



OPEN ACCESS

EDITED BY
Chao Shen,
Guangdong University of Petrochemical
Technology, China

REVIEWED BY
Simon Peter Abah,
National Root Crops Research Institute
(NRCRI), Nigeria
Weverton Gomes Costa,
Universidade Federal de Viçosa, Brazil

*CORRESPONDENCE
Kwame Obeng Dankwa
✉ anionlives@gmail.com

RECEIVED 10 February 2026
REVISED 31 March 2026
ACCEPTED 01 April 2026
PUBLISHED 22 April 2026

CITATION

Dankwa KO, Melendez JLL, Giraldo JC,
Sánchez-Sarria CE, Selvaraj M,
Olasanmi B and Gimode W (2026)
Genome-wide association study of post-
harvest physiological deterioration in
cassava (*Manihot esculenta* Crantz) using
visual and AI-powered phenotyping.
Front. Plant Sci. 17:1807180.
doi: 10.3389/fpls.2026.1807180

COPYRIGHT

© 2026 Dankwa, Melendez, Giraldo,
Sánchez-Sarria, Selvaraj, Olasanmi and
Gimode. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Genome-wide association study of post-harvest physiological deterioration in cassava (*Manihot esculenta* Crantz) using visual and AI-powered phenotyping

Kwame Obeng Dankwa ^{1,2*}, Jorge Luis Luna Melendez²,
Juan Camilo Giraldo², Camilo Enrique Sánchez-Sarria²,
Michael Selvaraj², Bunmi Olasanmi³ and Winnie Gimode²

¹Plant Breeding Program, Pan African University, Life and Earth Sciences Institute (Including Health and Agriculture), Ibadan, Oyo State, Nigeria, ²International Center for Tropical Agriculture (CIAT), Palmira, Colombia, ³Department of Crop and Horticultural Sciences, Faculty of Agriculture, University of Ibadan, Ibadan, Nigeria

Introduction: Postharvest physiological deterioration (PPD) is a rapid and severe process in cassava that causes root discoloration and spoilage soon after harvest, limiting shelf life and commercial value of the storage roots. Although environmental interventions can temporarily delay PPD, they are impractical for large-scale use. Previous studies have focused mainly on transcriptome, proteome, and candidate-gene analyses, with a few reporting robust and independently validated quantitative trait loci (QTLs). Identifying PPD-linked-QTLs is essential to breeding PPD-tolerant cassava varieties.

Methods: In this study, a genome-wide association study (GWAS) was conducted to identify genomic regions associated with PPD tolerance using both human visual scoring (VS) and an artificial intelligence (AI)-powered phenotyping method. A mapping population of 298 cassava accessions was genotyped using two platforms: DArTag mid-density panel and genotyping-by-sequencing (GBS).

Results and discussion: Across both genotyping platforms, 6 significant SNPs were identified using VS dataset and 6 using the AI phenotypic dataset. Consistent associations on chromosomes 1 and 12 across both phenotyping and genotyping platforms indicate potentially robust genomic regions influencing PPD response. Overall, the AI-powered phenotyping approach presents a standardized and reproducible PPD scoring procedure over the traditional visual scoring in future breeding programs for PPD. Despite differences in marker density, both DArTag and GBS markers capture comparable insights into the genetic structure of the GWAS panel. These findings provide insights into the genetic basis of PPD in cassava and offer valuable targets for marker-assisted and genomic selection toward developing cassava varieties with delayed PPD.

KEYWORDS

AI, cassava, genetics, GWAS, PPD, QTL, SNP markers

1 Introduction

Cassava (*Manihot esculenta* Crantz) is an important food security crop and a potential industrial crop in the world today. It is cultivated mainly in the tropics for its starch filled storage roots (Lokko et al., 2007). Cassava is a robust crop that can produce relatively reasonable root yield under marginal production conditions such as low soil fertility and irregular rainfall patterns, making it a risk buffer crop for smallholder farmers (Katz and Weaver, 2020; Ramcharan et al., 2017; Burns et al., 2010). It is the second most important staple crop in Sub-Saharan Africa with the highest per capita consumption of about 800g per person/day (Otun et al., 2023). Aside from gaining popularity as a staple food crop around the world, cassava now stands as an industrial and biofuel crop due to the high starch content of its storage roots (Dankwa and Pephrah, 2019; Ihemere et al., 2008; Munyikwa et al., 1997). Dried cassava roots can contain more than 80% starch of the total dry weight (Li et al., 2017; Baguma, 2004).

The entire cassava value chain heavily depends on the roots. However, post-harvest physiological deterioration (PPD) is a major constraint to cassava production affecting the root's marketability and processing quality. The phenomenon is characterized by discoloration of cassava storage roots from the natural white or cream flesh to blue-black or brown usually in response to physical bruise and thereby effectively reducing its quality (Beeching et al., 1998a; Noon and Booth, 1977; Wheatley and Gómez, 1985). This occurrence is swift, initiating a physiologically active response within 15 minutes of root injury (Houmani et al., 2018) and a visible discoloration appearing within 24–72 h of harvest (Iyer et al., 2010; Reilly et al., 2004; Beeching et al., 1998a, 1998b). This sometimes leads to further deterioration from microbial infection, causing roots to rot and soften, thereby rendering them useless (Luna et al., 2021).

Unfortunately, root wounding or mechanical damage is extremely hard to prevent and almost inevitable during cassava harvesting. This problem is exacerbated by the need to transport cassava roots to markets and processing factories which are usually far from the farms, especially in developing countries where road networks are poor and not easily plied (Reilly et al., 2004; Han et al., 2001). At the biochemical level, PPD is mainly caused by reactive oxygen species (ROS) production and accumulation in response to mechanical damage, a phenomenon called oxidative stress (Iyer et al., 2010; Reilly et al., 2004; Huang et al., 2001; Buschmann et al., 2000).

Advances in genomics through next-generation sequencing have led to the development of millions of new markers, which have been consistently utilized in studies of key agronomic traits (Edwards and Batley, 2010). Plant breeders increasingly use modern tools such as marker-assisted selection and genomic selection (GS) to enhance the speed and accuracy of developing crop varieties with desired traits (Crossa et al., 2017). In contrast, phenotype-based recurrent selection remains more challenging in cassava breeding due to its long maturity period (~10–12 months), low multiplication rate, and asynchronous flowering (Ceballos et al., 2012). Genome-wide association studies (GWAS) utilize linkage disequilibrium (LD) and high-density marker sets to dissect the genetic

architecture of quantitative traits (Nordborg and Weigel, 2008). It continues to play an important role in plant breeding and has been applied to many phenotypic traits in cassava including carotenoid and dry matter content (Rabbi et al., 2017), disease resistance (Wolfe et al., 2016; Kulakow et al., 2013), starch content, harvest index, leaf shape, and many more morphological and quality traits (Rabbi et al., 2022; Zhang et al., 2018). Nevertheless, there is a limited report on quantitative trait loci for PPD in cassava. Genetic intervention to the PPD problem in cassava is not only crucial but also the best way toward a sustainable cassava production and value chain. Moreover, the labor-intensive and subjective nature of PPD visual scoring presents a challenge, as different evaluators may assign different scores to the same deterioration level under similar conditions. To address this, the CIAT phenomics team developed AI-based algorithms for efficient analysis of large datasets, enabling rapid cassava root screening for PPD symptoms (Ayalde et al., 2024). This approach uses image and color identification techniques and offers a fast, reliable, and standardized method for PPD phenotyping.

Salcedo and Siritunga (2011) suggested that, for an effective solution to PPD in cassava, both basic and applied research should focus on long-term strategies such as developing improved varieties through conventional breeding and selection. However, it will be important to first develop standardized methods for evaluating PPD that are universal in both practice and quantification. A uniform evaluation tool would benefit scientists globally by enabling more accurate correlations between phenotypic PPD tolerance and findings from molecular and cellular studies.

In this study, PPD scoring was conducted using human visual assessment and AI-powered image analysis on 298 cassava genotypes in a clonal evaluation trial over two years at CIAT, Colombia. These clones were also genotyped using two distinct genotyping platforms: DArTag mid density and genotyping by sequencing (GBS). Separate GWAS analyses were performed to identify loci associated with PPD based on these two phenotypic datasets. The findings provide insights into genomic regions linked to PPD in cassava and highlight the potential of AI image analysis for enhanced and uniform PPD phenotyping.

2 Materials and methods

2.1 Experimental design and field establishment of mapping population

A field evaluation of a multi-parental mapping population comprising 298 genotypes was established in Palmira, Colombia. A partial replication design was used in the field layout. Each genotype was planted in a single-row plot containing six plants, with a spacing of 0.8 meters between plants and 1.2 meters between plots. The field evaluation was carried out in two seasons (2023/2024 and 2024/2025). Harvesting was done 11 months after planting. Although specific plot locations shifted between seasons to accommodate crop rotation, both trials were conducted at the Palmira station. During the 2023/2024 season, the average

temperature, relative humidity, and precipitation were 32.75 °C, 80.32%, and 2.43 mm, respectively, compared with 27.07 °C, 80.88%, and 0.07 mm in the 2024/2025 season.

2.2 PPD phenotyping

Two approaches were employed to phenotype PPD in this experiment; human visual scoring (VS) and AI-powered image analysis. The VS was carried out according to the method by Wheatley (1982), with some modifications (Dankwa et al., 2025; Sánchez et al., 2013). For each genotype, three plants were harvested and ten commercial-size roots from the three plants were used for PPD phenotyping. To induce deterioration, the distal and proximal extremes of each root were cut to obtain a 15-cm piece. The distal part of the root was covered with a plastic film and secured with a rubber band. The roots were then stored in a controlled chamber and PPD was evaluated on day 7 of storing the prepared roots, starting from the day of harvest as day zero (day 0). The temperature and relative humidity of the chamber were monitored using the Multi-use PDF Temp & RH –TempU03 Data Logger (Luna et al., 2021) and they were kept approximately at 25 °C and 70%, respectively. For the PPD scoring, the stored roots were transversely cut into seven slices (2cm thick), starting at the proximal end. Images for the slices were taken against a dark background (Figure 1) using FOTODIOX LED 770 STUDIO BOX and accessories. Each slice was assigned a PPD score between 0 and 10, corresponding to the percentage of the cut surface that exhibits discoloration of the parenchyma (0 = 0%, 1 = 10%, 2 = 20%, etc.) (Luna et al., 2021; Sánchez et al., 2013). The PPD score for each root was calculated using the average of the seven slices. Decomposed or microbial-damaged roots were discarded. The images of the root slices were subjected to a deep-learning high throughput technique using an AI-powered PPD scoring model with the PPDAnalyzer v2.0 (Ayalde et al., 2024).

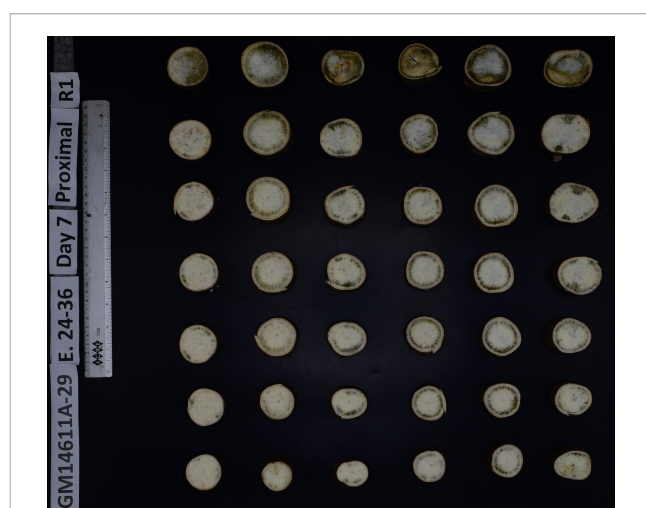


FIGURE 1
Representative image of cassava root slices for postharvest physiological deterioration (PPD) assessment. Slices are displayed against a high-contrast dark background. These standardized images are used for both human visual scoring and as input data for the AI-powered scoring model to evaluate PPD.

The AI-powered PPD scoring model (Ayalde et al., 2024) was developed using YOLO (You Only Look Once) foundation models to classify the severity of PPD from digital images. Among the tested architectures, YOLOv8 achieved the highest performance, with a mean average precision (mAP) of 84.0%, demonstrating a good balance between precision (0.77) and recall (0.831). To further improve the accuracy of damage estimation, the workflow incorporated the Segment Anything Model (SAM) for precise background removal and K-means clustering for automated quantification. The model was trained using a dataset of 1,142 images containing 23,934 expert-validated annotations of individual cassava root slices. The dataset was randomly divided into training (80%; 19,152 annotations), validation (15%; 3,584 annotations), and testing (5%; 1,198 annotations) sets to minimize bias and ensure robust model evaluation. Model performance was assessed using confusion matrices, which provided insight into class-specific predictions and misclassifications, and mean absolute error (MAE), which measured the absolute difference between the AI-predicted PPD percentage and the expert-assigned ground truth (Ayalde et al., 2024).

2.2.1 Phenotypic data analysis

For each PPD dataset (VS and AI), two years data were collapsed into a single best linear unbiased prediction (BLUPs) using a linear mixed model $Y_{ijk} = \mu + \beta_j + g_i + (g\beta)_{ij} + rk(j) + \epsilon_{ijk}$ within the lme4 package in RStudio (Boer and Rossum, 2022; R Core Team, 2021). Where: Y_{ijk} = observed PPD value for the i -th genotype, in the j -th year, and k -th rep. μ = overall intercept (grand mean). β_j = fixed effect of the j -th year (Year is fixed). g_i = random effect of the i -th genotype. $(g\beta)_{ij}$ = genotype $\times$ environment (year) interaction. $rk(j)$ = random effect of the k -th replication nested within the j -th year. ϵ_{ijk} = residual error.

The BLUPs were used to calculate the PPD genetic correlation using three models (MLMM, BLINK, and FarmCPU) in GAPIT (Wang et al., 2022). The same lme4 package was used to estimate PPD broad sense heritability as: $H^2 = \frac{\sigma_G^2}{\sigma_G^2 + (\sigma_{GE}^2/e) + (\sigma_E^2/re)}$. Where: H^2 is the broad-sense heritability, σ_G^2 is genotypic variance, σ_{GE}^2 is genotype $\times$ environment (year) interaction variance, σ_E^2 is Residual (error) variance, e is number of environments (years), r is number of replications per environment (year).

To calculate the coefficient of variation (CV) for each year, a mixed linear model $Y_{ik} = \mu + G_i + R_k + \epsilon_{ik}$ was fitted, where Y_{ik} is the observed PPD value of the i -th genotype in the k -th replication, μ is the overall mean, G_i is the random genotypic effect, R_k is the random replication effect, and ϵ_{ik} is the residual error. The CV for each year was calculated as the ratio of the residual standard deviation to the year-specific trait mean.

Pearson's correlation coefficient (r) was used to assess the relationship between human visual and AI-based PPD scores. Statistical significance was tested at $P < 0.05$. The result was visualized using ggplot in R.

2.3 DNA extraction and genotyping

Genomic DNA was extracted from cassava leaf samples using the CTAB method, following the DNA extraction protocol by Doyle (1991) with some modifications. DNA quantification was

performed via absorbance reading using a Nanodrop spectrophotometer. All DNA samples were assessed on a 1% agarose gel to verify quality before sequencing. The mapping population was genotyped using two distinct genotyping platforms; DArTag mid density panel and genotyping by sequencing (GBS) to identify single nucleotide polymorphism (SNP) markers associated with PPD in cassava. The DArTag mid density genotyping was done at Intertek, Australia while the genotyping by sequencing (GBS) was done at University of Minnesota Genomics Center.

2.4 SNP calling and filtering

For DArTag mid density, which is based on the *Manihot esculenta* version 7.1 genome, 3,002 single nucleotide polymorphisms (SNP) markers were obtained. These SNPs were filtered to exclude those with missing data $\geq 80\%$ and a minor allele frequency (MAF) < 0.05 , using the snpReady package in RStudio (Granato et al., 2018; R Core Team, 2021). For genotyping-by-sequencing (GBS), the raw paired-end sequence data trimming and resynchronization were performed on each sample using the gbsTrim and resync Perl scripts. All sample sequences were then aligned and indexed to version 8.1 of the *Manihot esculenta* reference genome using bwa mem (version 0.7.12), and variant discovery was carried out following the GATK Best Practices workflow (version 4.1.8) (DePristo et al., 2011), retaining only biallelic SNPs. The resulting dataset was filtered with vcftools (Danecek et al., 2011) using thresholds of 10% missing data, a minor allele frequency (MAF) of at least 5%, and a quality score (Q) greater than 30. After these filtering processes, 2,526 SNPs for the DArTag dataset and 30,046 SNPs for the GBS dataset were utilized to conduct a GWAS.

2.5 Linkage disequilibrium, population structure, and genetic relatedness analysis

To assess genome-wide marker distribution, SNP density was profiled for both DArTag and GBS datasets using a 1-Mb sliding window with the SRplot platform (Tang et al., 2023). For each genetic dataset (DArTag and GBS), linkage disequilibrium decay (LD decay) was calculated using PLINK version 1.9 (Purcell et al., 2007) and visualized with the ggplot2 R package (R Core Team, 2021). To account for genetic relationships between individuals, kinship was calculated with the ZHANG (Zones Harbored Adjustments of Negligent Genetic) algorithm in GAPIT (Zhang, 2014; Wang et al., 2022). To reveal potential subgroups within the population, principal component analysis (PCA) was performed using the genotypic marker datasets in TASSEL software (version 5.2.96) and visualized using the ggplot2 R package. Genotypes were categorized by common maternal parental background to provide biological context to the observed genetic clustering.

2.6 Association analyses and candidate gene exploration

The phenotypic (BLUP values) and genotypic data were analyzed to determine associations using the Genomic Association and Prediction Integrated Tool (GAPIT) package version 3.4.0 in R employing three models (MLMM, BLINK, and FarmCPU) (Wang and Zhang, 2021). Significant marker-trait associations were

identified using a Bonferroni-corrected significance threshold ($P < 0.05/m$), where m represents the total number of SNPs tested. The identified significant loci were explored for potential candidate genes using the *Manihot esculenta* v7.1 and v8.1 for DArTag mid-density and GBS, respectively. The corresponding v8.1 positions for DArTag v7.1 positions were explored as well.

The effectiveness of the models in controlling for systematic bias was evaluated by calculating the genomic inflation factor (λ). Defined as the ratio of the median of the observed χ^2 statistics to the expected median χ^2 under the null hypothesis. The P-values generated by GAPIT for each model were imported into RStudio to calculate the chi-square statistics and the resulting λ values.

3 Results

3.1 Phenotypic distribution

A comparison of the phenotypic data distributions for PPD scores obtained from human visual assessment (Figure 2A) and AI-powered image analysis (Figure 2B) revealed similarities and important methodological differences. The distribution of visually scored PPD was right skewed, with most genotypes clustering toward the lower end of the severity scale. The majority of observations fell between 0 – 20%. The mean visual PPD score was 15.47% and the distribution exhibited a long tail extending toward higher values ($>50\%$).

In contrast, the AI-powered phenotyping method produced a more symmetrical, near-normal distribution centered around moderate PPD levels. The AI-based PPD values ranged approximately from 20 to 75%, with most genotypes clustering near the mean of 43.84%. Unlike the visual score distribution, the AI-based scores occupied a broader range. The mean line fell near the center of the distribution, confirming a more even spread of phenotypes across the panel.

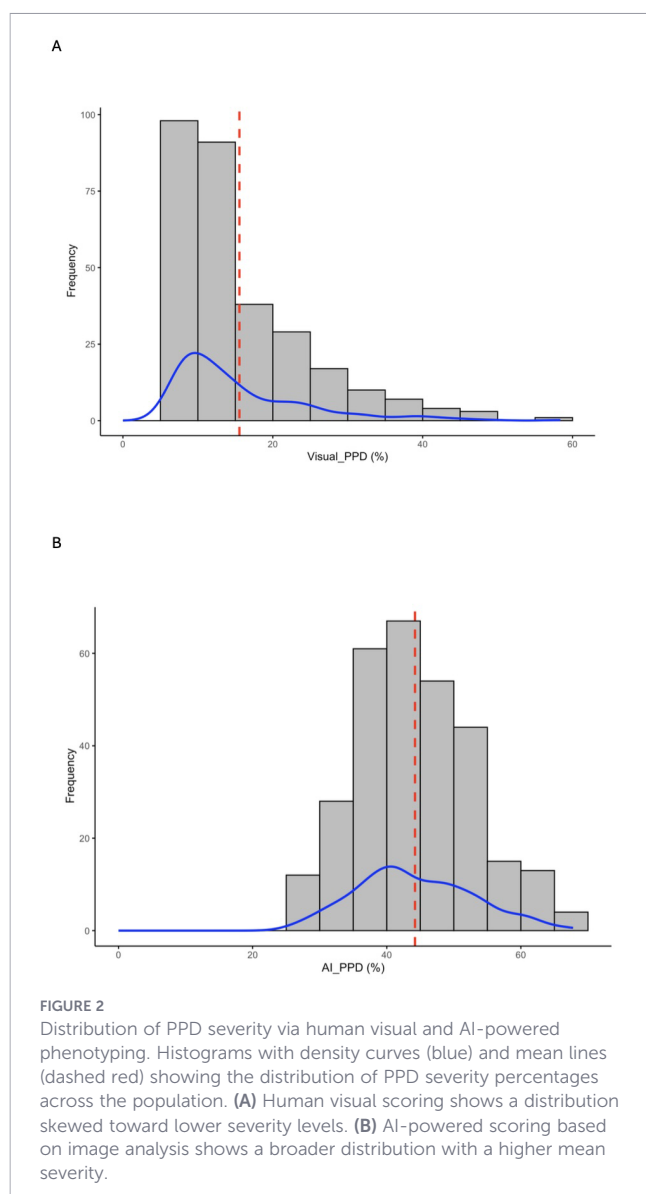
Across both approaches, the visual scoring method assigned lower percentage values for PPD severity relative to the AI-based system, particularly for genotypes exhibiting mild-to-moderate deterioration.

For visual PPD scores, the calculated broad-sense heritability was high ($H^2 = 0.71$), despite high coefficients of variation (55.5% in 2023/24 and 73.2% in 2024/25). In contrast, AI-based scores showed lower H^2 (0.64) but significantly more stable CV values (35.2% and 45.7%). Visual and AI-based PPD BLUPs were positively correlated ($r = 0.55$; Supplementary Figure 1). The elevated CV values likely reflect the inherent variability of PPD, which often manifests as significant variation between replicates of the same genotype or even among individual roots harvested from the same plant.

3.2 Genomic structure and landscape in the studied population

3.2.1 SNP density across the cassava genome

To assess genome-wide marker distribution, SNP density was profiled for both DArTag and GBS datasets using a 1-Mb sliding window (Figures 3A, 4A). The DArTag markers (Figure 3A)



exhibited a moderate and localized genomic distribution, with SNP density peaking at 16 SNPs/Mb. While higher densities were noted on some chromosomes such as chromosomes 12 and 15, large segments of the genome remained sparsely covered. This distribution reflects the targeted nature of the DArTag panel, where marker density is a function of probe design and the prioritization of specific loci. In contrast, GBS markers (Figure 4A) provided a significantly higher marker volume, with a nearly 8-fold increase in peak density (up to 126 SNPs/Mb). However, despite the higher density, GBS maintained a non-uniform distribution, with prominent marker-dense regions on chromosomes 6, 8, 11, and 14, where windows exceeded 98 SNPs/Mb, with some regions showing reduced marker density. The GBS distribution reflects the inherent bias of restriction-site-associated DNA sequencing, where marker density depends on the frequency and accessibility of restriction enzyme recognition sites. Overall, GBS provided greater genomic resolution which was approximately 6 to 8-fold higher SNP counts per window compared to DArTag – while DArTag provided targeted coverage of high-priority loci at a lower density.

3.2.2 Linkage disequilibrium decay patterns

Linkage disequilibrium (LD) decay analysis revealed highly consistent patterns across both the DArTag and GBS marker platforms (Figures 3A, 4B). The smoothed r^2 curve intersected the background LD threshold ($r^2 \approx 0.10$) at approximately 685,090 bp for DArTag and 625,262 bp for GBS, suggesting that both platforms are effectively capturing the same underlying genomic architecture and recombination landscape of the population. The scatter patterns and fitted curves indicate that both datasets capture strong LD at short distances, followed by a gradual and comparable decline over larger genomic distances. While GBS exhibited a marginally steeper initial decline, offering a slightly refined resolution for identifying narrow QTLs, the overall widespread recombination indicated in both panels suggests that both datasets are well-suited for robust association mapping in this cassava germplasm.

3.2.3 Principal component analysis

Principal component analysis (PCA) was used to assess genetic structure in the cassava panel using DArTag (Figure 3C) and GBS markers (Figure 4C). The analysis revealed a high level of congruency in clustering patterns between the DArTag and GBS marker platforms. In both analyses, the first two principal components accounted for a comparable proportion of the total genetic variance (21.3–22.2%), indicating similar large-scale population structure across marker systems. Both platforms resolved the same primary genetic groupings with accessions generally clustered according to shared parental background. Notably, groups such as ARG73, PER183, CM9460-40, HMC1, SM3559-11, and SM2773-32 formed well-defined clusters across both marker systems, though some overlap was also observed among certain families. While the absolute orientation of the clusters shifted along the axes—a common artifact of PCA coordinate calculation—the relative genetic distances between the families remained stable.

In the DArTag analysis, PC1 and PC2 explained 15.2% and 7.0% of the variance, respectively, while the GBS-based PCA explained 14.6% of the variance in PC1 and 6.7% in PC2, closely mirroring the DArTag results. Despite slight differences in resolution, the overall clustering remained comparable between platforms, confirming that both DArTag and GBS are robust for population structure characterization.

Notably, the SM2792-31 maternal group appeared as two distinct subgroups. This separation is attributed to the use of different paternal parents (PER183 and PER221) in crosses with the SM2792-31 female parent. The PCA effectively captured this paternal influence, splitting the maternal family into two discrete clusters positioned toward their respective paternal donors. This further confirms that both DArTag and GBS were robust and sensitive enough to resolve pedigree-level substructure.

3.3 Marker loci linked to PPD

The GWAS conducted using both DArTag mid-density and GBS markers identified loci associated with post-harvest physiological deterioration (PPD) under human visual and AI-powered

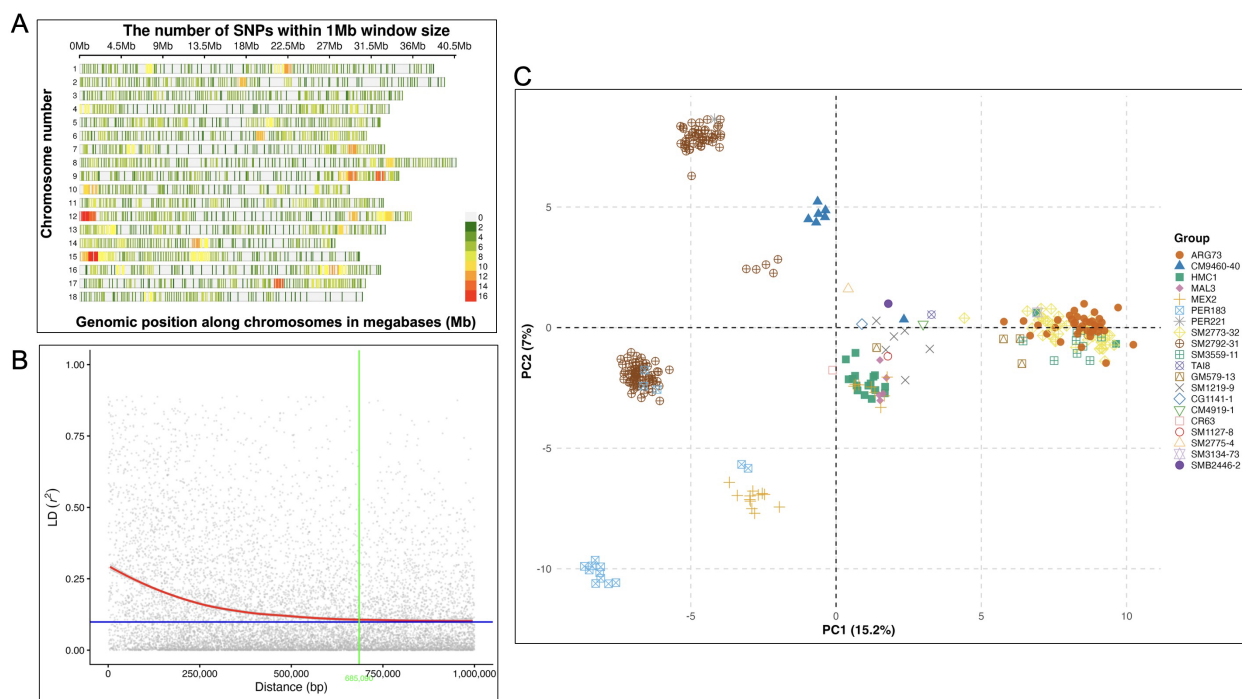


FIGURE 3

Population structure of 298 cassava genotypes used in genome-wide association studies (GWAS) based on DArTag markers. **(A)** Marker distribution and density across the 18 chromosomes, **(B)** linkage disequilibrium decay, and **(C)** clustering results visualized on a principal component analysis (PCA) plot.

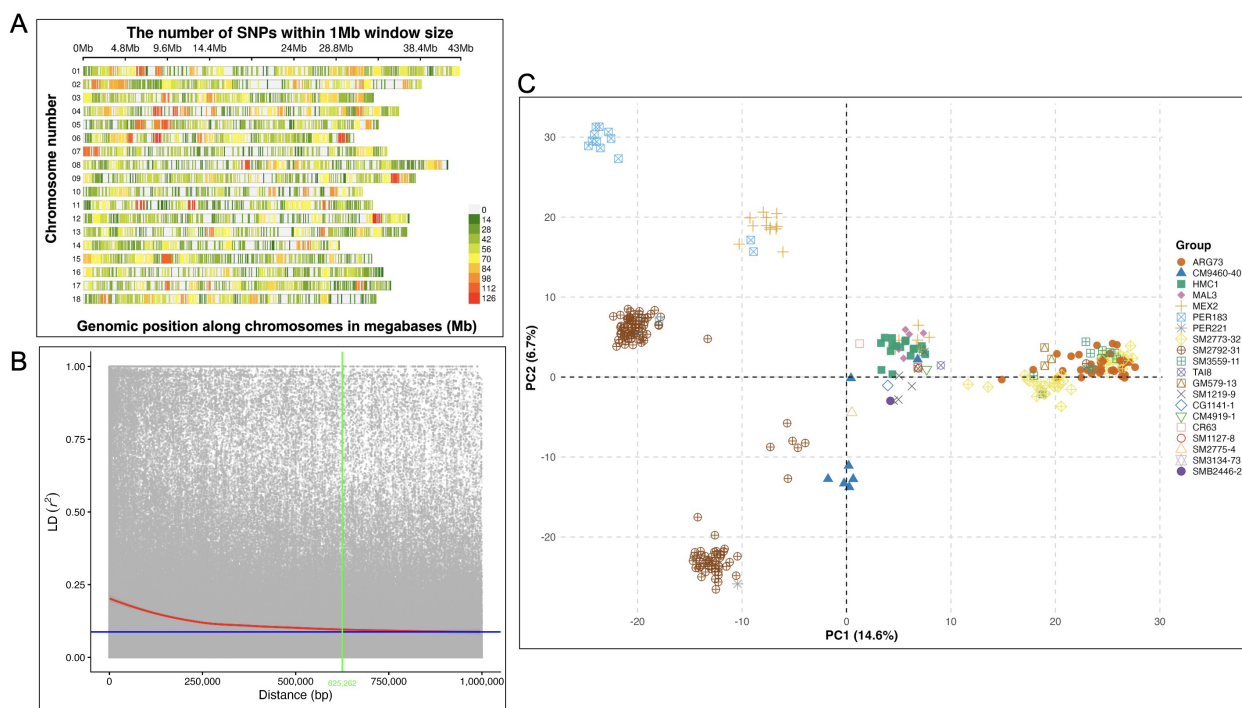


FIGURE 4

Population structure of 298 cassava genotypes used in genome-wide association studies (GWAS) based on GBS markers. **(A)** marker distribution and density across the 18 chromosomes, **(B)** linkage disequilibrium decay, and **(C)** clustering results visualized on a principal component analysis (PCA) plot.

phenotyping approaches. The DArTag-based GWAS revealed several significant SNPs distributed across multiple chromosomes for both scoring methods (Table 1; Figure 5). For the human visual dataset, significant associations were detected on chromosomes 12 and 15 with the SNPs CassV7_chr12_32503122 and CassV7_chr15_3648525 explaining 5.18% and 0.75% of the phenotypic variance, respectively. The AI-powered scoring approach identified loci on chromosomes 1 and 8, with CassV7_chr08_901823 explaining 0.96% of the phenotypic variance, and the other associated locus on chromosome 1 exhibiting negligible (0%) individual PVE values (Table 1). The Manhattan plots (Figures 5A, B) displayed clear peaks corresponding to these associations, while the Q-Q plots indicated proper control of population structure. However, the BLINK model exhibited mild genomic inflation ($\lambda = 1.22$) for the AI-powered assessment, though the SNP (CassV7_chr01_28520241) was corroborated by the more conservative FarmCPU model ($\lambda = 0.92$).

Similarly, GWAS using GBS markers identified several significant SNPs associated with PPD across the two phenotyping methods (Table 2; Figure 6). For the human visual assessment, major association peaks were observed on chromosomes 2, 12 and 15 with S02_9881815 consistently detected across MLM and BLINK models. This marker explained a high phenotypic variance of 35.5% and 44.8% for MLM and BLINK models respectively. The AI-powered dataset revealed additional significant loci on chromosomes 1, 12, and 14 with S14_17899979 explaining phenotypic variance of 57.7%. The Manhattan and Q-Q plots (Figures 6A, B) confirmed these associations and demonstrated good model fit with limited false positives. Notably, the BLINK model for AI-powered traits showed mild genomic inflation ($\lambda = 1.18$) while the visual scoring model exhibited a more conservative, slightly deflated value ($\lambda = 0.87$).

4 Discussion

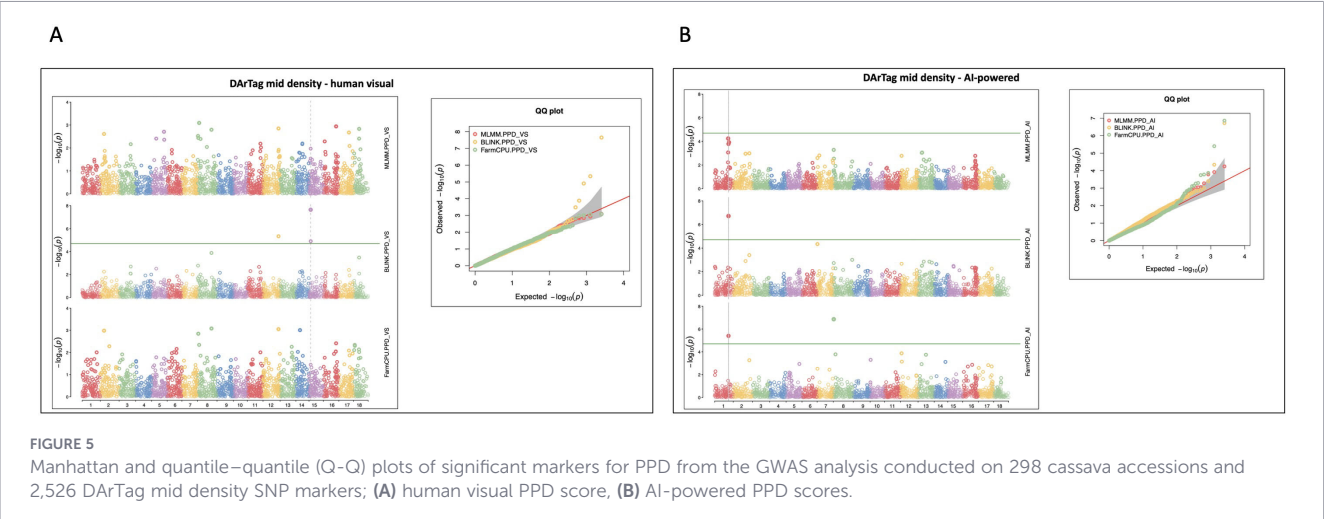
Artificial intelligence (AI) has had an exciting application in modern plant breeding, both demonstration and potential; data mining, phenotyping, monitoring, multi-omics, environment, sustainability, and economics (Farooq et al., 2024; Sangjan et al., 2025). In winter wheat breeding, artificial intelligence (AI) serves as a link

between scientific innovation and practical application by addressing limitations of traditional statistical approaches and enabling the generation of high-quality phenotypic data (Chang-Brahim et al., 2024). Successful genome-wide association studies (GWAS) rely on accurate and reliable phenotypes, which AI facilitates through high-throughput remote phenotyping platforms (Huang et al., 2025). Beyond phenotyping, AI and machine learning have also emerged as essential tools for integrating heterogeneous data sources, including single-cell omics and protein–interaction networks, to identify disease drivers (Hazelett, 2025). The strategy of utilizing AI-derived phenotypes for GWAS is established in recent literature across diverse crops. For example, deep learning models were successfully employed in lettuce to map 11 QTLs linked to seven distinct seed traits (Seki and Toda, 2022). This study demonstrated the effectiveness of image-based phenotyping in accelerating the discovery of the genetic basis for small, complex morphological features—such as seed size and shape—that are difficult to quantify manually. Similarly, research in soybean has utilized deep learning to capture complex phenotypic traits directly from field imagery. Rairdin et al. (2022) compared image-based phenotypic data with manually collected visual scores for GWAS and obtained highly comparable results which were validated by detecting genomic regions near previously reported loci associated with sudden death syndrome in soybean.

The method used to evaluate a trait is critical, as it can influence subsequent decisions made on that trait. Variability in phenotyping methods, particularly differences between specific-gravity and oven-drying approaches for estimating dry matter content (DMC) has been shown to reduce the predictive accuracy of DMC in genomic selection studies (Esuma et al., 2021). The two phenotyping methodologies employed in this study clearly demonstrated the effect of methodology. The human visual PPD scores seem to indicate that visible deterioration symptoms were generally mild during the assessment period and a right-skewed distribution suggests the presence of a smaller subset of highly susceptible genotypes (Figure 2A). With the AI-based PPD scores, a broader spread and nearly even distribution with a higher mean indicates higher incidence of PPD. This pattern implies that the AI system may be more sensitive to subtle deterioration features and therefore produces greater phenotypic differentiation among genotypes.

TABLE 1 Significant SNPs identified from human visual and AI-powered PPD scoring methods using DArTag mid density markers with genomic inflation factor (λ) for each model.

PPD scoring method	Model	SNP_ID	Chromosome	Position (v7.1)	Position (v8.1)	P. value	Minor allele frequency	Phenotypic variance explained (%)
Visual	BLINK ($\lambda = 0.97$)	CassV7_chr12_32503122	12	32503122	33211138	4.60E-06	0.32	5.18
		CassV7_chr15_3536121	15	3536121	3393477	2.24E-08	0.47	0
		CassV7_chr15_3648525	15	3648525	3505542	1.24E-05	0.36	0.75
AI-powered	BLINK ($\lambda = 1.22$)	CassV7_chr01_28520241	1	28520241	32684118	1.92E-07	0.18	0
	FarmCPU ($\lambda = 0.92$)	CassV7_chr01_28520241	1	28520241	32684118	3.97E-06	0.18	0
		CassV7_chr08_901823	8	901823	647907	1.37E-07	0.35	0.96



The visual PPD scoring is labor-intensive and more prone to subjectivity or human errors that may further limit the power to detect genomic regions underlying the trait—a challenge that AI-based image analysis is designed to overcome. These contrasting distribution patterns indicate that although both methods capture meaningful biological variation, the AI-based scoring may provide enhanced phenotypic resolution, which is likely beneficial for downstream GWAS, genomic selection, or association mapping analyses. The complexity of human visual scoring likely contributes to the fact that limited QTL for PPD in cassava have been reported (Cortés et al., 2002), with subsequent research relying largely on transcriptomic, proteomic, and candidate-gene evidence rather than robust, replicated QTLs.

The positive association observed between visual PPD BLUPs and AI-derived PPD BLUPs (Pearson’s $r = 0.55$) (Supplementary Figure 1) indicates a general agreement between the two approaches, with higher visual PPD scores corresponding to higher AI-predicted PPD values. Despite this positive relationship, some dispersion around the fitted regression line was evident, suggesting that the two methods are not perfectly concordant. The marked experimental variability observed between seasons underscores the significant environmental dependence of PPD expression. Consequently, multilocation and multiseasonal evaluation are

essential to ensure the robustness and stability of PPD phenotyping data.

PPD is one of the most important constraints in cassava production (Naziri et al., 2014; Sayre et al., 2011). It is uniquely rapid in cassava, commencing a few hours after harvesting, especially when the storage roots are injured during harvesting and post-harvest handling. The early phase of PPD, usually referred to as the primary stage, is basically a complex physiological and biochemical response that is triggered by root injury and begins at the wound site a few hours after injury (Bayoumi et al., 2008; Reilly et al., 2004). Root wounding or mechanical damage is extremely hard to prevent and almost inevitable during cassava harvesting and transportation. Hence, identifying QTLs linked to delay in PPD and deploying them through marker-assisted selection (MAS) could strengthen the cassava-value-chain through rapid breeding pipelines to reduce postharvest losses.

Markers from the two genotyping platforms used in this study yielded congruent and comparable results. Differences in SNP density reflect the inherent contrasts between DArTag and GBS sequencing approaches (Figures 3A, 4A) and have implications for downstream analyses such as LD decay, haplotype detection, and GWAS resolution. However, despite these differences in marker density and distribution, both platforms revealed similar LD decay

TABLE 2 Significant SNPs identified from human visual and AI-powered PPD scoring methods using GBS markers with genomic inflation factor (λ) for each model.

PPD scoring method	Model	SNP_ID	Chromosome	Position (v8.1)	P. value	Minor allele frequency	Phenotypic variance explained (%)
Visual	MLMM ($\lambda = 1.01$)	S02_9881815	2	9881815	1.59E-08	0.09	35.95
		S15_9885471	15	9885471	1.92E-07	0.11	17.41
	BLINK ($\lambda = 0.87$)	S02_9881815	2	9881815	2.38E-07	0.09	44.80
		S12_20118969	12	20118969	1.86E-07	0.08	9.45
AI-powered	BLINK ($\lambda = 1.18$)	S01_31747016	1	31747016	4.02E-10	0.06	21.30
		S12_8904674	12	8904674	9.21E-09	0.48	6.75
		S14_17899979	14	17899979	9.26E-07	0.108	57.70

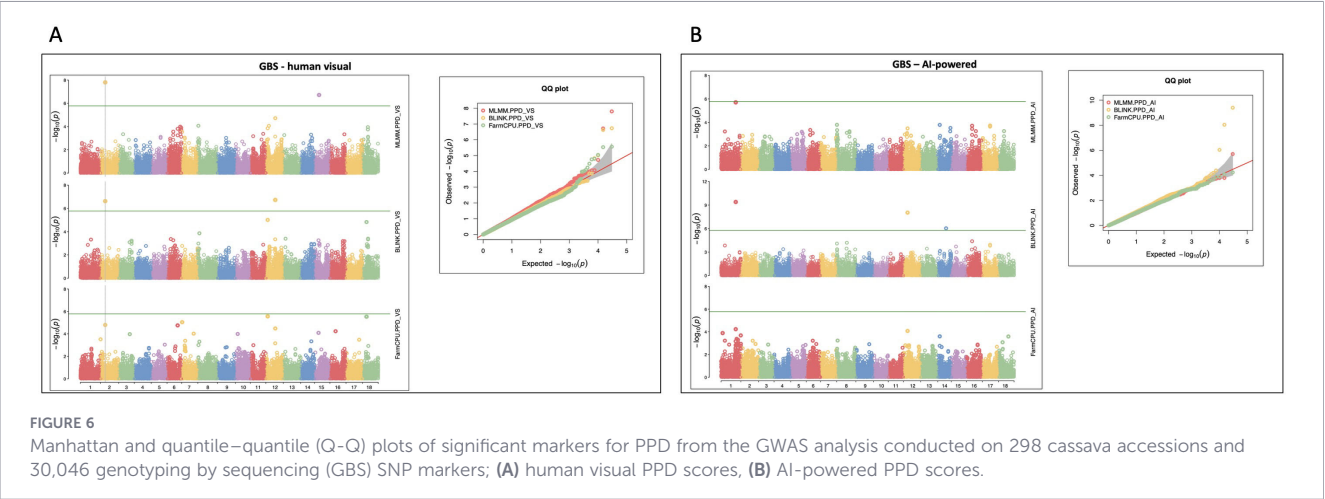


TABLE 3 Putative gene annotations of significant SNPs directly mapped to annotated genes using DArTag and GBS markers, aligned to cassava reference genome versions 7.1 and 8.1, respectively.

PPD scoring type	SNP platform	SNP ID	Chromosome	Position	Gene ID	Gene annotation
Visual	DArTag	CassV7_chr12_32503122	12	32503122 (v7.1)	Manes.12G126100	KINASE INTERACTING (KIP1-LIKE) FAMILY PROTEIN
		CassV7_chr15_3536121	15	3536121 (v7.1)	Manes.15G043900	F19P19.6 PROTEIN
		CassV7_chr15_3648525	15	3648525 (v7.1)	Manes.15G046000	OS02G0131700 PROTEIN
AI-powered		CassV7_chr01_28520241	1	28520241 (v7.1)	Manes.01G133800	DUF1195 family protein
AI-powered	GBS	S01_31747016	1	31747016 (v8.1)	Manes.01G121600	Nonsense-mediated mRNA decay protein 3 (NMD3)

patterns. Linkage disequilibrium in both datasets declined with physical distance, reaching a stable r^2 plateau (0.10) at approximately 625–685 kb. This indicates similar haplotype block structure and recombination rates across both marker platforms (Figures 3B, 4B). This harmony suggests that both DArTag and GBS provide consistent estimates of genome-wide LD structure, supporting their joint use and comparability in downstream analyses, including GWAS and population structure inference.

The DArTag and GBS marker analyses identified unique genomic regions associated with PPD tolerance. The identified loci using AI-based PPD scores highlights the possibility of using AI-assisted phenotyping approaches in capturing PPD-related genetic variation in cassava. While the GBS significant SNPs with high PVE (S01_31747016, S02_9881815, S14_17899979: Table 2) met our minor allele threshold ($MAF \geq 0.05$), they exhibit limited allele representation in our population. Specifically, the alternate or favorable alleles were primarily present in the heterozygous state and were rarely found in the homozygous state. This likely contributed to inflated PVE estimates and the observed genomic inflation observed especially in the BLINK model ($\lambda = 1.18 - 1.22$). While these signals clearly identify PPD-associated regions, further validation in larger, independent populations is required to confirm the magnitude of these genetic effects.

As a biochemical response, PPD is linked to a build-up of reactive oxygen species (ROS) resulting in oxidative stress (Reilly et al., 2004) and enzymes such as catalase, peroxidase, and superoxide dismutase that regulates the ROS levels (Buschmann et al., 2000; Iyer et al., 2010; Reilly et al., 2001). In plants, ROS are produced as a by-product of aerobic respiration and play crucial roles in essential processes such as photosynthesis and other cellular metabolism (Singh et al., 2016; Apel and Hirt, 2004). Under typical conditions plants possess several mechanisms to neutralize ROS, preventing or mitigating their toxicity (Choudhary et al., 2017). However, when under stress (biotic or abiotic), the rate of ROS production far supersedes the rate at which they are dissipated (also called detoxification), causing a disequilibrium and a rapid buildup of ROS, a phenomenon known as oxidative burst (Apostol et al., 1989). This disequilibrium between ROS production rate and detoxification can result in DNA, proteins, and lipids damage (Tripathy and Oelmüller, 2012). In well studied grain crops like wheat and maize, biotic and abiotic factors that could result in large production of ROS include but not limited to drought, extreme temperatures, salinity, ultraviolet-B radiation, heavy metals, and pathogens (Tripathi et al., 2017; Singh et al., 2016; Arora et al., 2002). In cassava, cultivars with high levels of substances that reduce ROS accumulation such as β -carotene (Sanchez et al.,

2005) show delayed PPD while overexpression of a cyanide-resistant terminal oxidase in plants (alternative oxidase) that reduces cyanide accumulation has similar effect (Zidenga et al., 2012).

While the candidate genes identified are biologically plausible based on their co-mapping with significant SNPs (Table 3), they remain putative until functional validation can be performed. Generally, these annotations (Table 3) point to a model in which wound-triggered signaling (Ca^{2+} fluxes and ion channels), vacuolar trafficking and proteostasis, and RNA surveillance contribute to the early cellular responses that may influence PPD occurrence and progress. Cassava PPD is characterized by rapid production and accumulation of reactive oxygen species (ROS), calcium (Ca^{2+}) signaling, and programmed cell death (PCD). Ca^{2+} -related genes are noteworthy candidates, as cyclic-nucleotide-gated channels potentially mediate Ca^{2+} influx that can regulate ROS generation and PCD soon after harvest, linking external stress cues to internal deterioration processes (Wang et al., 2021; Djabou et al., 2017).

Based on Arabidopsis functional annotation, the F19P19 protein is putatively associated with plant defense responses through transcriptional regulation of defense-related genes and is hypothesized to play a role in innate immune signaling pathways (Reiser et al., 2024). The OS02G0131700 protein belongs to the DUF794 domain family, which is predominantly found in plants and remains functionally uncharacterized (Finn et al., 2016). Some annotation databases further associate this protein with the RNA recognition motif (RRM) family, particularly the RRM1 subtype. RRM-containing proteins in rice have been widely implicated in post-transcriptional regulation and responses to environmental stresses, including pathogen infection, suggesting a potential role for this gene in stress-related regulatory processes (Zhang et al., 2016; Lorković, 2009).

The link with an NMD3-like gene points toward RNA quality control as a potential factor influencing PPD. The NMD genes help regulate faulty or stress-related RNA molecules, and changes in their activity can alter how plants respond to stress (Nyikó et al., 2013). Hence changes in NMD activities can reprogram defense and stress pathways and could therefore influence the magnitude or timing of PPD responses.

Ultimately, the SNPs and candidate genes associated with these diverse biological pathways underscore the complex polygenic nature of cassava PPD. While these specific mechanisms require further functional validation, their identification through congruent GWAS signals may provide a robust genetic framework for future breeding efforts.

5 Conclusion

This study provides valuable insights into the genetic basis of post-harvest physiological deterioration (PPD) in cassava using two different phenotypic and genotypic approaches. The AI-based approach may serve as a useful, objective tool for standardizing PPD evaluation in high-throughput breeding contexts. The loci and candidate genes identified provide a plausible framework for understanding PPD tolerance. Collectively, these findings contribute to the genomic resources available for developing cassava varieties with improved post-harvest stability.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

KD: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. JM: Data curation, Investigation, Methodology, Writing – review & editing. JG: Formal analysis, Investigation, Writing – review & editing. CS-S: Data curation, Formal analysis, Writing – review & editing. MS: Formal analysis, Writing – review & editing. BO: Conceptualization, Supervision, Writing – review & editing. WG: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research work was fully funded by the Gates Foundation at CIAT. The PhD studies of the lead author in Plant breeding at the Pan African University, Life and Earth Sciences Institute (Including Health and Agriculture), Ibadan, Nigeria was sponsored by the African Union Commission (AUC) scholarship.

Acknowledgments

The authors are grateful to the staff of CIAT cassava program for their support and effort.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure

accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product

that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1807180/full#supplementary-material>

References

- Apel, K., and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701
- Apostol, I., Heinstein, P. F., and Low, P. S. (1989). Rapid stimulation of an oxidative burst during elicitation of cultured plant cells: role in defense and signal transduction. *Plant Physiol.* 90, 109–116. doi: 10.1104/pp.90.1.109
- Arora, A., Sairam, R. K., and Srivastava, G. C. (2002). Oxidative stress and antioxidative system in plants. *Curr. Sci.* 82 (10), 1227–12.
- Ayalde, D. G., Londoño, J. C. G., Mosquera, A. Q., Melendez, J. L. L., Gimode, W., Tran, T., et al. (2024). AI-powered detection and quantification of post-harvest physiological deterioration (PPD) in cassava using YOLO foundation models and K-means clustering. *Plant Methods* 20, 178. doi: 10.1186/s13007-024-01309-w
- Baguma, Y. (2004). *Regulation of starch synthesis in cassava*. Swedish University of Agricultural Sciences – Uppsala. Doctoral thesis, Swedish University of Agricultural Sciences – Uppsala.
- Bayoumi, S. A., Rowan, M. G., Beeching, J. R., and Blagbrough, I. S. (2008). Investigation of biosynthetic pathways to hydroxycoumarins during post-harvest physiological deterioration in cassava roots by using stable isotope labelling. *ChemBioChem* 9, 3013–3022. doi: 10.1002/cbic.200800515
- Beeching, J. R., Han, Y., Gomez-vasquez, R., Day, R. C., and Cooper, R. M. (1998a). Wound and defense responses in cassava as related to postharvest physiological deterioration. *Phytochem. Sig. Plant Microbe Interact.* 32, 231–248. doi: 10.1007/978-1-4615-5329-8_12
- Beeching, J. R., Li, H., Han, H., Cooper, R. M., Buschmann, H., Gomez-Vasquez, R., et al. (1998b). "Post-harvest physiological deterioration in cassava", In L. J. C. B. Carvalho, A. M. Thro and A. D. Vilarinhos (Eds.), *Cassava biotechnology: Proceedings of the IVth Scientific Meeting of the Cassava Biotechnology Network*, (Salvador, Brazil: International Center for Tropical Agriculture (CIAT)), 5–13.
- Boer, M., and Rossum, B. (2022). *LMMsolver: linear mixed model solver*, R package version 1.0.4.9000. (Biometris, Wageningen University & Research). Available online at: <https://CRAN.R-project.org/package=LMMsolver>.
- Burns, A., Gleadow, R., Cliff, J., Zacarias, A., and Cavagnaro, T. (2010). Cassava: the drought, war and famine crop in a changing world. *Sustainability* 2, 3572–3607. doi: 10.3390/su2113572
- Buschmann, H., Reilly, K., Rodriguez, M. X., Tohme, J., and Beeching, J. R. (2000). Hydrogen peroxide and flavan-3-ols in storage roots of cassava (*Manihot esculenta* Crantz) during postharvest deterioration. *J. Agric. Food. Chem.* 48, 5522–5529. doi: 10.1021/jf000513p
- Ceballos, H., Kulakow, P., and Hershey, C. (2012). Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. *Trop. Plant Biol.* 5, 73–87. doi: 10.1007/s12042-012-9094-9
- Chang-Brahim, I., Koppensteiner, L. J., Beltrame, L., Bodner, G., Saranti, A., Salzinger, J., et al. (2024). Reviewing the essential roles of remote phenotyping, GWAS and explainable AI in practical marker-assisted selection for drought-tolerant winter wheat breeding. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1319938
- Choudhary, K. K., Chaudhary, N., Agrawal, S. B., and Agrawal, M. (2017). "Reactive oxygen species: generation, damage, and quenching in plants during stress," In V. P. Singh, S. Singh, D. K. Tripathi, S. M. Prasad and D. K. Chauhan (Eds.), *Reactive Oxygen Species in Plants: Boon or Bane – Revisiting the Role of ROS* (John Wiley & Sons, Inc.), 89–115.
- Cortés Bermúdez, D. F., Reilly, K., Okogbenin, E., Beeching, J. R., Iglesias, C., and Tohme, J. (2002). Mapping wound-response genes involved in post-harvest physiological deterioration (PPD) of cassava (*Manihot esculenta* Crantz). *Euphytica* 128, 47–53. doi: 10.1023/A:1020695719706
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Dankwa, K. O., Gimode, W., and Olanmi, B. (2025). A review of global cassava (*Manihot esculenta* Crantz) production trends, post-harvest physiological deterioration (PPD) challenge, and control strategies. *Disc. Food* 5, 307. doi: 10.1007/s44187-025-00629-7
- Dankwa, K. O., and Peprah, B. B. (2019). Industrialization of cassava sector in Ghana: progress and the role of developing high starch cassava varieties. *Ghana. J. Agric. Sci.* 54, 79–85. doi: 10.4314/gjas.v54i2.8
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Djabou, A. S. M., Carvalho, L. J. C. B., Li, Q. X., Niemenak, N., and Chen, S. (2017). Cassava postharvest physiological deterioration: a complex phenomenon involving calcium signaling, reactive oxygen species and programmed cell death. *Acta Physiol. Plant* 39, 91. doi: 10.1007/s11738-017-2382-0
- Doyle, J. (1991). "DNA protocols for plants," in *Molecular techniques in taxonomy* (Springer Berlin Heidelberg, Berlin, Heidelberg), 283–293.
- Edwards, D., and Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J.* 8, 2–9. doi: 10.1111/j.1467-7652.2009.00459.x
- Esuma, W., Ozimati, A., Kulakow, P., Gore, M. A., Wolfe, M. D., Nuwamanya, E., et al. (2021). Effectiveness of genomic selection for improving provitamin A carotenoid content and associated traits in cassava. *G3 Genes|Genomes|Genetics* 11, jkab160. doi: 10.1093/g3journal/jkab160
- Farooq, M. A., Gao, S., Hassan, M. A., Huang, Z., Rasheed, A., Hearne, S., et al. (2024). Artificial intelligence in plant breeding. *Trends Genet.* 40, 891–908. doi: 10.1016/j.tig.2024.07.001
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Granato, I. S., Galli, G., de Oliveira Couto, E. G., e Souza, M. B., Mendonça, L. F., and Fritsche-Neto, R. (2018). snpReady: a tool to assist breeders in genomic analysis. *Mol. Breed.* 38, 1–7. doi: 10.1007/s11032-018-0844-8
- Han, Y., Gomez-Vasquez, R., Reilly, K., Li, H., Tohme, J., Cooper, R. M., et al. (2001). Hydroxyproline-rich glycoproteins expressed during stress responses in cassava. *Euphytica* 120, 59–70. doi: 10.1023/a:1017547419332
- Hazelett, D. J. (2025). Rethinking GWAS: how lessons from genetic screens and artificial intelligence could reveal biological mechanisms. *Bioinformatics* 41, btaf153. doi: 10.1093/bioinformatics/btaf153
- Houmani, H., Rodríguez-Ruiz, M., Palma, J. M., and Corpas, F. J. (2018). Mechanical wounding promotes local and long distance response in the halophyte *Cakile maritima* through the involvement of the ROS and RNS metabolism. *Nitric. Oxide* 74, 93–101. doi: 10.1016/j.niox.2017.06.008
- Huang, J., Bachem, C., Jacobsen, E., and Visser, R. G. F. (2001). Molecular analysis of differentially expressed genes during postharvest deterioration in cassava (*Manihot esculenta* Crantz) tuberous roots. *Euphytica* 120, 85–93. doi: 10.1023/a:101755605219
- Huang, J., McLean, G. R., and Franke, A. (2025). Twenty years of genome-wide association studies: Health translation challenges and AI opportunities. *Eur. J. Hum. Genet.* 33 (12), 1–6. doi: 10.1038/s41431-025-01951-5
- Ihemere, U., Siritunga, D., and Sayre, R. T. (2008). "Cassava," in *Compendium of transgenic crop plants. 7. Transgenic sugar, tuber and fiber crops*. Eds. C. Kole and T. C. Hall (Wiley-Blackwell, West Sussex, UK), 177–198.
- Iyer, S., Mattinson, D. S., and Fellman, J. K. (2010). Study of the early events leading to cassava root postharvest deterioration. *Trop. Plant Biol.* 3, 151–165. doi: 10.1007/s12042-010-9052-3

- Katz, S. H., and Weaver, W. W. (2020). *Cassava encyclopedia of food and culture* (New York: Schribner). Available online at: <https://www.newworldencyclopedia.org/p/index.php?title=Cassava&oldid=1030031> (Accessed December 12, 2025).
- Kulakow, P., Gedil, M., Rabbi, I., Hamblin, M., Jannink, J. L., and Ly, D. (2013). "Linkage and genome-wide association mapping using next-generation genotyping-by-sequencing in clonally propagated cassava", in: *Proceedings of the International Plant and Animal Genome Conference XXI* (San Diego, CA: Scherago International).
- Li, S., Cui, Y., Zhou, Y., Luo, Z., Liu, J., and Zhao, M. (2017). The industrial applications of cassava: current status, opportunities and prospects. *J. Sci. Food Agric.* 97, 2282–2290. doi: 10.1002/jsfa.8287
- Lokko, Y., OkogBenin, E., Mba, C., Dixon, A., Raji, A., and Fregene, M. (2007). "Cassava," in *Genome mapping and molecular breeding in plants, vol 3. Pulses, sugar and tuber crops*. Ed. C. Kole (Springer-Verlag, Berlin), 249–269.
- Lorković, Z. J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 14, 229–236. doi: 10.1016/j.tplants.2009.01.007
- Luna, J., Dufour, D., Tran, T., Pizarro, M., Calle, F., Domínguez, M. G., et al. (2021). Post-harvest physiological deterioration in several cassava genotypes over sequential harvests and effect of pruning prior to harvest. *Int. J. Food Sci. Technol.* 56, 1322–1332. doi: 10.1111/ijfs.14711
- Munyikwa, T. R. I., Langeveld, S., Salehuzzaman, S. N. I. M., Jacobsen, E., and Visser, R. G. F. (1997). Cassava starch biosynthesis: new avenues for modifying starch quantity and quality. *Euphytica* 96, 65–75. doi: 10.1023/a:1002935603412
- Naziri, D., Quayle, W., Siwoku, B., Wanlapatit, S., Viet Phu, T., and Bennett, B. (2014). The diversity of postharvest losses in cassava value chains in selected developing countries. *J. Agric. Rural Dev. Trop. Subtrop.* 115, 111–123.
- Noon, R., and Booth, R. H. (1977). Nature of post-harvest deterioration of cassava roots. *Trans. Br. Mycol. Soc.* 69, 287–290. doi: 10.1016/s0007-1536(77)80049-1
- Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. *Nature* 456, 720–723. doi: 10.1038/nature07629
- Nyikó, T., Kerényi, F., Szabadkai, L., Benkovics, A. H., Major, P., Sonkoly, B., et al. (2013). Plant nonsense-mediated mRNA decay is controlled by different autoregulatory circuits and can be induced by an EJC-like complex. *Nucleic Acids Res.* 41, 6715–6728. doi: 10.1093/nar/gkt366
- Otun, S., Eschrich, A., Achilonu, I., Rauwane, M., Lerma-Escalera, J. A., Morones-Ramírez, J. R., et al. (2023). The future of cassava in the era of biotechnology in Southern Africa. *Crit. Rev. Biotechnol.* 43, 594–612. doi: 10.1080/07388551.2022.2048791
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Yusuf, M., Aghogho, C. I., Ogunpaimo, K., et al. (2022). Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant Mol. Biol.* 109, 195–213. doi: 10.1007/s11103-020-01038-3
- Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A., et al. (2017). Genome-wide association mapping of correlated traits in cassava: Dry matter and total carotenoid content. *Plant Genome* 10, 10.3835/plantgenome2016.09.0094. doi: 10.3835/plantgenome2016.09.0094
- Rairdin, A., Fotouhi, F., Zhang, J., Mueller, D. S., Ganapathysubramanian, B., Singh, A. K., et al. (2022). Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean. *Front. Plant Sci.* 13, 966244. doi: 10.3389/fpls.2022.966244
- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01852
- R Core Team (2021). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available online at: <https://www.R-project.org/> (Accessed December 20, 2025).
- Reilly, K., Gomez-Vasquez, R., Buschmann, H., Tohme, J., and Beeching, J. R. (2004). Oxidative stress responses during cassava post-harvest physiological deterioration. *Plant Mol. Biol.* 56, 625–641. doi: 10.1007/s11103-005-2271-6
- Reilly, K., Han, Y., Tohme, J., and Beeching, J. R. (2001). Isolation and characterisation of a cassava catalase expressed during post-harvest physiological deterioration. *Biochim. Biophys. Acta* 1518, 317–323. doi: 10.1016/s0167-4781(01)00195-6
- Reiser, L., Bakker, E., Subramaniam, S., Chen, X., Sawant, S., Khosa, K., et al. (2024). The Arabidopsis information resource in 2024. *Genetics* 227, iyae027. doi: 10.1093/genetics/iyae027
- Salcedo, A., and Siritunga, D. (2011). Insights into the physiological, biochemical and molecular basis of postharvest deterioration in cassava (*Manihot esculenta*) roots. *Am. J. Exp. Agric.* 1, 414. doi: 10.9734/ajea/2011/784
- Sanchez, T., Chávez, A. L., Ceballos, H., Rodriguez-Amaya, D. B., Nestel, P., and Ishitani, M. (2005). Reduction or delay of post-harvest physiological deterioration in cassava roots with higher carotenoid content. *J. Sci. Food Agric.* 86, 634–639.
- Sánchez, T., Dufour, D., Moreno, J. L., Pizarro, M., Aragón, I. J., Domínguez, M., et al. (2013). Changes in extended shelf life of cassava roots during storage in ambient conditions. *Postharv. Biol. Technol.* 86, 520–528. doi: 10.1016/j.postharvbio.2013.07.014
- Sangjan, W., Kick, D. R., and Washburn, J. D. (2025). Improving plant breeding through AI-supported data integration. *Theor. Appl. Genet.* 138, 132. doi: 10.1007/s00122-025-04910-2
- Sayre, R., Beeching, J. R., Cahoon, E. B., Egesi, C., Fauquet, C., Fellman, J., et al. (2011). The BioCassava Plus Program: Biofortification of cassava for Sub-Saharan Africa. *Annual Rev. Plant Biol.* 62, 251–272.
- Seki, K., and Toda, Y. (2022). QTL mapping for seed morphology using the instance segmentation neural network in *Lactuca* spp. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.949470
- Singh, R., Singh, S., Parihar, P., Mishra, R. K., Tripathi, D. K., Singh, V. P., et al. (2016). Reactive oxygen species (ROS): Beneficial companions of plants' developmental processes. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01299
- Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., et al. (2023). SRplot: A free online platform for data visualization and graphing. *PloS One* 18, e0294236. doi: 10.1371/journal.pone.0294236
- Tripathi, D. K., Singh, S., Singh, V. P., Prasad, S. M., Dubey, N. K., and Chauhan, D. K. (2017). Silicon nanoparticles more effectively alleviated UV-B stress than silicon in wheat (*Triticum aestivum*) seedlings. *Plant Physiol. Biochem.* 110, 70–81. doi: 10.1016/j.plaphy.2016.06.026
- Tripathy, B. C., and Oelmüller, R. (2012). Reactive oxygen species generation and signaling in plants. *Plant Signaling Behav.* 7, 1621–1633. doi: 10.4161/psb.22455
- Wang, X., Feng, C., Tian, L., Hou, C., Tian, W., Hu, B., et al. (2021). A transceptor-channel complex couples nitrate sensing to calcium signaling in *Arabidopsis*. *Mol. Plant* 14, 774–786. doi: 10.1016/j.molp.2021.02.005
- Wang, J., Tang, Y., and Zhang, Z. (2022). "Performing genome-wide association studies with multiple models using GAPIT," in *Genome-wide association studies* (Springer US, New York, NY), 199–217.
- Wang, J., and Zhang, Z. (2021). GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genom. Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005
- Wheatley, C. (1982). Studies on cassava (*Manihot esculenta* Crantz) root post-harvest deterioration. University of London, UK, 246. Ph.D. thesis.
- Wheatley, C., and Gomez, G. (1985). Evaluation of some quality characteristics in cassava storage roots. *Plant Food. Hum. Nutr.* 35, 121–129. doi: 10.1007/bf01092127
- Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9, plantgenome2015.11.0118. doi: 10.3835/plantgenome2015.11.0118
- Zhang, Z. (2014). "GAPIT.kinship.ZHANG.R," in *Zhiwu-zhang-lab/GAPIT*. Available online at: <https://github.com/Zhiwu-Zhang-Lab/GAPIT/blob/master/GAPIT.kinship.ZHANG.R> (Accessed October 25, 2014).
- Zhang, S., Chen, X., Lu, C., Ye, J., Zou, M., Lu, K., et al. (2018). Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* Crantz). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00503
- Zhang, Y., Yang, J., Showalter, A. M., and Zhao, J. (2016). Interactions between plant RRM-containing proteins and stress responses. *Plant Physiol. Biochem.* 109, 134–142. doi: 10.1016/j.plaphy.2016.09.008
- Zidenga, T., Leyva-Guerrero, E., Moon, H., Siritunga, D., and Sayre, R. (2012). Extending cassava root shelf life via reduction of reactive oxygen species production. *Plant Physiol.* 159, 1396–1407. doi: 10.1104/pp.112.200345