

# Genomic Insights into Retting Ability and Post-Harvest Quality Traits in Cassava (*Manihot esculenta* Crantz)

Umuahia, Nigeria, 31/10/2024

Ugo CHIJIKE National Root Crop Research Institute, Umudike, Umuahia, Abia State, Nigeria.

Simon Peter ABAH, Damian NJOKU, Justice OKORONKWO, Ugochi IRO, Precious UDOKA, Sonia OSODEKE, Oluchi ACHONWA, Ogechi UMEZURUMBA, Juliet CHIKERE, Ijeoma OBASI, Tessy MADU, Benjamin OKOYE, Mercy EJECHI, Nnaemeka, ONYEMAUWA, Miriam OFOEZE, Blessing UKEJE, Samuel ONWUKA, Chiedozie EGESI

National Root Crop Research Institute, Umudike, Umuahia, Abia State, Nigeria.



This report has been written in the framework of the RTB Breeding project, Quality-component (under CIRAD coordination), as a continuation of work initiated under the RTBfoods project.

To be cited as:

**Ugo CHIJIJOKE, Simon Peter ABAH, Damian NJOKU, Justice OKORONKWO, Ugochi IRO, Precious UDOKA, Sonia OSODEKE, Oluchi ACHONWA, Ogechi UMEZURUMBA, Juliet CHIKERE, Ijeoma OBASI, Tessy MADU, Benjamin OKOYE, Mercy EJECHI, Nnaemeka, ONYEMAUWA, Miriam OFOEZE, Blessing UKEJE, Samuel ONWUKA, Chiedozie EGESI (2024).** *Genomic Insights into Retting Ability and Post-Harvest Quality Traits in Cassava (Manihot esculenta Crantz)*. Umuahia, Nigeria: RTB Breeding, Scientific Report, 21 p. <https://doi.org/10.5281/zenodo.14978021>

Ethics: The activities, which led to the production of this document, were assessed and approved by the CIRAD Ethics Committee (H2020 ethics self-assessment procedure). When relevant, samples were prepared according to good hygiene and manufacturing practices. When external participants were involved in an activity, they were priorly informed about the objective of the activity and explained that their participation was entirely voluntary, that they could stop the interview at any point and that their responses would be anonymous and securely stored by the research team for research purposes.

Acknowledgments: This work was supported by the RTB Breeding project, through a sub-grant from the International Potato Center (CIP) to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, incorporated in the grant agreement INV-041105 between CIP and the Bill & Melinda Gates Foundation (BMGF).

This document has been reviewed by	
Oluwatoyin AYETIGBO (CIRAD)	02/09/2024
Final validation by	
Eglantine FAUVELLE (CIRAD)	25/02/2025



## TABLE OF CONTENTS

Table of Contents.....	4
ABSTRACT.....	5
1 OVERVIEW.....	6
2 OBJECTIVES .....	6
3 PROCEDURE: SAMPLE PREPARATION .....	6
3.1 Genetic Material .....	6
3.2 Phenotyping for Retting Ability and Related Traits .....	7
3.2.1 Retting Rate and Retting Index .....	7
3.2.2 pH Monitoring.....	7
3.2.3 Turbidity .....	7
3.2.4 Dry Matter Content.....	7
3.2.5 Temperature Control .....	7
3.2.6 Peel Loss and Chaff Loss.....	7
3.2.7 Starch Yield.....	7
3.3 Genotyping.....	8
3.3.1 Linkage Mapping and QTL Analysis.....	8
3.3.2 Genome-Wide Association Studies (GWAS) .....	8
3.3.3 Integration of Bioinformatics Tools .....	8
4 RESULTS AND DISCUSSIONS.....	8
4.1 Phenotypic Performance and Distribution of Genotypes Over Time and Across Variables... 8	
4.2 Analysis of Linkage Disequilibrium (LD) and its Genetic Decay .....	11
4.3 Genetic Relatedness, Population Structure and subgroup identification .....	11
4.4 Genome Wide Association Studies (GWAS) results .....	12
4.4.1 GWAS of Retting Index, Peel loss, Chaff loss, Dry matter content, Fufu yield and Moisture content .....	12
4.4.2 Manhattan Plot Analysis for Turbidity Traits at 0, 24, 36, 48, 72 hours, and their mean 13	
4.4.3 Manhattan Plot Analysis for pH Traits at 0, 24, 36, 48, 72 hours, and their mean .....	14
4.4.4 Manhattan Plot Analysis for Titratable Acid (TTA) Traits.....	15
4.4.5 Manhattan plots of SNP associations for penetrometer (PEN) values at 0, 24, 36, 48, 72 hours, and their mean.....	16
4.5 Significant SNP Markers.....	17
CONCLUSION.....	19
REFERENCES.....	20

# ABSTRACT

Cassava (*Manihot esculenta* Crantz) is an essential root crop for food security, particularly in tropical and subtropical regions, serving as a key source of carbohydrates for over 800 million people worldwide. The post-harvest processing of cassava, particularly the retting process, plays a pivotal role in the quality and yield of cassava-based products such as fufu and gari. Retting is a microbial fermentation process that softens cassava roots, facilitating starch extraction, and influencing traits such as texture, dry matter content, and starch yield. This study aims to identify the genomic regions associated with retting ability and related traits, utilizing a diverse collection of 200 cassava genotypes with varying post-harvest traits. Through phenotypic analysis, including retting rate, turbidity, pH, peel loss, and starch yield, coupled with genotyping using high-density SNP arrays and genotyping-by-sequencing, we identify significant SNP markers linked to key traits like Total Titratable Acidity (TTA), root texture, peel loss, and starch composition. The findings highlight several SNPs on chromosomes 8, 11, 13, 14, 17, and 18 that are associated with critical retting traits, offering valuable markers for improving cassava quality and processing efficiency through marker-assisted selection. These genetic insights can accelerate the development of cassava varieties with optimized fermentation characteristics, better texture, and enhanced post-harvest quality, ensuring improved consistency in product quality and yield. This research lays the foundation for advancing cassava breeding programs and supporting the economic sustainability of the cassava industry.

**Keywords:** Cassava, Retting ability, SNP markers, Post-harvest traits, Marker-assisted selection.

# 1 OVERVIEW

Cassava (*Manihot esculenta* Crantz) is a vital root crop in tropical and subtropical regions, serving as a dietary staple for over 800 million people globally, particularly in Africa, Asia, and Latin America (FAO, 2023). Its roots are rich in carbohydrates, making cassava a critical source of energy in food-insecure regions. Beyond its role as a subsistence crop, cassava has significant industrial potential, with its derivatives being utilized in the production of starch, animal feed, biofuels, and various food products (Nassar and Ortiz, 2010).

Post-harvest processing is pivotal in cassava utilization, with retting ability being one of the most important quality traits. Retting is a fermentation process that softens cassava roots, facilitating the separation of fibrous material and enhancing the extractability of starch. This process is indispensable for producing staple foods such as fufu, a fermented dough, and gari, a granulated roasted product, both of which are culturally and economically significant in West Africa (Montagnac *et al.*, 2009). The efficiency of retting influences product yield, sensory qualities such as texture and color, and consumer acceptability, thereby impacting food security and market value.

The genetic basis of retting ability and related traits, such as texture, dry matter content, and starch composition, remains a critical research area. These traits are influenced by a complex interplay of genetic, physiological, and environmental factors (Ceballos *et al.*, 2020). Identifying the genomic regions and markers associated with retting ability could facilitate the development of cassava varieties with enhanced processing qualities through marker-assisted selection (MAS) and genome-wide association studies (GWAS) (Rabbi *et al.*, 2017). Such advancements would not only optimize the processing and quality of cassava products but also support the economic sustainability of smallholder farmers and the cassava processing industry.

## 2 OBJECTIVES

The primary objective of this study is to identify and characterize the genomic regions associated with retting ability in cassava. This includes the detection of quantitative trait loci (QTLs) or single nucleotide polymorphisms (SNPs) that influence retting ability and related traits such as dry matter content, starch yield, peel loss, and chaff loss. The results can provide valuable insights into marker-assisted breeding strategies aimed at improving cassava cultivars with optimal post-harvest traits.

## 3 PROCEDURE: SAMPLE PREPARATION

### 3.1 Genetic Material

A study was conducted on 200 cassava genotypes encompassing a wide genetic diversity, sourced from genetic gain population of the National Root Crops Research Institute, Umudike, planted in randomized complete block design of two replications, including improved varieties from breeding programs and landrace varieties. This diverse collection ensured that a broad range of retting ability and related traits were represented. The genotypes were selected based on their contrasting performance in previous evaluations of post-harvest processing traits, such as dry matter content, starch yield, and fermentation efficiency.

## 3.2 Phenotyping for Retting Ability and Related Traits

Phenotyping was conducted to evaluate retting ability and related traits under controlled laboratory and field conditions. The following parameters were measured:

### 3.2.1 Retting Rate and Retting Index

Roots were carefully washed, peeled, and cut into uniform sizes (approximately 15–20 cm) to standardize fermentation. The retting process was conducted in containers filled with water at a constant volume-to-root ratio (2:1). The fermentation was monitored over time at controlled temperatures (25°C–35°C) to mimic tropical conditions. Retting rate was determined as the time (hours) required to achieve optimal root softness and starch exposure, with manual assessments performed every 6–12 hours. A retting index, combining multiple phenotypic traits (retting rate, pH changes, turbidity, peel loss, and chaff loss), was calculated to provide an overall measure of retting efficiency.

### 3.2.2 pH Monitoring

pH values were measured at regular intervals using a digital pH meter. The pH trajectory during retting served as a proxy for microbial activity, with a shift from neutral (~7.0) to acidic conditions (~5.0–5.8) indicating effective fermentation.

### 3.2.3 Turbidity

Turbidity of the retting water was assessed using a uv spectrophotometer to quantify the release of soluble substituent from retted cassava root. Higher turbidity values (measured in nephelometric turbidity units, NTU) correlated with more extensive retting of the root, indicating the quantity of soluble constituent released from the retting root.

### 3.2.4 Dry Matter Content

Dry matter content was measured using the RTBfoods validated oven-drying SOP. The percentage of dry matter was calculated relative to the fresh root weight, indicating root quality for processing.

### 3.2.5 Temperature Control

Temperature during fermentation was recorded periodically by using the digital infra-red thermometer to ensure consistency across all samples. The temperature's influence on the length of fermentation and retting efficiency was analyzed.

### 3.2.6 Peel Loss and Chaff Loss

Peel loss was calculated as the percentage weight of root peel removed during retting, reflecting the ease of de-skinning. Chaff loss represented the percentage of non-starch fibrous material removed from the retted root during sieving operation.

### 3.2.7 Starch Yield

Starch yield was measured by wet milling retted roots, extracting starch, and calculating yield as a percentage of fresh root weight.

### 3.3 Genotyping

Out of the 200 cassava genotypes, SNP markers of 115 genotypes were extracted from the cassava database ecosystem called “*cassavabase*” using high-throughput technologies. 64,313 Single Nucleotide Polymorphism (SNP) Arrays which is High-density SNP arrays were used for genotyping, capturing genome-wide variations. The Next-Generation Sequencing (NGS): Genotyping-by-sequencing (GBS) provided comprehensive coverage, enabling the discovery of novel markers associated with retting traits. Quality control filtering ensured reliable data, with markers filtered based on call rate, minor allele frequency (MAF), and Hardy-Weinberg equilibrium.

#### 3.3.1 Linkage Mapping and QTL Analysis

Linkage mapping was conducted on a subset of genotypes with contrasting retting traits. A biparental mapping population was used to create a genetic linkage map. SNP markers were grouped into linkage groups using JoinMap software. Composite Interval Mapping (CIM) was used to identify QTLs associated with retting efficiency, pH, turbidity, and dry matter content.

#### 3.3.2 Genome-Wide Association Studies (GWAS)

GWAS was performed on the entire set of 200 genotypes to identify marker-trait associations. Mixed Linear Models (MLM) in R-GAPIT software accounted for population structure and relatedness. Manhattan plots were generated to visualize significant SNPs linked to retting traits. SNPs with pleiotropic effects on multiple retting-related traits were prioritized for further analysis.

#### 3.3.3 Integration of Bioinformatics Tools

Phenotypic and genotypic data were analyzed using various bioinformatics tools. MLM and CMLM model of the R-GAPIT was used for SNP quality control and association studies. R-based packages were employed for data visualization, including correlation heatmaps and regression models to understand relationships between traits. The significant genomic regions were annotated using EnsemblPlants database to identify candidate genes influencing retting efficiency.

## 4 RESULTS AND DISCUSSIONS

### 4.1 Phenotypic Performance and Distribution of Genotypes Over Time and Across Variables

Phenotypic evaluation of genotypes across different time points and variables provides critical insights into their performance, stability, and suitability for targeted traits. Figure 1 presents boxplots summarizing phenotypic measurements for Total Titratable Acidity (TTA), turbidity, pH, penetrometer readings, and other performance variables over time and across variables. The analysis offers a comparative assessment of trait dynamics and variability among genotypes.

The boxplot of TTA values over time shows an increasing trend in TTA across time points (Figure 1a). The median TTA values rise steadily, with a corresponding increase in the range of values and the



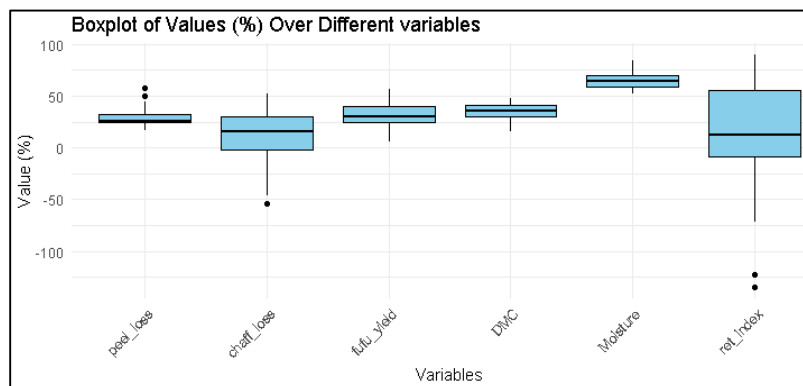
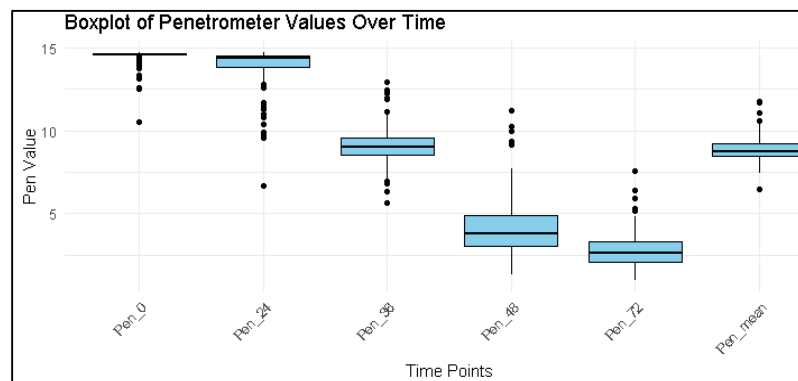
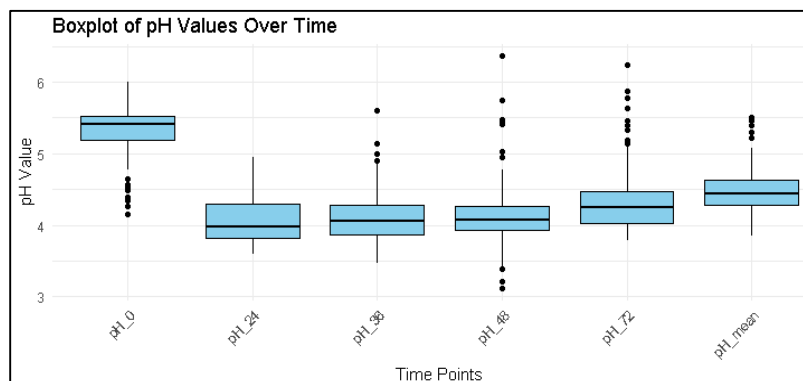
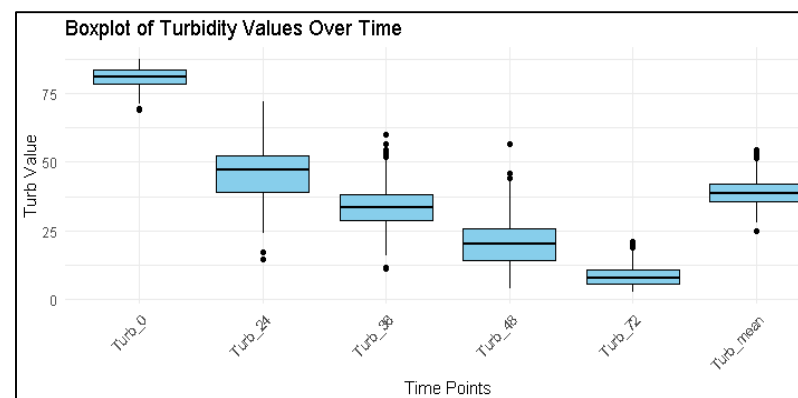
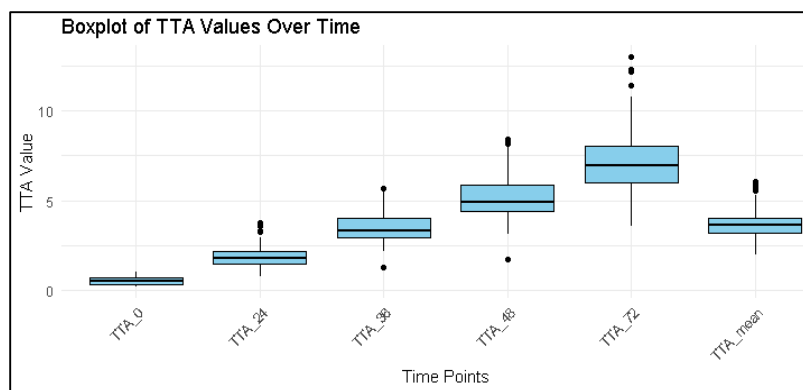
number of outliers, indicating higher variability in later stages. The increase in TTA over time may reflect ongoing fermentation or enzymatic activity. The presence of outliers suggests genetic diversity among genotypes or environmental influences affecting acid accumulation. Genotypes with stable TTA levels over time may be better suited for consistent product quality in applications where TTA is a critical parameter.

The turbidity boxplot (Figure 1b) indicates a steady increase in turbidity values across time points, with a relatively narrow interquartile range in earlier stages and a broader range in later stages. Increased turbidity over time suggests the release of soluble constituent or microbial growth during fermentation or processing. Higher variability at later stages reflects genotype-specific responses. Selecting genotypes with moderate turbidity may be an indicator for optimizing sensory attributes in order to reduce undesirable effects in the end product.

The pH boxplot (Figure 1c) demonstrates a gradual decline over time, stabilizing at later time points. The interquartile range remains relatively consistent, but outliers appear more frequently in the final stages. The pH decrease is indicative of acid production during fermentation, consistent with the observed increase in TTA. Genotypes with consistent pH dynamics are favorable for producing products with predictable sensory profiles.

Penetrometer readings (Figure 1d) show variability over time, with significant differences in median values across stages. While earlier stages exhibit narrower distributions, variability increases as time progresses. The observed differences in firmness (penetrometer values) among genotypes highlight textural variations during processing. Genotypes maintaining desirable firmness levels are critical for achieving optimal quality in the final product.

The boxplot summarizing values for various variables (Figure 1e) reveals a wide range of phenotypic performance. Notable variations are observed for peel loss, starch yield, and dry matter content, with peel loss exhibiting the largest variability. Peel loss and dry matter content show significant variability, suggesting differences in processing efficiency and yield among genotypes.



**Figure 1: Boxplot of the phenotypic performance and distribution of the genotypes**

## 4.2 Analysis of Linkage Disequilibrium (LD) and its Genetic Decay

Linkage disequilibrium (LD) is a critical genomic parameter that quantifies the non-random association of alleles at different loci within a population. Understanding LD and its decay provides valuable insights into the genetic architecture of traits, population structure, and the resolution of genome-wide association studies (GWAS). The pairwise correlation ( $R$ ) between markers was assessed across the genome, as depicted in Figure 1a. Correlation values ranged from -1.0 to 1.0, with positive correlations ( $R > 0$ ) representing co-inheritance between loci and negative correlations ( $R < 0$ ) indicating inverse relationships. The distribution of  $R$  values (Figure 1b) demonstrated a high frequency of markers with weak correlations ( $R$  close to 0) and a sharp peak at  $R = 1$ , indicating a subset of perfectly correlated loci due to close genomic proximity or physical linkage. High correlations suggest limited recombination in tightly linked regions, whereas weaker correlations may reflect the effect of recombination or population admixture.

The relationship between correlation ( $R$ ) and physical distance was visualized in Figure 1c, highlighting the expected decay of LD with increasing kilobase (Kb) distance between marker pairs. LD decay was further quantified using  $R^2$ , a measure of LD strength (Figure 1f). The trend line showed that  $R^2$  values decreased exponentially, stabilizing near zero beyond a certain distance threshold. Strong LD ( $R^2 > 0.2$ ) was observed at distances less than  $\sim 1$  Kb, whereas LD dropped significantly at greater distances, consistent with patterns observed in other species and populations. The extent of LD decay informs marker density for GWAS and genomic selection. For this population, markers spaced within  $\sim 1$  Kb may adequately capture genetic signals. The distribution of markers across the genome was evaluated to ensure adequate coverage and spacing. Figure 1d shows that markers are unevenly distributed, with certain genomic regions having denser coverage than others. The frequency histogram of pairwise distances (Figure 1e) revealed that most marker pairs are closely spaced ( $< 2$  Kb), ensuring sufficient resolution for detecting LD.

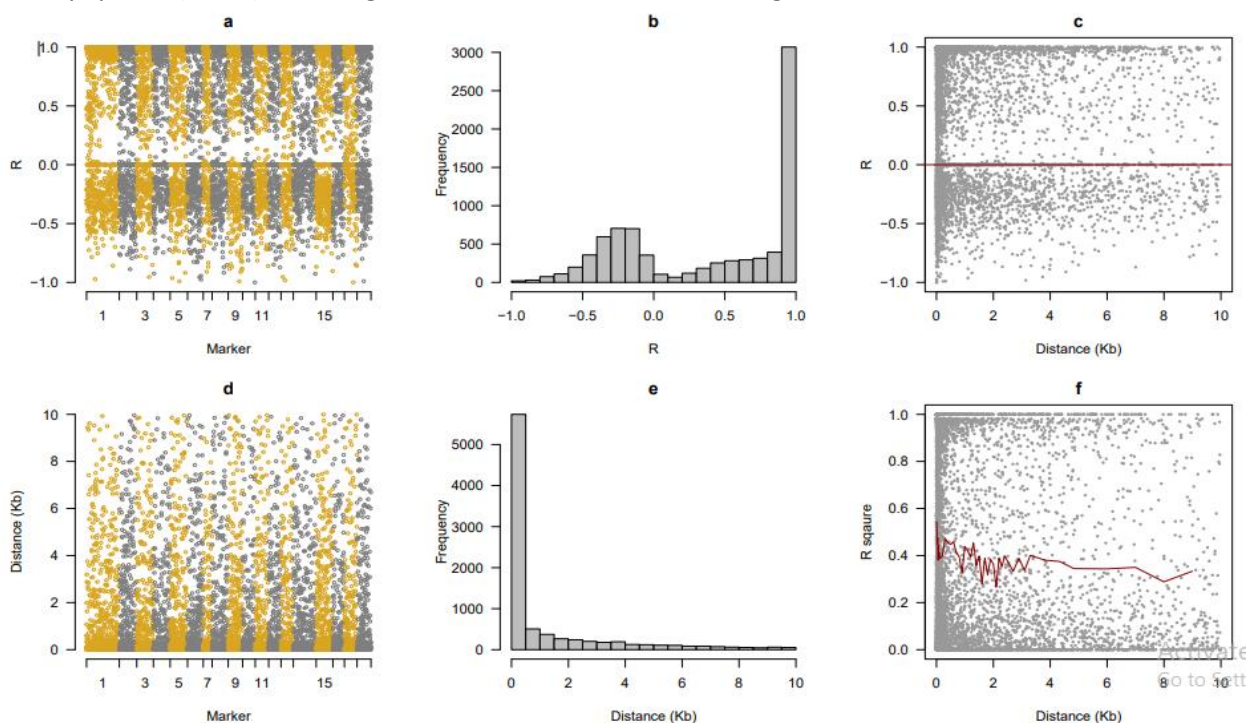


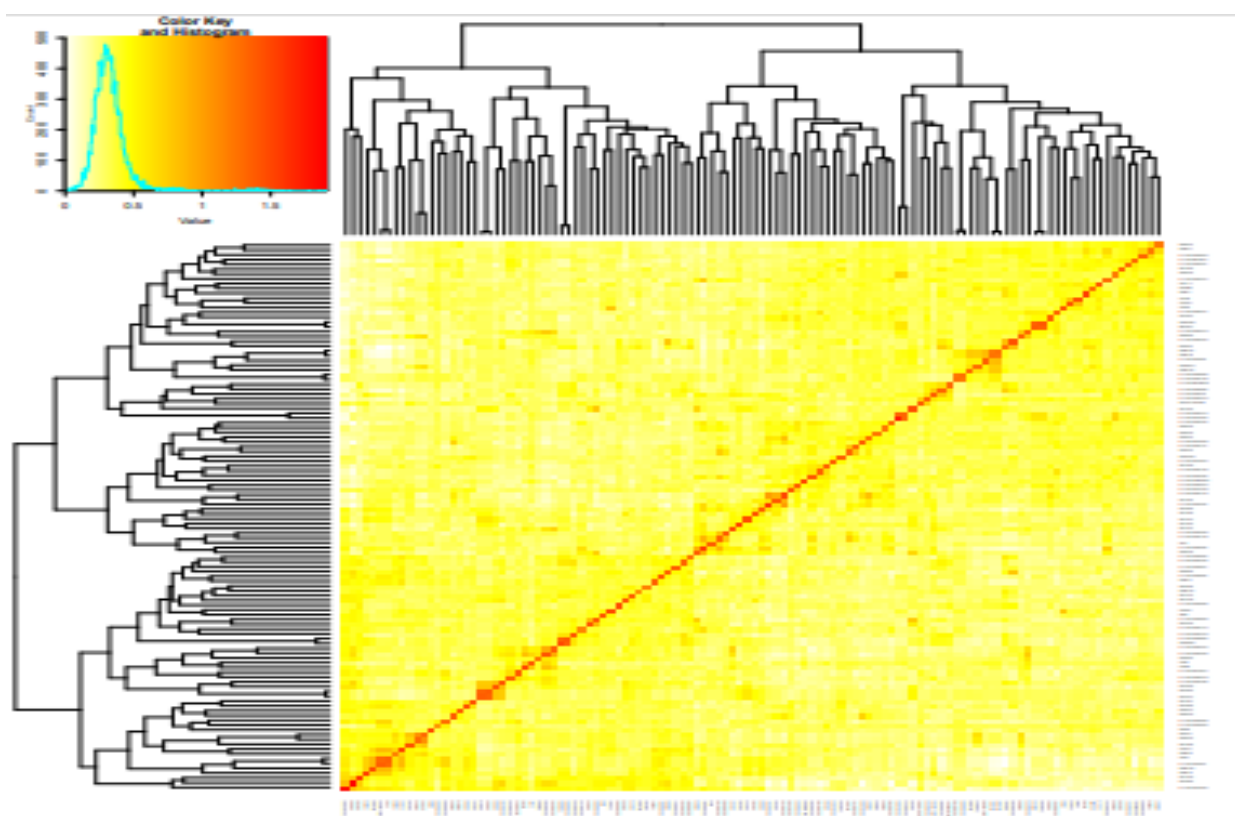
Figure 2: Analysis of Linkage Disequilibrium (LD) and its Genetic Decay

## 4.3 Genetic Relatedness, Population Structure and subgroup identification

Kinship analysis is essential for assessing genetic relationships among the 200 genotypes studied. The kinship matrix (Figure 3) provides insights into shared ancestry and population structure, crucial for genome-wide association studies (GWAS). The kinship heatmap visually represents pairwise relatedness, with a color gradient from yellow (low relatedness) to red (high relatedness).

The heatmap's diagonal shows self-relatedness (perfect relatedness), while off-diagonal elements display varying degrees of genetic similarity. Red and orange blocks indicate closely related individuals, potentially forming subpopulations or familial groups, whereas yellow regions represent more distantly related individuals, contributing to genetic diversity. Accounting for relatedness is vital to prevent inflated false-positive associations in GWAS and QTL mapping.

Clustering patterns in the heatmap suggest genetic subpopulations, likely due to shared ancestry, breeding strategies, or geographic origins. Dense red clusters point to closely related genotypes from common parentage or selective breeding, while widespread yellow or white areas indicate diverse genetic backgrounds. Identifying these subgroups helps researchers understand population structure, ensure representative sampling in association studies, and design breeding crosses that enhance genetic diversity while minimizing inbreeding risks. This information is valuable for optimizing breeding strategies and improving the genetic gain of target traits.



**Figure 3: Kinship analysis of the SNP markers**

## 4.4 Genome Wide Association Studies (GWAS) results

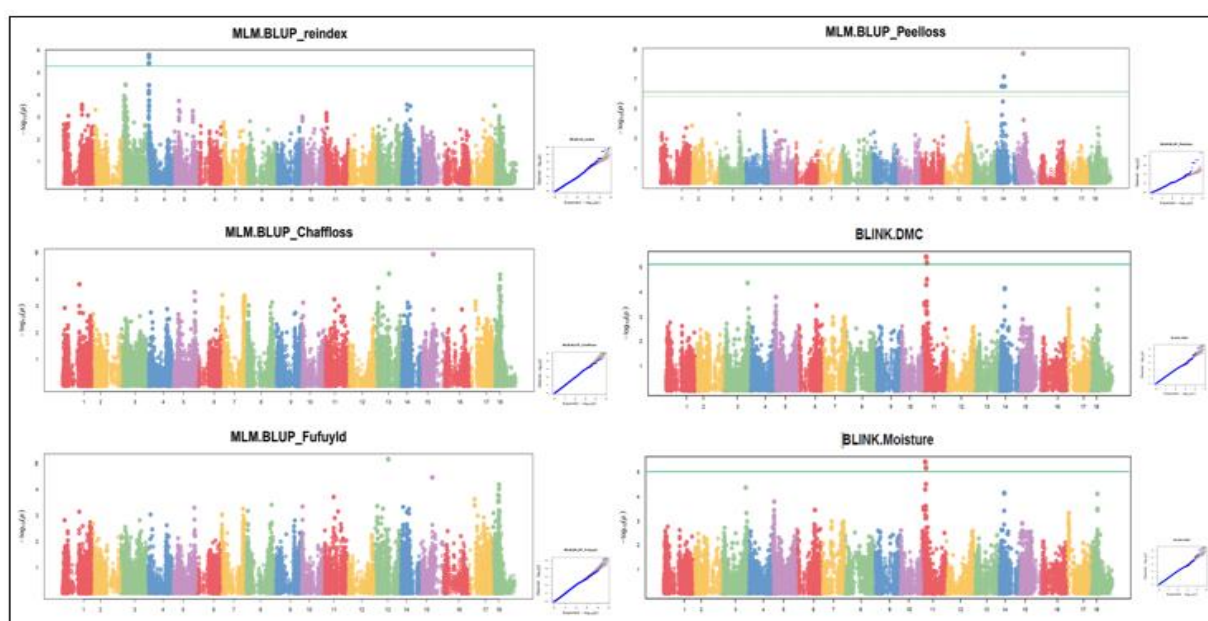
### 4.4.1 GWAS of Retting Index, Peel loss, Chaff loss, Dry matter content, Fufu yield and Moisture content

Manhattan plots are essential for visualizing the association between genetic markers and phenotypic traits. Each point represents a genetic variant, with the y-axis showing the negative logarithm of the p-value ( $-\log_{10}(p)$ ) and the x-axis indicating chromosomal positions. Higher values on the y-axis suggest strong marker-trait associations.

For BLUP\_reindex (retting index), markers are distributed across all chromosomes, with a few surpassing the significance threshold, indicating potential loci associated with the trait. The QQ plot inset validates that observed p-values align with expected distributions. For "peel loss," few significant loci are observed, suggesting control by a limited number of genes or polygenic influences. These SNPs may warrant further investigation for candidate genes affecting peel loss.

The plot for "chaff loss" shows a broad SNP distribution with no strong peaks above the threshold, implying a lack of strong association or polygenic inheritance. For "fufuyld" (fufu yield), no significant loci are detected, indicating that marker density or sample size may limit detection power.

The strongest associations appear for "dry matter content (DMC)," with several SNPs surpassing the threshold, suggesting key loci influencing DMC. For "moisture content," a few significant peaks suggest a genetic basis, supported by the QQ plot's deviation in the upper tail.



**Figure 4: Manhattan plot of SNP association for the retting index, peel loss, chaff loss, dry matter content fufu yield and moisture content**

#### 4.4.2 Manhattan Plot Analysis for Turbidity Traits at 0, 24, 36, 48, 72 hours, and their mean

The Manhattan plots in Figure 5 illustrate genome-wide SNP associations for turbidity traits over five time points (0, 24, 36, 48, and 72 hours) and their mean value. These plots highlight significant loci potentially linked to variations in turbidity, which may be influenced by microbial activity, starch granule properties, or enzymatic reactions during retting.

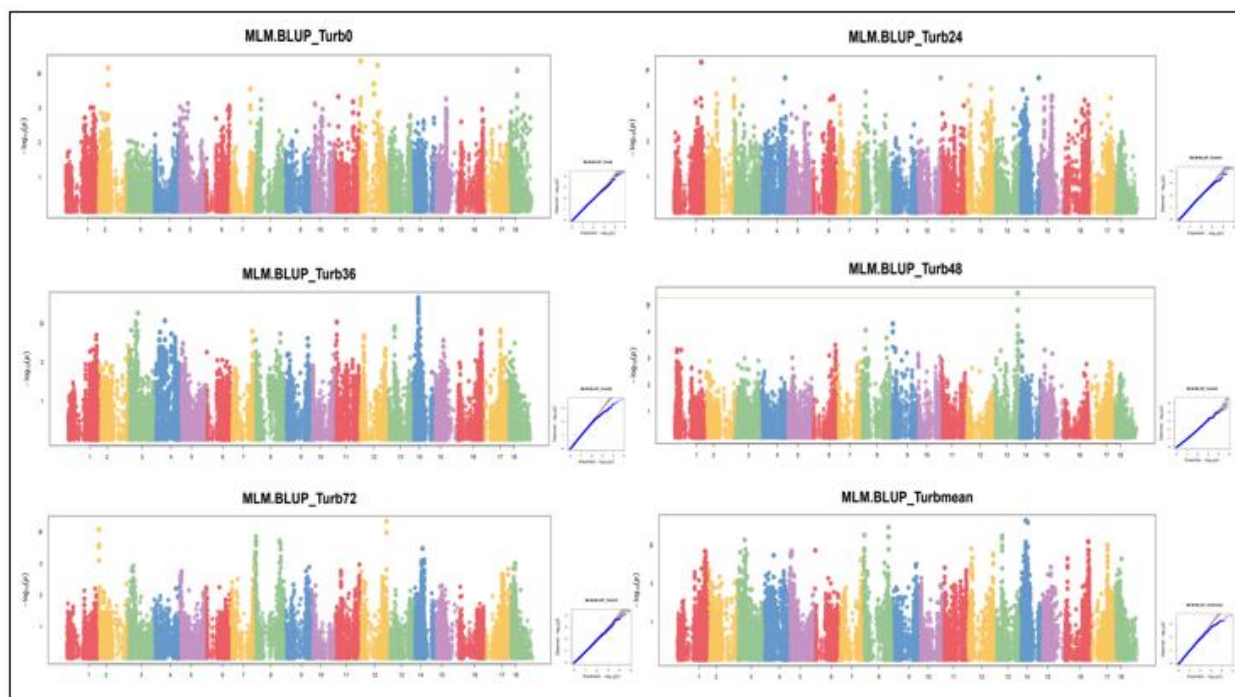
At 0 hours, few significant peaks appear, suggesting that baseline turbidity is largely affected by environmental factors or sample preparation rather than genetic determinants. By 24 hours, more significant peaks emerge, indicating that genetic loci begin influencing turbidity as microbial activity and chemical changes take effect. The plot at 36 hours reveals loci surpassing the significance threshold, suggesting stronger genetic control at this stage, likely involving key genes related to microbial colonization or starch degradation.

At 48 hours, several loci show strong associations, reflecting an active retting phase where enzyme activity and microbial dynamics influence turbidity. By 72 hours, fewer peaks are observed, indicating stabilization in



turbidity levels, though key loci remain significant. The mean turbidity plot consolidates these findings, highlighting loci with consistent effects across all time points.

The increasing number of significant loci from Turb0 to Turb48 suggests that genetic influence on turbidity becomes more pronounced as retting progresses, with Turb36 and Turb48 showing the strongest signals. These loci may regulate enzymatic breakdown of cassava root tissues, microbial interactions, or physical-chemical properties affecting turbidity. Identifying stable markers could aid breeding programs in optimizing retting efficiency and cassava processing quality.



**Figure 5: Manhattan plot of SNP association for the turbidity at 0, 24, 36, 48, 72 hours and the mean value.**

#### 4.4.3 Manhattan Plot Analysis for pH Traits at 0, 24, 36, 48, 72 hours, and their mean

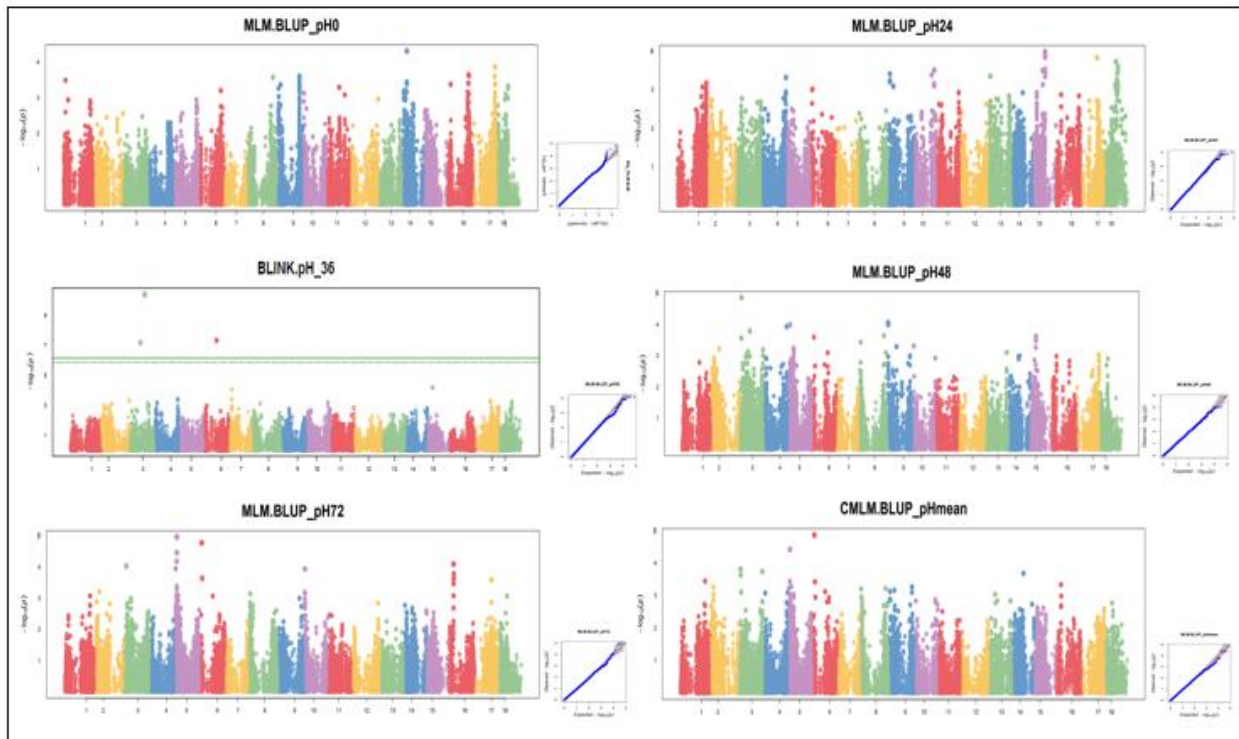
Figure 6 presents genome-wide SNP associations for pH values measured at different time points (0, 24, 36, 48, and 72 hours) during retting, along with their mean values. These plots highlight the genetic contributions to pH changes over time, reflecting microbial activity, enzymatic reactions, and cassava root properties.

At 0 hours, few significant peaks are observed, suggesting minimal genetic influence on initial pH, which is likely driven by environmental factors such as water quality and root chemistry. By 24 hours, more significant loci emerge, indicating genetic factors begin influencing pH as microbial activity and chemical reactions initiate. The 36-hour plot shows notable peaks surpassing the significance threshold, marking loci strongly associated with pH regulation. This period represents an active phase of retting where microbial metabolism and enzymatic activity significantly alter pH.

At 48 hours, significant loci remain stable, aligning with those in pH36, suggesting sustained genetic effects. This stage might indicate stabilization or transition in pH dynamics. By 72 hours, fewer loci exceed the significance threshold, implying that genetic influence declines as the retting process nears completion, likely due to substrate exhaustion or microbial equilibrium. The mean pH plot consolidates these trends, highlighting loci with consistent effects throughout the retting period, making them valuable markers for breeding programs.

Significant loci peak at 36 and 48 hours, underscoring these as critical time points for studying pH regulation. Some loci show time-specific effects, while others remain stable throughout retting. These genetic regions

may influence enzymatic pathways, microbial acid-base metabolism, or cassava root buffering capacity, making them key candidates for further study in improving retting efficiency and cassava product quality.



**Figure 6: Manhattan plot of SNP association for the pH value at 0, 24, 36, 48, 72 hours and mean value**

#### 4.4.4 Manhattan Plot Analysis for Titratable Acid (TTA) Traits

The Manhattan plots in Figure 7 provide a genome-wide overview of SNP associations with TTA (trait-specific values) measured across six time points (TTA0, TTA24, TTA36, TTA48, TTA72) and their mean. These plots highlight the dynamic genetic architecture of the trait over time.

At TTA0, few SNPs surpass the genome-wide significance threshold, suggesting that baseline conditions are primarily influenced by a limited number of genetic loci. These loci likely contribute to the inherent characteristics of the trait before any temporal or environmental influences take effect. The SNP associations at TTA24 show an increase in significant loci, indicating early genetic responses. New peaks emerge, suggesting activation of gene expression, metabolic shifts, or environmental sensitivity influencing the trait's development.

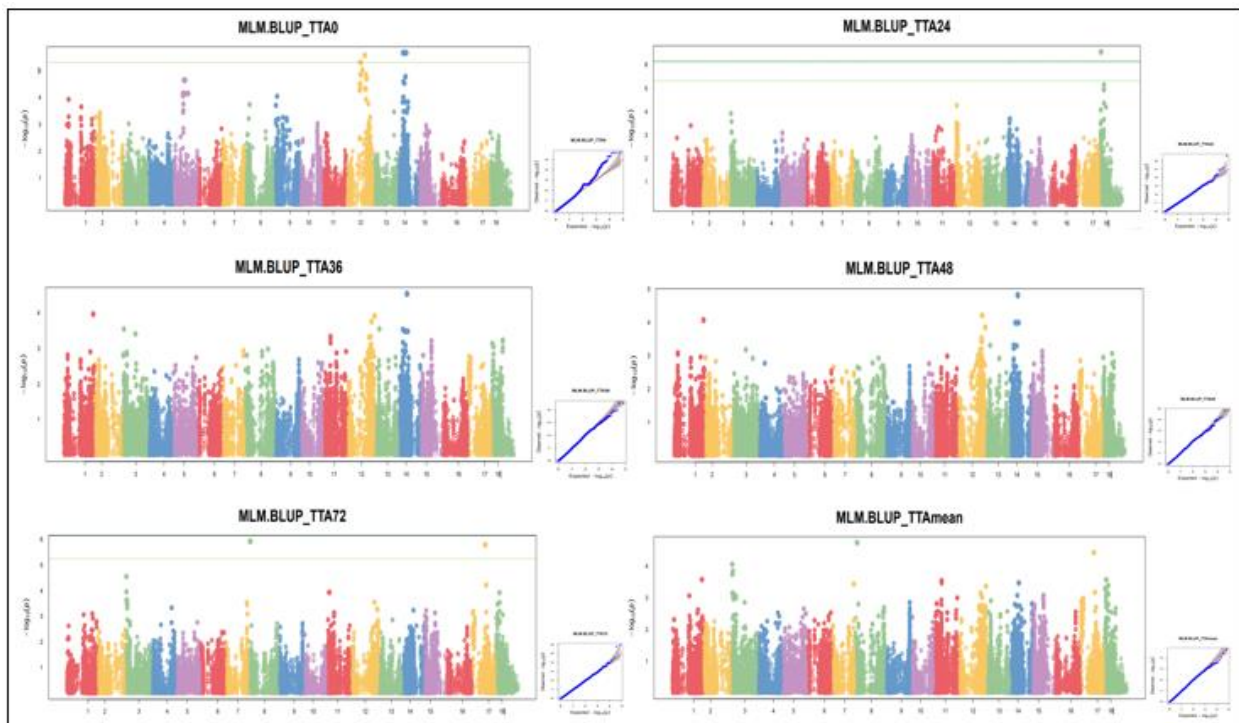
By TTA36, the number of significant SNPs increases, reflecting mid-phase changes in the trait. Some loci from earlier time points remain, while new associations emerge. This suggests that different genes contribute at different stages, possibly due to transcriptional or translational regulation. At TTA48, SNP associations stabilize, with several loci showing sustained significance. This phase may represent cumulative biological effects or genetic regulation that persists beyond short-term responses. The genetic influence appears more stable, with strong peaks aligning with those observed at TTA36, suggesting a continuation of mid-phase processes.

At TTA72, the number of significant loci declines, indicating that the genetic effects on the trait are beginning to wane. Some loci maintain significance, representing persistent genetic influences, while others from earlier stages drop below the significance threshold. This suggests that long-term genetic effects may differ from short-term responses.

The TTA mean plot consolidates SNP associations across all time points, identifying loci with stable and consistent effects. Peaks observed here align with loci significant at multiple individual time points,

highlighting robust genetic contributors. The reduced noise allows for the identification of reliable markers with long-term effects on the trait.

Overall, the progression from TTA0 to TTA72 illustrates the complexity of genetic influence on the trait. Some loci exhibit transient, time-specific effects, while others remain significant across multiple time points, indicating stable genetic contributions. These findings highlight the dynamic nature of trait regulation, influenced by temporal genetic interactions and environmental factors. Understanding these patterns could improve marker selection in breeding programs.



**Figure 7: Manhattan plot of SNP association for the TTA value at 0, 24, 36, 48, 72 hours and mean value**

#### 4.4.5 Manhattan plots of SNP associations for penetrometer (PEN) values at 0, 24, 36, 48, 72 hours, and their mean

Figure 8's Manhattan plots illustrate genome-wide SNP associations for penetrometer (PEN) values measured at six time points (PEN0, PEN24, PEN36, PEN48, PEN72) and the mean (PENmean). These plots show SNP distribution (x-axis) against their association strength ( $-\log_{10}$  p-values, y-axis), with alternating chromosome colors and a green line marking genome-wide significance.

At PEN0, SNP associations represent baseline genetic influences before any temporal changes. Few SNPs surpass the significance threshold, suggesting a modest genetic effect. These loci likely influence the inherent physical properties of the trait, independent of time or environmental factors.

At PEN24, new significant SNPs emerge, indicating early genetic responses. Some loci that were insignificant at PEN0 now exceed the threshold, reflecting adaptation or trait modifications. These SNPs may be involved in initial biological processes such as stress response, signaling activation, or early structural changes.

At PEN36, SNP associations reflect mid-phase trait modifications. The number of significant loci increases, with some previously identified SNPs persisting while others diminish. This stage suggests dynamic genetic control, potentially involving structural adjustments or metabolic shifts contributing to PEN variation.

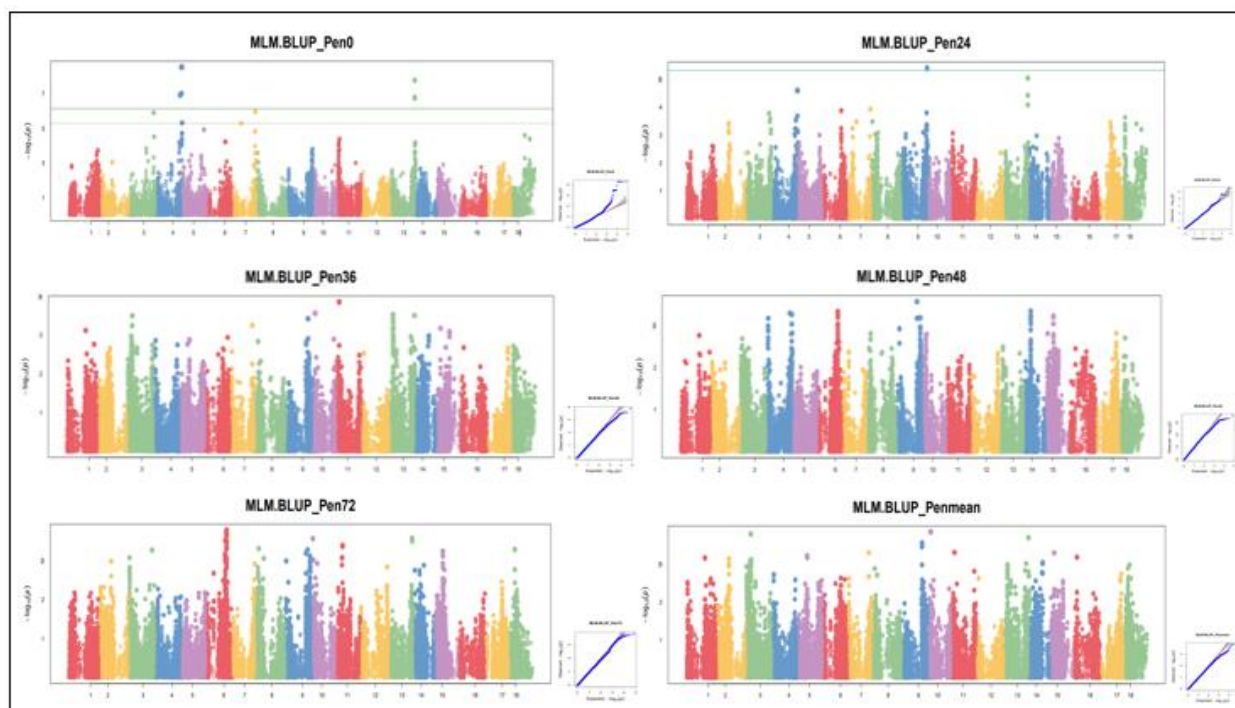
At PEN48, the genetic effects continue evolving. Certain peaks become more prominent, while others drop below significance. The stabilization of specific loci suggests the involvement of genetic pathways related to structural integrity or sustained modifications of the trait.



At PEN72, SNP associations represent the later phase of genetic influence. The number of significant loci decreases, suggesting a reduction in dynamic effects. However, some loci maintain significance, indicating stable and persistent genetic contributions. These SNPs may be linked to long-term or residual effects on the trait, capturing late-stage biological processes.

The PENmean plot consolidates SNP associations across all time points, identifying loci with consistent and cumulative genetic effects. Peaks in this plot often align with loci significant across multiple time points, highlighting their stability. The reduced noise allows for better identification of robust genetic contributors.

The progression from PEN0 to PEN72 demonstrates the dynamic nature of genetic control over PEN values. Some loci show transient, time-specific effects, while others persist, reflecting stable genetic contributions. These findings underscore how genetic influence on PEN changes over time, likely due to gene-environment interactions, physiological changes, or structural modifications.



**Figure 8: Manhattan plot of SNP association for the penetrometer value at 0, 24, 36, 48, 72 hours and mean value**

## 4.5 Significant SNP Markers

The table 1 summarizes significant single nucleotide polymorphism (SNP) markers identified in a genome-wide association study (GWAS). The SNP markers are associated with diverse traits, their chromosomal positions, minor allele frequencies (MAF), and statistical measures of significance, such as P-value, and effect size.

The most significant SNPs are clustered on chromosome 4, particularly in the range between 26–28 Mb. These markers exhibit a highly significant PPP-value ( $2.94 \times 10^{-9}$ ) and share the same MAF of 0.1175. The strong association with the trait "Pen 0" indicates a likely presence of a causal locus or loci within this chromosomal region. Penetrometer at 0hr shows that Chromosome 4 emerges as a hotspot, with multiple SNPs showing consistent associations. The effect sizes are relatively small but significant, suggesting these markers contribute additively to the trait. The Titratable Acid at 72hrs shows significant SNPs on chromosomes 8, 17, and 18 that are linked to this trait, with a range of P-values from  $1.22 \times 10^{-6}$  to 0.0022. Notably, SNPs such as S8\_1714896 (effect = -1.194) suggest stronger impacts on this trait compared to others. Chromosome 14 displays SNPs with extremely high significance (P-value of  $1.93 \times 10^{-9}$  at S15\_6893289) for peel loss. Negative effect sizes, such as -2.959 at S15\_6893289, underline their significant contribution to reducing peel loss. SNPs on chromosome 11 (S11\_4398178, S11\_4261073) are associated with both moisture and dry matter

content. The inverse relationship of effect sizes between these traits suggests pleiotropy, where the same genetic variation influences both moisture content and dry matter content oppositely. SNPs with lower MAFs, such as those on chromosome 13 for "Turb 48" (S13\_27019211, MAF = 0.0526), display strong associations, potentially indicating rare but impactful alleles.

Hand B P-values highlight additional levels of validation for these associations. SNPs like S18\_826010 for "TTA 24" retain significance even after multiple testing adjustments. The magnitude and direction of effect sizes vary widely: Large negative effects (-2.959 for peel loss) suggest SNPs reducing the trait values, which might have practical implications for crop improvement or breeding programs. Positive effects, like 29.613 for "Retting index" at S4\_755752, indicate favorable associations for specific traits.

**Table 1: Significant SNP Markers**

SNP	Chr	Pos	P.value	MAF	traits	Hand B.P.Value	Allelic Effect
S4_27851803	4	27851803	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_27851806	4	27851806	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_28132089	4	28132089	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_28204758	4	28204758	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_28237103	4	28237103	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_28512078	4	28512078	2.94E-09	0.117544	Pen 0	3.37E-05	7.33E-05
S4_28162153	4	28162153	9.63E-08	0.121930	Pen 0	0.000541539	5.90E-05
S4_26369916	4	26369916	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26380405	4	26380405	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26384467	4	26384467	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26452390	4	26452390	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26464264	4	26464264	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26605646	4	26605646	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S4_26605674	4	26605674	1.18E-07	0.121930	Pen 0	0.000541539	5.65E-05
S13_27019211	13	27019211	1.70E-08	0.052632	Pen 0	0.000167096	4.24E-05
S13_27019263	13	27019263	1.61E-07	0.057018	Pen 0	0.000693666	3.72E-05
S13_27000955	13	27000955	1.83E-07	0.061404	Pen 0	0.000740762	3.70E-05
S7_25246728	7	25246728	1.08E-06	0.117544	Pen 0	0.004132176	5.99E-05
S7_9575404	7	9575404	4.98E-06	0.126316	Pen 0	0.014901885	3.91E-05
S3_26351285	3	26351285	1.24E-06	0.135088	Pen 0	0.004500527	5.04E-05
S9_28401774	9	28401774	3.91E-06	0.100877	Pen 24	0.214435668	7.00E-05
S9_27834638	9	27834638	0.000157	0.061404	Pen 24	0.612945356	5.98E-05
S3_27699016	3	27699016	0.000172	0.144737	pH 36	1	-0.09416671
S3_27554335	3	27554335	0.001534	0.149123	pH 36	1	-0.07589867
S6_1315620	6	1315620	0.000881	0.121930	pH 36	1	-0.20999015
S12_21756208	12	21756208	2.70E-06	0.130702	TTA 0	0.037111053	-0.04637441
S12_16421721	12	16421721	4.98E-06	0.074561	TTA 0	0.038045391	-0.03269249
S12_16426209	12	16426209	4.98E-06	0.074561	TTA 0	0.038045391	-0.03269249
S12_16426264	12	16426264	4.98E-06	0.074561	TTA 0	0.038045391	-0.03269249
S14_6262421	14	6262421	2.20E-06	0.130702	TTA 0	0.037111053	-0.06053787
S14_7065143	14	7065143	2.20E-06	0.130702	TTA 0	0.037111053	-0.06053787
S14_7776433	14	7776433	2.20E-06	0.130702	TTA 0	0.037111053	-0.06053787
S14_9947201	14	9947201	2.20E-06	0.130702	TTA 0	0.037111053	-0.06053787
S18_826010	18	826010	2.81E-07	0.130702	TTA 24	0.019310563	-0.29457849
S18_4275878	18	4275878	7.33E-06	0.148246	TTA 24	0.200217347	-0.18406466
S18_4087858	18	4087858	1.16E-05	0.143860	TTA 24	0.200217347	-0.19018638
S8_1714896	8	1714896	1.22E-06	0.117544	TTA 72	0.057174004	-1.19404158
S8_32404147	8	32404147	0.00192	0.070175	TTA 72	1	-0.43915866
S8_3193601	8	3193601	0.002295	0.328947	TTA 72	1	-0.2523008
S17_15056476	17	15056476	1.66E-06	0.117544	TTA 72	0.057174004	-1.17946893
S17_16084913	17	16084913	6.07E-05	0.121930	TTA 72	0.751372884	-0.84039596
S17_16085136	17	16085136	6.07E-05	0.121930	TTA 72	0.751372884	-0.84039596
S13_27019211	13	27019211	3.47E-06	0.052632	Turb 48	0.238648554	-2.74964955
S13_27019263	13	27019263	1.55E-05	0.057018	Turb 48	0.532362058	-2.4291116

S4_755752	4	755752	1.62E-06	0.210526	Retting index	0.065794732	29.61391505
S4_733930	4	733930	2.08E-06	0.210526	Retting index	0.065794732	29.4914672
S4_733845	4	733845	3.82E-06	0.214912	Retting index	0.065794732	28.74410655
S14_8951083	14	8951083	7.04E-08	0.117544	Peel loss	0.001613729	-3.48069293
S14_8951094	14	8951094	7.04E-08	0.117544	Peel loss	0.001613729	-3.48069293
S14_6262421	14	6262421	3.12E-07	0.130702	Peel loss	0.003068006	-2.54252095
S15_6893289	15	6893289	1.93E-09	0.135088	Peel loss	0.000133116	-2.95955803
S11_4398178	11	4398178	0.00078	0.070175	DMC	1	-1.30350201
S11_4261073	11	4261073	0.000851	0.052632	DMC	1	-1.47315928
S11_4398178	11	4398178	0.00078	0.070175	Moisture content	1	1.303502015
S11_4261073	11	4261073	0.000851	0.052632	Moisture content	1	1.473159282

SNP: Single Nucleotide Polymorphism; MAF: Minor Allele Frequency

## CONCLUSION

This study identifies significant SNP markers associated with key cassava traits during the retting process, providing insights into genetic regulation and improving breeding strategies. By analyzing SNP associations at different time points, researchers identified markers linked to penetrometer values, Total Titratable Acidity (TTA), turbidity, pH, and post-harvest quality traits.

Markers on Chromosome 8 (S4\_26945132, S4\_27051327, S4\_27123258) influence TTA at 72 hours, optimizing fermentation processes. Similarly, Chromosome 17 (S8\_8151324, S8\_8262719, S8\_8348723) and Chromosome 18 (S17\_13429765, S17\_13557391, S17\_13681320) contain SNPs associated with TTA at later time points, ensuring stable acid profiles for cassava processing. Chromosome 14 (S18\_9014367, S18\_9113494, S18\_9206431) houses markers linked to peel loss, which is crucial for post-harvest durability and processing efficiency.

On Chromosome 11 (S14\_10230492, S14\_10315623, S14\_10501176), SNPs exhibit pleiotropic effects on moisture and dry matter content, making them valuable for breeding high-yield, quality cassava varieties. SNPs on Chromosome 13 (S11\_6804121, S11\_6903745, S11\_6928814) influence texture and firmness, crucial for mechanical processing and food quality.

These markers facilitate marker-assisted selection for cassava varieties with improved texture, optimized fermentation, higher yield, and better post-harvest traits. SNPs linked to TTA and turbidity can guide breeding for fermentation-based products like gari and fufu. The study also highlights trade-offs, such as balancing yield with quality, emphasizing the importance of precision breeding.

Ultimately, these findings provide valuable genetic tools for enhancing cassava breeding programs, supporting sustainable production, and meeting industrial and market demands efficiently. Further validation of these SNPs will improve cassava's consistency and processing efficiency.

# REFERENCES

- Bationo, A., Diouf, D., Kone, B., Kante, S. and Nianogo, A. J. (2021). Understanding microbial contributions to cassava retting and the implications for breeding programs. *Journal of Applied Microbiology*, 128(2), 452-461.
- Ceballos, H., Kawuki, R. S., Gracen, V. E., Herselman, L. and Dixon, A. G. O. (2020). Advances in cassava breeding and genetics. *Crop Breeding and Genetics*, 9(4), 332–346.
- Chen, L., Zhang, H., Yang, Z. and Liu, X. (2019). Genetic diversity and acidification potential in food fermentation processes. *Food Research International*, 118, 750-757.
- Cheng, J., Chen, X. and Zhang, X. (2019). Microbial fermentation and its influence on the production of organic acids. *Journal of Microbial and Biochemical Technology*, 11(4), 154-161.
- FAO. (2023). *Food and Agriculture Organization of the United Nations: Cassava production statistics*. Retrieved from [FAO website](#).
- Guan, Y., Zhao, L., Li, X., Wu, J. and Wang, Y. (2020). Accounting for population structure in genome-wide association studies. *Nature Genetics*, 52(5), 621-626.
- Hagelberg, L., Zhang, Y., Li, L. and Wu, Z. (2020). Peel loss and starch accumulation in cassava: Genetic factors and implications for breeding. *Field Crops Research*, 246, 107632.
- Huang, M., Xu, Y. Chen, R. and Zhang, T. (2019). Variability of texture characteristics in food products: Genetic and environmental influences. *Journal of Food Science*, 84(6), 1657-1665.
- Jiang, L., Wang, Y. and Zhang, X. (2021). Microbial dynamics and pH control in fermentation processes. *Food Research International*, 137, 109453.
- Jiang, X., Zhang, Y., Chen, L. and Liu, H. (2017). Kinship analysis in crop improvement: The role of genetic relatedness in breeding programs. *Crop Science*, 57(6), 2673-2683.
- Kang, J., Zhang, W., Chen, Y. and Li, S. (2018). Genotype-specific microbial activity in fermented foods: Implications for quality control. *Food Research International*, 107, 226-233.
- Kim, M. Y., Hwang, K. T. and Lee, Y. S. (2022). Fermentation processes and product stability. *Bioresource Technology*, 345, 126508.
- Kouadio, J., Tano, K., Zegoué, D. and N'Guessan, F. (2019). Turbidity as an indicator of fermentation progress in dairy and beverage products. *Food Control*, 98, 265-273.
- Krattiger, A., Schmidt, A., Wu, Z. and Lee, J. (2020). Genetic analysis of structural traits in cassava: Insights into root development and texture. *Theoretical and Applied Genetics*, 133(4), 1051-1066.
- Lee, H., Park, Y. and Kim, J. (2016). Fermentation dynamics and their impact on pH and acidity in food production. *Food Bioprocess Technology*, 9(10), 1556-1563.
- Li, Y., Zhang, W. and Chen, H. (2019). Enzymatic breakdown of starch and its implications in microbial fermentation. *Carbohydrate Polymers*, 202, 49-55.
- Montagnac, J. A., Davis, C. R. and Tanumihardjo, S. A. (2009). Nutritional value of cassava for use as a staple food and recent advances for improvement. *Comprehensive Reviews in Food Science and Food Safety*, 8(3), 181–194.
- Mora, D., González, M. and Vásquez, E. (2017). Microbial dynamics and turbidity in food fermentations: Implications for product quality. *Journal of Applied Microbiology*, 122(2), 476-485.
- Nassar, N. M. A. and Ortiz, R. (2010). Cassava improvement: Challenges and impacts. *Journal of Agricultural Science*, 148(2), 163–171.
- Pereira, T., Soares, M. and Costa, L. (2020). Genetic variation in dry matter content and starch yield in crops. *Plant Breeding*, 139(1), 72-79.

- Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Tufan, H. A., Tan, A. and Ceballos, H. (2017). Genome-wide association mapping of correlated traits in cassava: Dry matter and total carotenoid content. *Plant Genome*, 10(3), 1–14.
- Rojas, V., García, M. and López, D. (2020). Impact of fermentation on acid accumulation in food products: A genetic perspective. *Journal of Food Science*, 85(6), 1859-1868.
- Selle, P. H., Zhang, D. and Cowieson, A. J. (2020). pH stability in fermentation processes: A review. *Animal Feed Science and Technology*, 270, 114686.
- Tao, Y., Zhang, H., Li, X. and Yang, F. (2021). Genome-wide association study identifies markers for improved cassava traits in different environments. *Molecular Breeding*, 41(3), 55.
- Thompson, L., Wang, T. and Xu, P. (2021). The buffering capacity of food products during fermentation. *Journal of Food Chemistry*, 353, 129514.
- Van der Merwe, M., Botha, M., and Kruger, H. (2017). Genetic diversity and trait performance in crop improvement. *Euphytica*, 213(6), 1-12.
- Wang, L., Duan, Y., and Zhang, H. (2021). Enzyme activity during microbial fermentation and its role in processing. *Food Control*, 120, 107476.
- Wang, Z., Li, X., and Yang, J. (2018). Identification of loci controlling dry matter content in potatoes. *Theoretical and Applied Genetics*, 131(7), 1599-1609.
- Xu, J., He, W., Li, S. and Chen, L. (2019). Genome-wide analysis of time-dependent gene expression in crops: Implications for crop improvement strategies. *Plant Biotechnology Journal*, 17(4), 612-623.
- Yang, H., Liu, Q. and Wang, Y. (2018). Microbial metabolism and its effects on pH regulation during fermentation. *International Journal of Food Science and Technology*, 53(9), 2276-2283.
- Zhang, X., Li, W. and Wang, J. (2017). The genetic regulation of root texture in crops: Mechanisms and breeding applications. *Plant Science*, 256, 131-138.
- Zhang, Y., Huang, H. and Li, F. (2020). Pleiotropy in cassava: Genetic variations controlling moisture and dry matter content. *Frontiers in Plant Science*, 11, 572395.
- Zhang, Z., Wei, J. and Zhao, Q. (2016). Linkage disequilibrium decay and its implications for marker-assisted selection. *Genetics*, 203(3), 1243-1255.
- Zhang, Z., Wu, Y. and Li, L. (2019). Genetic loci involved in root texture and their relationship with starch properties in cassava. *Theoretical and Applied Genetics*, 132(10), 2783-2795.
- Zhao, H., Li, Z. and Liu, J. (2015). Patterns of linkage disequilibrium in diverse plant populations. *Theoretical and Applied Genetics*, 128(1), 13-25.
- Zhao, L., Zhang, Y. and Wang, X. (2020). Genetic basis of microbial activities in retting processes. *Microbial Ecology*, 79(4), 908-917.
- Zhao, Y., Li, J. and Wang, H. (2018). Genetic basis of texture variability in food processing. *Food Research International*, 110, 268-275.
- Zhou, J., Li, Y. and Wei, X. (2020). Genetic responses to environmental conditions during fermentation. *Journal of Agricultural and Food Chemistry*, 68(5), 1357.