels*, *13*, pp.1-16.

Odoch, M., Buys, E.M. and Taylor, J.R., 2017. Mechanism of cassava tuber cell wall weakening by dilute sodium hydroxide steeping. *Food chemistry*, *228*, pp.338-347.

Ogbonna AC, Andrade LRB, Rabbi IY, Mueller LA, Oliveira Ej, Bauchet GJ. Large-scale genome-wide association study, using historical data, identifies conserved genetic architecture of cyanogenic glucoside content in cassava (Manihot esculenta Crantz) root. The Plant Journal. 2021; 105(3): 754–770. doi: 10.1111/tpj.15071

Ostos Garrido FJ, Pistón F, Gómez LD, McQueen-Mason SJ (2018) Biomass recalcitrance in barley, wheat and triticale straw: Correlation of biomass quality with classic agronomical traits. PLOS ONE 13(11): e0205880.

Panahabadi, R., Ahmadikhah, A., McKee, L.S., Ingvarsson, P.K. and Farrokhi, N., 2022. Genome‐wide association study for lignocellulosic compounds and fermentable sugar in rice straw. *The Plant Genome*, *15*(1), p.e20174.

Parmar, A., Sturm, B. and Hensel, O., 2017. Crops that feed the world: Production and improvement of cassava for food, feed, and industrial uses. *Food Security*, *9*, pp.907-927.

Patle S, Lal B. Investigation of the potential of agro-industrial material as low cost substrate for ethanol production by using Candida tropicalis and Zymomonas mobilis.Biomass Bioenerg. 2008;32(7):596-602.

Peterson, B.G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., Cornilly, D., Hung, E., Lestel, M., Balkissoon, K. and Wuertz, D., 2018. Package ‘performanceanalytics’. *R Team Cooperation*, *3*, pp.13-14.

Petti, C., Shearer, A., Tateno, M., Ruwaya, M., Nokes, S., Brutnell, T. and DeBolt, S., 2013. Comparative feedstock analysis in Setaria viridis L. as a model for C4 bioenergy grasses and Panicoid crop species. *Frontiers in plant science*, *4*, p.181.

Phumichai, C., Aiemnaka, P., Nathaisong, P., Hunsawattanakul, S., Fungfoo, P., Rojanaridpiched, C., Vichukit, V., Kongsil, P., Kittipadakul, P., Wannarat, W. and Chunwongse, J., 2022. Genome-wide association mapping and genomic prediction of yield-related traits and starch pasting properties in cassava. *Theoretical and Applied Genetics*, pp.1-27.

Poonsrisawat, A., Paemanee, A., Wanlapatit, S., Piyachomkwan, K., Eurwilaichitr, L. and Champreda, V., 2017. Simultaneous saccharification and viscosity reduction of cassava pulp using a multi-component starch-and cell-wall degrading enzyme for bioethanol production. *3 Biotech*, *7*, pp.1-10.

Rabbi, I.Y., Kayondo, S.I., Bauchet, G., Yusuf, M., Aghogho, C.I., Ogunpaimo, K., Uwugiaren, R., Smith, I.A., Peteti, P., Agbona, A. and Parkes, E., 2022. Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant Molecular Biology*, *109*(3), pp.195-213.

Rabbi, I.Y., Udoh, L.I., Wolfe, M., Parkes, E.Y., Gedil, M.A., Dixon, A., Ramu, P., Jannink, J.L. and Kulakow, P., 2017. Genome‐wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *The Plant Genome*, *10*(3), pp.plantgenome2016-09.

Rao, X., Chen, X., Shen, H., Ma, Q., Li, G., Tang, Y., Pena, M., York, W., Frazier, T.P., Lenaghan, S. and Xiao, X., 2019. Gene regulatory networks for lignin biosynthesis in switchgrass (Panicum virgatum). *Plant Biotechnology Journal*, *17*(3), pp.580-593.

Saeman, J.F., 1945. Kinetics of wood saccharification-hydrolysis of cellulose and decomposition of sugars in dilute acid at high temperature. *Industrial & Engineering Chemistry*, *37*(1), pp.43-52.

Santos, C.S.D., Sousa, M.B., Brito, A.C., De Oliveira, L.A., Carvalho, C.W.P. and De Oliveira, E.J., 2022. Genome-wide association study of cassava starch paste properties. *Plos one*, *17*(1), p.e0262888.

Scutenaire, J., Deragon, J.M., Jean, V., Benhamed, M., Raynaud, C., Favory, J.J., Merret, R. and Bousquet-Antonelli, C., 2018. The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in Arabidopsis. *The Plant Cell*, *30*(5), pp.986-1005.

Sriroth, K., Chollakup, R., Chotineeranat, S., Piyachomkwan, K. and Oates, C.G., 2000. Processing of cassava waste for improved biomass utilization. *Bioresource technology*, *71*(1), pp.63-69.

Sun, J., Hui, K., Guo, Z., Li, Y. and Fan, X., 2023. Cellulose and lignin contents are negatively correlated with starch accumulation, and their correlation characteristics vary across cassava varieties. *Journal of Plant Growth Regulation*, *42*(2), pp.658-669.

Sunvittayakul, P., Wonnapinij, P., Chanchay, P., Wannitikul, P., Sathitnaitham, S., Phanthanong, P., Changwitchukarn, K., Suttangkakul, A., Ceballos, H., Gomez, L.D. and Kittipadakul, P., 2024. Genome-Wide Association Studies of Three-Dimensional (3D) Cassava Root Crowns and Agronomic Traits Using Partially Inbred Populations. *Agronomy*, *14*(3), p.591.

Theander O, Aman P, Westerlund E, Andersson R, Pettersson D. Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason lignin (the Uppsala method): collaborative study. J AOAC Int. 1995 Jul-Aug;78(4):1030-44. PMID: 7580315.

Uchendu, K., Njoku, D.N., Paterne, A., Rabbi, I.Y., Dzidzienyo, D., Tongoona, P., Offei, S. and Egesi, C., 2021. Genome-wide association study of root mealiness and other texture-associated traits in cassava. *Frontiers in Plant Science*, *12*, p.770434.

Virunanon, C., Ouephanit, C., Burapatana, V. and Chulalaksananukul, W., 2013. Cassava pulp enzymatic hydrolysis process as a preliminary step in bio-alcohols production from waste starchy resources. *Journal of Cleaner Production*, *39*, pp.273-279.

Wang, Y.H., Acharya, A., Burrell, A.M., Klein, R.R., Klein, P.E. and Hasenstein, K.H., 2013. Mapping and candidate genes associated with saccharification yield in sorghum. *Genome*, *56*(11), pp.659-665.

Wattanagonniyom, T., Lee, W.C., Tolieng, V., Tanasupawat, S. and Akaracharanya, A., 2017. Co-fermentation of cassava waste pulp hydrolysate with molasses to ethanol for economic optimization. *Annals of Microbiology*, *67*, pp.157-163.

Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P, et al. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. The Plant Genome. 2016; 9(2). doi: 10.3835/plantgenome2015.11.0118

Wu, Z., Hao, H., Tu, Y., Hu, Z., Wei, F., Liu, Y., Zhou, Y., Wang, Y., Xie, G., Gao, C. and Cai, X., 2014. Diverse cell wall composition and varied biomass digestibility in wheat straw for bioenergy feedstock. *biomass and bioenergy*, *70*, pp.347-355.

Xu, N., Zhang, W., Ren, S., Liu, F., Zhao, C., Liao, H., Xu, Z., Huang, J., Li, Q., Tu, Y. and Yu, B., 2012. Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H 2 SO 4 pretreatments in Miscanthus. *Biotechnology for biofuels*, *5*, pp.1-12.

Yang, X., Todd, J., Arundale, R., Binder, J.B., Luo, Z., Islam, M.S., Sood, S. and Wang, J., 2019. Identifying loci controlling fiber composition in polyploid sugarcane (Saccharum spp.) through genome-wide association study. *Industrial crops and products*, *130*, pp.598-605.

Zargar SM, Raatz B, Sonah H. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. Journal of Crop Science and Biotechnology. 2015; 18(5): 293–308. Zhang et al., 2016

Zhang, S., Chen, X., Lu, C., Ye, J., Zou, M., Lu, K., Feng, S., Pei, J., Liu, C., Zhou, X. and Ma, P.A., 2018. Genome-wide association studies of 11 agronomic traits in cassava (Manihot esculenta Crantz). *Frontiers in Plant Science*, *9*, p.503.

Zhang, W., Qin, W., Li, H. and Wu, A.M., 2021. Biosynthesis and transport of nucleotide sugars for plant hemicellulose. *Frontiers in plant science*, *12*, p.723128.