ISSN: 2317-0840

Index for identifying relevant and stable genomic regions in analysis with multiple genomic association methods

Renata Dourado Roque[1], Brenda Vieira de Oliveira[1], Guilherme Flaviano Pereira[1], Camila Ferreira Azevedo[1†]

[1] *Federal University of Viçosa; Center for Exact and Technological Sciences; Graduate Program in Applied Statistics and Biometry; Viçosa – Minas Gerais, Brazil.*

Abstract: *Genome-Wide Association Studies (GWAS) aim to identify associations between quantitative trait loci (QTL) and phenotypes. These studies are of great interest to genetic breeding programs as they enable the identification of markers associated with agronomically important traits. Commonly used GWAS methods include the Mixed Linear Model (MLM), Compressed MLM (CMLM), General Linear Model (GLM), Settlement of MLM Under Progressively Exclusive Relationship (SUPER), Multiple Locus Mixed Linear Model (MLMM), and FarmCPU. These methods differ in their theoretical foundations, leading to variations in the genomic regions detected. This study aimed to develop two indices, one unweighted and one weighted based on the importance of each region, to identify significant and stable genomic regions across different statistical methods applied to GWAS. The objective was to minimize the detection of potentially false-positives regions. The analysis was conducted on a dataset comprising 413 Asian rice (Oryza sativa) individuals genotyped with 36,901 single nucleotide polymorphisms (SNPs). A total of 11 phenotypic traits were evaluated, and the six aforementioned methods were compared. Among the analyzed traits, four stood out for exhibiting a higher number of significant genomic regions detected by multiple methods. The SUPER and GLM methods proved to be the most effective in idntifying genomic associations. The proposed indices identified genomic regions previously described in the literature, associated with traits such as blast resistance, fertility, and plant height. The weighted index demonstrated greater sensitivity in detecting significant genomic regions due to its differentiated weighting, which prioritized the most important regions.*

Keywords: *genotyping; molecular markers; genetic breeding.*

Introduction

Advances in biotechnology, especially in the development and application of molecular markers such as Single Nucleotide Polymorphisms (SNPs), have enabled the identification of individuals based on their genetic variations, driving Genome-Wide Association Studies (GWAS) (TIBBS CORTES; ZHANG; YU, 2021). GWAS aims to identify associations between Quantitative Trait Loci (QTL) and genetic values associated with traits of interest. In practice, these associations are established between SNPs and phenotypes based on the assumption of Linkage Disequilibrium (LD) between markers and QTLs. This approach provides detailed insights into genomic regions that directly influence target traits and is, therefore, of great interest to genetic breeding programs.

In the context of genetic breeding, GWAS stands out as an essential tool for understanding the genetic architecture of complex traits, enabling the identification of associated genes (*gene mining*) and the comprehension of the molecular mechanisms that control these traits. Additionally, GWAS facilitates the application of marker-assisted selection (MAS), which uses molecular markers to efficiently identify and select individuals

[†] **Corresponding author: camila.azevedo@ufv.br**

Manuscript received: 23/12/2024; Revised: 14/03/2025; Accepted: 25/03/2025.

with higher genetic potential at a lower cost compared to genomic prediction (BOOPATHI, 2020).

Several statistical approaches have been proposed for GWAS, with mixed linear model (MLM)-based methods standing out due to their ability to control factors that can lead to false positives in analyses, such as population structure and familial relatedness. This is because individuals within the same population often share large genomic blocks due to common ancestry (ZHANG et al., 2010). The *General Linear Model* (GLM) is the simplest GWAS method, incorporating population structure correction without including random effects in the model. On the other hand, widely used mixed-model-based methods include the *Mixed Linear Model* (MLM), *Compressed MLM* (CMLM), *Settlement of MLM Under Progressively Exclusive Relationship* (SUPER), *Multiple Locus Mixed Linear Model* (MLMM), and *FarmCPU.*

Each of these methods presents theoretical distinctions and different approaches to correcting population structure and relatedness, consequently influencing the control of false positives and computational time (TIBBS CORTES; ZHANG; YU, 2021). As a result, variations occur in the genomic regions detected. These methods have been applied to various phenotypic traits and crops, such as grain yield and traits in Indian mustard (AKHATAR; BANGA, 2015), salt tolerance in soybean germplasm lines (ZENG et al., 2017), yield-related traits in wheat (MALIK et al., 2021), chlorophyll content in maize (XIONG et al., 2023), genetic mechanisms regulating panicle architecture in rice (ZHONG et al., 2021), and milk production in French dairy cattle breeds (TEISSIER et al., 2018).

The different genomic regions detected by distinct statistical methodologies can complicate the selection of genomic regions to be used in breeding programs. Consequently, there is a need to develop procedures that facilitate decision-making for breeders. In this context, inspired by selection indices developed for choosing genetically superior individuals (HAZEL, 1943; SMITH, 1936), an index for selecting genomic regions has been proposed. This index aims to identify significant and stable genomic regions, reducing effort, time, and resources in analyses that could lead to false positives while prioritizing those that truly deserve further investigation.

The present study used a dataset consisting of 413 Asian rice (*Oryza sativa*) individuals, genotyped for 36,901 SNP markers. Eleven phenotypic traits related to productivity, morphology, grain quality, and disease resistance were evaluated to assess the proposed selection indices.

Materials and Methods

Data

The dataset used in this research consists to 413 individuals of Asian rice (*Oryza sativa*), genotyped for 36,901 SNP markers. The data are publicly available as part of the OryzaSNP Project and the OMAP Project (AMMIRAJU et al., 2006; ZHAO et al., 2011) and can be accessed at https://ricediversity.org/data/. Quality control of the markers was performed by excluding those with a call rate lower than 70% and a minor allele frequency below 1%.

The experiments were conducted in Arkansas, United States, from May to fall in 2006 and 2007. Each experiment included two replications per year, arranged in a randomized

complete block design. Each plot consisted of rows 5 meters long, with a spacing of 25 cm between plants and 50 cm between rows (ZHAO et al., 2011).

The evaluated phenotypic traits were related to productivity, morphology, grain quality, and disease resistance. These included flag leaf dimensions (length and width), number of flowers and seeds per panicle, panicle fertility and length, number of panicles per plant, plant height, number of primary branches on the panicle, protein content, and blast resistance.

Six statistical methods were evaluated: the Mixed Linear Model (MLM), Compressed MLM (CMLM), General Linear Model (GLM), Settlement of MLM Under Progressively Exclusive Relationship (SUPER), Multiple Locus Mixed Linear Model (MLMM), and FarmCPU, all implemented using the GAPIT package in R (WANG; ZHANG, 2021). These methods were selected based on their theoretical differences and specific modeling approaches to control the false positive rate and improve accuracy in detecting significant genomic regions.

GLM and MLM are widely used as they account for population structure, reducing the false positive rate (PRICE et al., 2006). CMLM optimizes MLM by grouping similar individuals, thereby reducing computational effort (ZHANG et al., 2010). MLMM facilitates the mapping of complex traits controlled by multiple genetic loci (KALER et al., 2020). SUPER selects a subset of markers to define the genomic relationship matrix, increasing statistical power (WANG et al., 2014). FarmCPU separates fixed and random effects in the model, enhancing the control of confounding factors between test markers and kinship, thus enhancing computational efficiency (LIU et al., 2016).

Methods used for GWAS analysis

General Linear Model

The General Linear Model (GLM) is a regression approach applied to individual markers, incorporating corrections for population structure by extracting the genomic relationship matrix and the first principal components to reduce spurious associations (PRICE et al., 2006). It is defined as:

$$y = 1\mu + Qq + M_i m_i + e$$

where y is the vector of phenotypic observations, 1 is the vector whose elements are all equal to 1, μ is the overall mean, q is the vector containing the first q principal components of the genomic relationship matrix (K), obtained as proposed by Vanraden (2008), and m_i is the fixed effect of the i-th marker. Q and $M_i$are the incidence matrices of their respective effects. The random error term follows a normal distribution: $e\,N(0, I\sigma_e^2)$, where I is the identity matrix and σ_e^2 is the residual variance.

Mixed Linear Model

The Mixed Linear Model (MLM) is a regression approach applied to individual markers, incorporating corrections for population structure by extracting the first principal components of the genomic relationship matrix, reducing spurious associations (PRICE et al., 2006). Additionally, random polygenic effects are included to account for kinship differences

between individuals, further reducing the occurrence of false positives (YU et al., 2006). This model is defined by:

$$y = 1\mu + Qq + Zu + M_i m_i + e$$

where y is the vector of phenotypic observations, 1 is the vector whose elements are all equal to 1, μ is the overall mean, q is the vector containing the first q principal components of the genomic relationship matrix (K), u is the vector of polygenic effects, and m_i is the fixed effect of the i-th marker. Q, $M_i$and Z are the incidence matrices of their respective effects. The probability distributions of the random effects are given by: $u\,N(0, K\sigma_u^2)$ and $e\,N(0, I\sigma_e^2)$, where I is the identity matrix**,** K is the genomic relationship matrix**,** σ_u^2 is the variance of the polygenic effects and σ_e^2 is the residual variance.

Compressed MLM

The Compressed Mixed Linear Model (CMLM) was developed to prevent the double counting of relationship information in the MLM, which is accounted for both in the principal components and in the polygenic effects (LI et al., 2014). In this method, a clustering analysis is performed using the genomic relationship matrix as a measure of similarity, applaying the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm to group similar individuals. Once individuals are assigned to groups, summary statistics of the relationship between and within groups are used to construct a reduced relationship matrix. This procedure is repeated for different compression levels, generating a reduced relationship matrix ($K_{reduzida}$).

Multiple Locus Mixed Linear

The Multiple Locus Mixed Linear Model (MLMM) combines MLM regression with an interactive stepwise selection process (forward-backward) (SEGURA et al., 2012). This approach enables the identification of multiple genetic loci by incorporating the effects of associated markers as covariates in the model, helping to control false positives.

SUPER

The SUPER method (Settlement of MLM Under Progressively Exclusive Relationship) was developed to enhance the detection of significant genomic regions, control the false positive rate, and reduce p-value inflation (KALER et al., 2020). In this method, the genome is divided into segments of equal size, with each segment represented by the most significant marker within it. Additionally, the size and number of selected segments are estimated using the maximum likelihood method in a random model, where the relationship matrix is derived from these segments. SUPER constructs a relationship matrix using only the associated markers while excluding those in strong linkage disequilibrium (LD) with the tested markers.

FarmCPU

The FarmCPU model (*Fixed and Random Model Circulating Probability Unification*) (LIU et al., 2016), developed based on the MLMM method, is an approach that optimizes the selection of associated markers, known as pseudo-quantitative trait nucleotides (pseudo-QTNs), and uses them as covariates to reduce confounding between test markers and those assumed as covariates. The method operates in two main stages: i) Fixed Effect Model (FEM): Each marker is tested individually while incorporating previously identified associated markers as covariates, helping to control false positives. ii) Random Effect Model (REM): The effects of the markers are estimated, and these estimates are used to define the relationship matrix. These stages are executed iteratively, allowing the progressive refinement of associated markers and improving the accuracy of the analyses throughout the process.

Selection Index

The index I is inspired by selection indices developed for identifying genetically superior individuals (HAZEL, 1943; SMITH, 1936) and is expressed as a linear combination of the statistical methods applied in the GWAS analysis, as described by:

$$I_j = b_{1j}m_{1j} + b_{2j}m_{2j} + b_{3j}m_{3j} + b_{4j}m_{4j} + b_{5j}m_{5j} + b_{6j}m_{6j}$$

where I_j is the index value for the j-th marker ($j = 1, \ldots, 36.901$), b_{ij} are the weighting coefficients of the i-th marker and j-th method ($i = 1, \ldots, 6$), and m_{ij} is an indicator variable defined as:

$$m_{ij} = \begin{cases} 1, \textit{if SNP}_j \textit{ was detected by method } j \\ 0, \textit{if SNP}_j \textit{ was not detected by method } j \end{cases}$$

Two selection indices for genomic regions were created. The first index assigned equal weights ($b_{ij} = 1$), considering only the frequency of detection of the regions. In the second index, weights were assigned to reflect the importance of each region, defined as:

$$b_{ij} = \frac{-\log\left(p - value_{ij}\right)}{\sum_{i=1}^{6} \log\left(p - value_{ij}\right)}$$

where $p - value_{ij}$ is the p-value associated with the i-th method and j-th marker.

Genomic regions were defined across the 12 chromosomes, each with a size of 0.69 Mb, as evaluated by Suela et al. (2022).

Results and Discussion

When comparing the performance of the methods, GLM and SUPER identified the highest number of associations, while MLM and CMLM detected fewer (Figure 1). High linkage disequilibrium (LD) in a given genomic region can result in inflated estimates of SNP effects (TRYNKA et al., 2015). As a result, SNPs may be erroneously identified as associated with a phenotype even in the absence of a true causal relationship. This occurs because the SNPs may be in proximity to a QTL (*Quantitative Trait Locus*), but the observed association arises from the correlation between markers rather than direct causation. According to Suela et al. (2022), the average LD for markers within a 0.69 Mb region is approximately 0.20 in this genomic dataset.

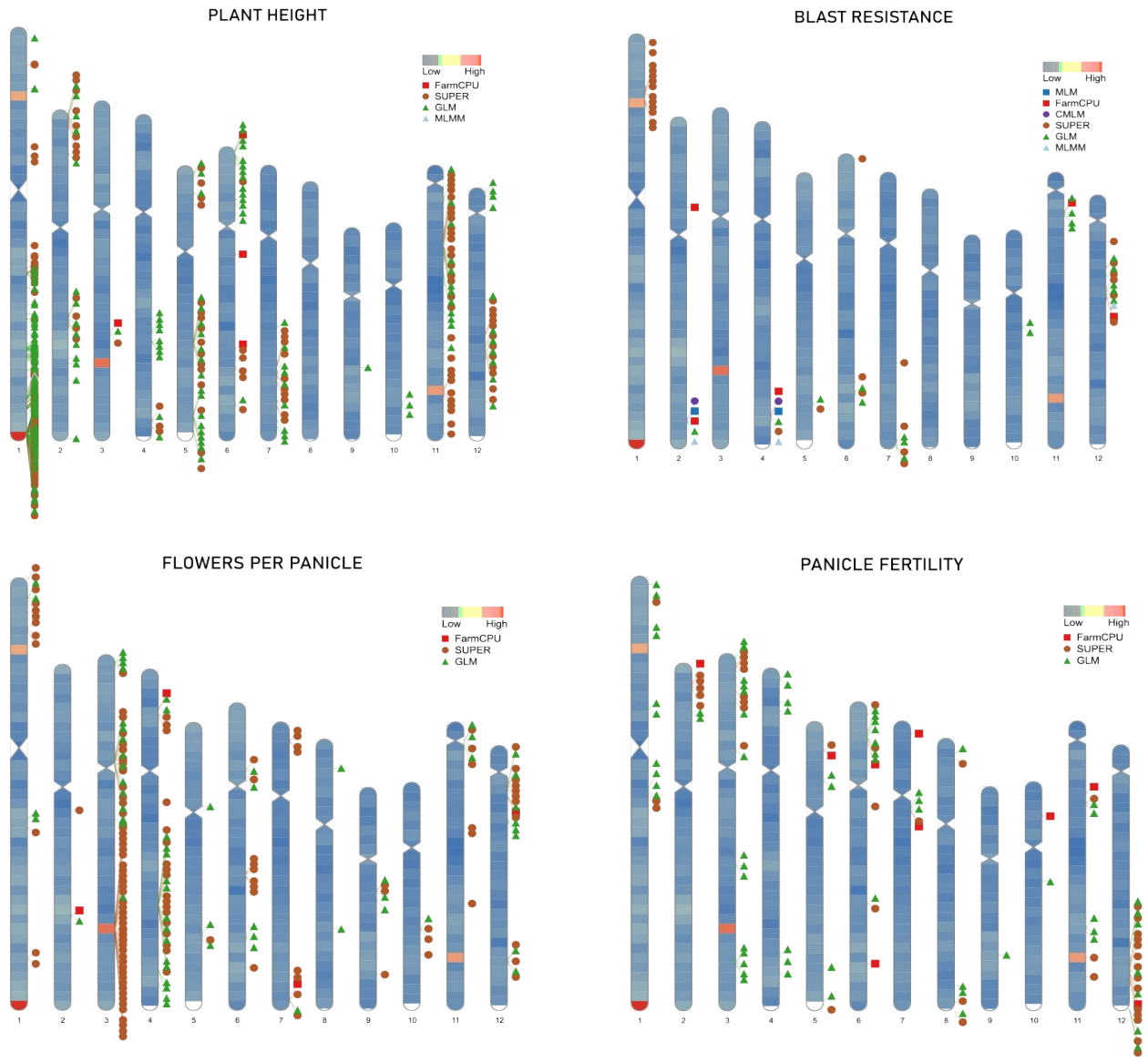
For example, for the traits plant height and number of flowers per panicle, a high detection of genomic regions was observed on chromosomes 1 and 3, respectively, using the GLM and SUPER methods. Another biological factor contributing to a high rate of false positives is population heterogeneity, which arises from population structure and relatedness among individuals (HALDAR; GHOSH, 2012). Genetic differences between subgroups within a population can lead to spurious associations, where the observed associations are caused by population variation rather than a true relationship between the SNP and the phenotypic trait. While the GLM corrects only for population structure, the SUPER method eliminates markers with high LD, which may still be insufficient to fully minimize false positives.

Among the phenotypic traits analyzed, four stood out for having a larger number of significant genomic regions. In particular, a region on chromosome 4, associated with blast resistance, was identified by all methods, as illustrated in Figure 1.

Figures 2 and 3 illustrate the intensity of the indices for each genomic region using a color scale. The color of each region reflects its relevance according to each index, with red indicating the regions that breeders should prioritize. When comparing the color/intensity patterns of the indices, variations across traits were observed. For blast resistance and the number of flowers per panicle, only one region appeared in red in both indices, located in the initial portions of chromosomes 4 and 7, respectively. However, for panicle fertility, the highlighted red regions diverged. In the unweighted index, these regions were identified on chromosomes 7 and 12, while in the weighted index, they were located on chromosomes 5 and 11.

It is important to note that the detection of a genomic region by statistical methods does not necessarily indicate its importance for the genetic variation of the trait. The p-value, widely used in the literature (ESPOSITO et al., 2023; SU et al., 2014), serve as a key indicator of marker relevance for phenotypic variation. Markers identified through GWAS, representing potential QTLs, can be effectively utilized in marker-assisted selection only if they explain a significant proportion of the trait's genetic variation (O'CONNOR et al., 2020). In such cases, individuals can be discriminated and categorized into selected and non-selected groups based on their allele dosages.

Figure 1: SNP density, represented on a scale from high (red) to low (blue), with geometric shapes indicating significant genomic regions detected by each method.



Source: from the authors (2025).

The indices applied in this study identified highly relevant genomic regions (marked in red), which had been previously reported in the literature. Traits for which already cataloged regions were detected using the unweighted index (Figure 2) included resistance to blast, fertility, and plant height. The weighted index (Figure 3), in turn, identified a higher number of highly relevant genomic regions associated with these same traits.

For blast resistance, significant regions were observed on chromosomes 4 and 12 (FUKUOKA; OKUNO, 2001; WANG et al., 1994), with the latter region showing greater relevance only in the weighted index. Notably, blast resistance is one of the most critical traits due to the substantial impact of this disease on productivity.

Regarding plant height, associations were observed on chromosomes 1 and 6, with greater relevance detected exclusively in the weighted index. These regions had been previously described by Marri et al. (2005) and Mei et al. (2003), emphasizing the importance of this phenotype in rice yield (ZHANG et al., 2017).

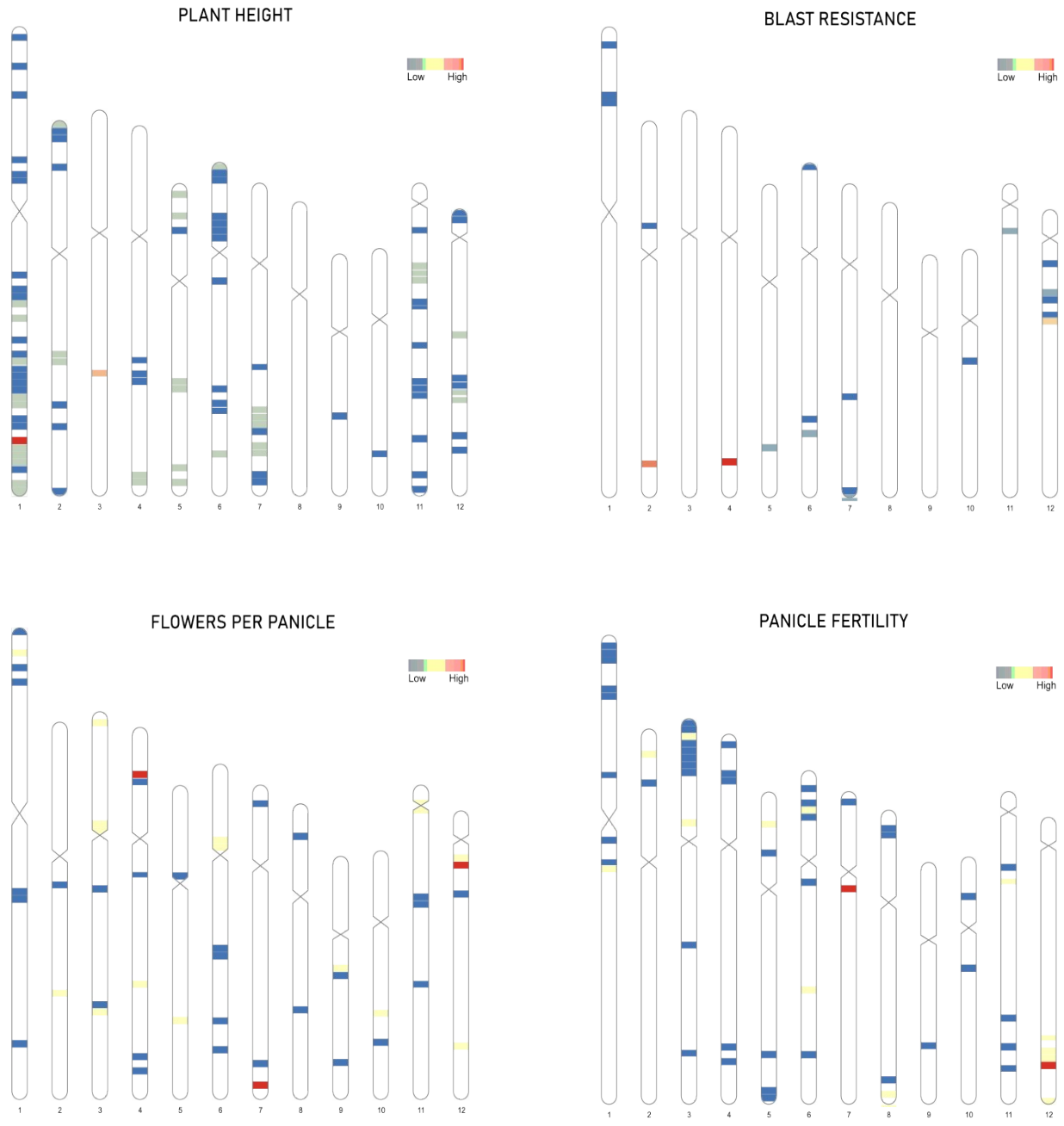
For fertility, high intensity regions were mapped on chromosomes 5, 8, and 11. On chromosome 5, relevance was detected only in the weighted index, as was the case for chromosome 11. These regions had been previously reported by He Yu-Qing (2000) and Mei et al. (2003).

Conversely, none of the methodologies employed in this study were able to identify stable and highly relevant regions associated with the number of flowers per panicle.

This study introduced indices that enabled the identification of significant and stable genomic regions based on the different methods evaluated. The indices were weighted according to the p-values of marker effects on the phenotype, following the rationale that these values reflect the importance of each marker for the trait. However, future studies could explore alternative weighting methods, such as the coefficient of determination, genetic variance attributed to the region, or the effect size of the region on the phenotype.

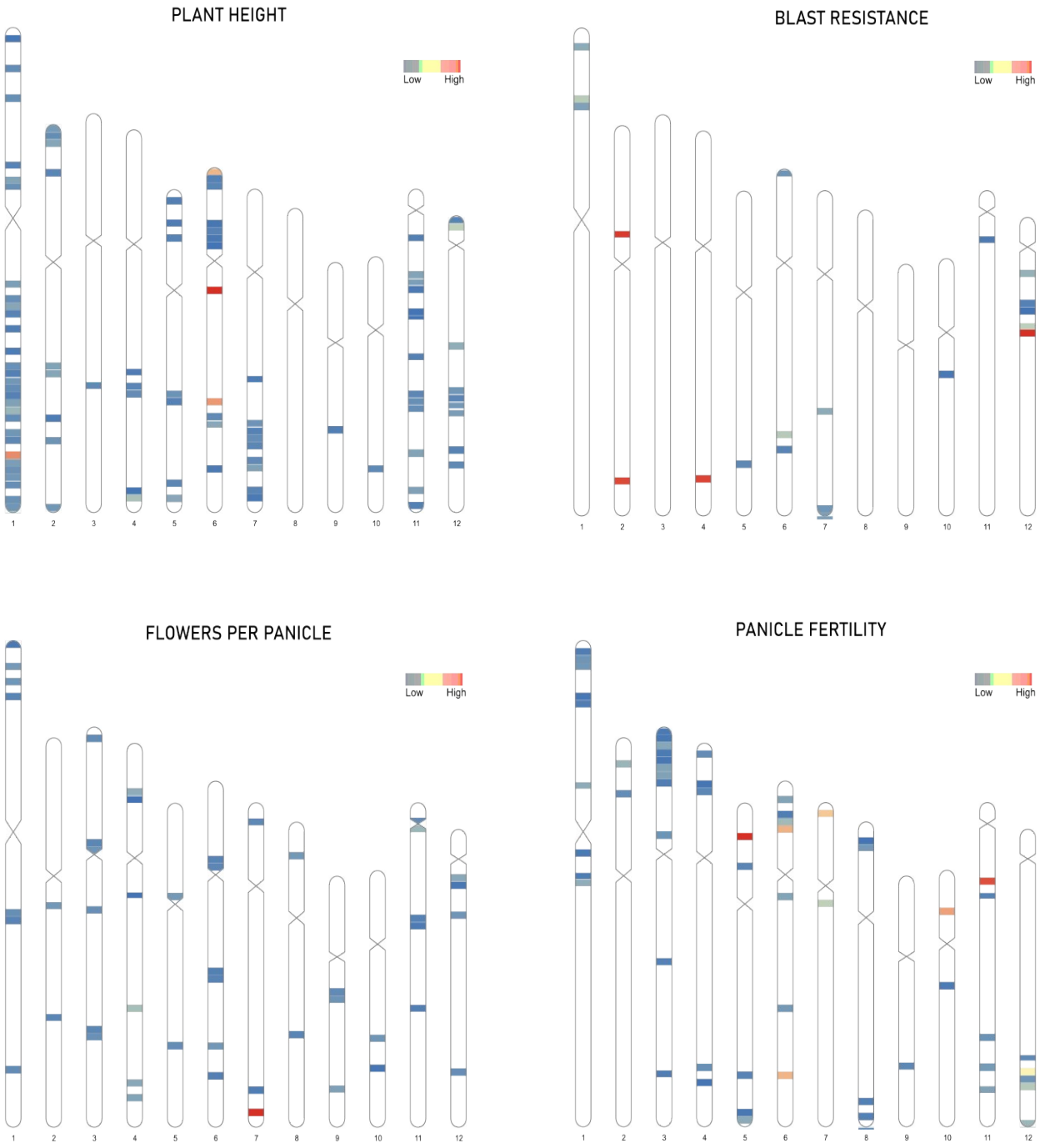
The stability considered here pertains to the statistical methods and the underlying theories. However, genetic stability is another crucial factor. To expand our understanding of genome-wide association, future studies should not be limited to a single generation or a single year of evaluation. Genetic variations specific to different populations can directly influencing the significant genomic regions detected, which may fluctuate due to genetic structure and environmental factors (WOJCIK et al., 2019). A more comprehensive analysis would require applying these indices to diverse populations to assess whether important genomic regions remain stable across both statistical methods and genetics contexts.

Figure 2: Index intensity without weighing each genomic region across the 12 chromosomes, considering a 0.69 Mb window size, displayed on a scale from high (red) to low (blue).



Source: from the authors (2025).

Figure 3: Index intensity weighted for each genomic region across the 12 chromosomes, considering a 0.69 Mb window size, displayed on a scale from high (red) to low (blue).



Source: from the authors (2025).

Conclusions

This research highlights the importance of using multiple statistical methods in association studies to more efficiently identify potential significant regions while reducing efforts on false positives. The indices, with and without weighting, present distinct patterns regarding the relevance of genomic regions. However, weighting the regions in the index is recommended to indicate that a given region explains a reasonable proportion of the genetic variation of the traits.

Overall, the creation of these indices provided better visualization of potential associated genomic regions, with the weighted index demonstrating greater sensitivity in detecting significant regions due to the different weights assigned, emphasizing the most relevant regions. These findings offer valuable insights that can be explored in future research and applied in breeding programs to facilitate the identification of individuals with desirable traits.

Future studies should evaluate alternative weighting approaches and apply these indices across different generations of a breeding program to assess the stability of genomic regions over time, contributing to a deeper understanding of the genetic architecture of these traits.

Acknowledgments

The authors thank the National Council for Scientific and Technological Development (CNPq) and the Research Support Foundation of the State of Minas Gerais (FAPEMIG) for their financial support.

References

AKHATAR, J.; BANGA, S. S. Genome-wide association mapping for grain yield components and root traits in Brassica juncea (L.) Czern & Coss. **Molecular Breeding**, v. 35, n. 1, p. 48, 2015.

AMMIRAJU, J. S. S. et al. The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. **Genome research**, v. 16, n. 1, p. 140–147, jan. 2006.

BOOPATHI, N. M. Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits. In: Singapore: Springer, 2020. p. 343–388.

ESPOSITO, S. et al. Simultaneous improvement of grain yield and grain protein concentration in durum wheat by using association tests and weighted GBLUP. **TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik**, v. 136, n. 12, 1 dez. 2023.

FUKUOKA, S.; OKUNO, K. QTL analysis and mapping of pi21, a recessive gene for field resistance to rice blast in Japanese upland rice. **Theoretical and Applied Genetics**, v. 103, n. 2, p. 185–190, 2001.

HALDAR, T.; GHOSH, S. Effect of Population Stratification On False Positive Rates Of Population-based Association Analyses Of Quantitative Traits. **Annals of Human Genetics**, v. 76, n. 3, p. 237, maio 2012.

HAZEL, L. N. THE GENETIC BASIS FOR CONSTRUCTING SELECTION INDEXES. **Genetics**, v. 28, n. 6, p. 476–490, 20 nov. 1943.

KALER, A. S. et al. Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. **Frontiers in Plant Science**, v. 10, 2020.

LI, M. et al. Enrichment of statistical power for genome-wide association studies. **BMC Biology**, v. 12, n. 1, p. 73, 2014.

LIU, X. et al. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. **PLOS Genetics**, v. 12, n. 2, p. e1005767, 1 fev. 2016.

MALIK, P. et al. Single-trait, multi-locus and multi-trait GWAS using four different models for yield traits in bread wheat. **Molecular Breeding**, v. 41, n. 7, p. 46, 2021.

MARRI, P. R. et al. Identification and mapping of yield and yield related QTLs from an Indian accession of Oryza rufipogon. **BMC Genetics**, v. 6, n. 1, p. 33, 2005.

MEI, H. W. et al. Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. **Theoretical and Applied Genetics**, v. 107, n. 1, p. 89–101, 2003.

O'CONNOR, K. et al. Genome-wide association studies for yield component traits in a macadamia breeding population. **BMC Genomics**, v. 21, n. 1, p. 199, 2020.

PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics 2006 38:8**, v. 38, n. 8, p. 904–909, 23 jul. 2006.

SEGURA, V. et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. **Nature Genetics**, v. 44, n. 7, p. 825–830, 2012.

SMITH, H. F. A DISCRIMINANT FUNCTION FOR PLANT SELECTION. **Annals of Eugenics**, v. 7, n. 3, p. 240–250, 1 nov. 1936.

SU, G. et al. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. **Journal of Dairy Science**, v. 97, n. 10, p. 6547–6559, 2014.

SUELA, M. M. et al. Regional heritability mapping and genome-wide association identify loci for rice traits. **Crop Science**, v. 62, n. 2, p. 839–858, 1 mar. 2022.

TEISSIER, M. et al. Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: An application in milk production of French dairy cattle breeds. **Journal of Dairy Science**, v. 101, n. 4, p. 3126–3139, 2018.

TIBBS CORTES, L.; ZHANG, Z.; YU, J. Status and prospects of genome-wide association studies in plants. **The Plant Genome**, v. 14, n. 1, p. e20077, 1 mar. 2021.

TRYNKA, G. et al. Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. **The American Journal of Human Genetics**, v. 97, n. 1, p. 139–152, 2 jul. 2015.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008.

WANG, G. L. et al. RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. **Genetics**, v. 136, n. 4, p. 1421–1434, 1 abr. 1994.

WANG, J.; ZHANG, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. **Genomics, Proteomics & Bioinformatics**, v. 19, n. 4, p. 629–640, 1 ago. 2021.

WANG, Q. et al. A SUPER Powerful Method for Genome Wide Association Study. **PLoS ONE**, v. 9, n. 9, p. e107684, 23 set. 2014.

WOJCIK, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. **Nature**, v. 570, n. 7762, p. 514–518, 2019.

XIONG, X. et al. Genetic dissection of maize (Zea mays L.) chlorophyll content using multi-locus genome-wide association studies. **BMC Genomics**, v. 24, n. 1, p. 384, 2023.

YU, J. et al. A unified mixed-model method for association mapping that accounts for

multiple levels of relatedness. **Nature Genetics**, v. 38, n. 2, p. 203–208, 2006.
ZENG, A. et al. Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. **Molecular Breeding**, v. 37, n. 3, p. 30, 2017.
ZHANG, Y. et al. Os MPH1 regulates plant height and improves grain yield in rice. **PLOS ONE**, v. 12, n. 7, p. e0180825, 14 jul. 2017.
ZHANG, Z. et al. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, v. 42, n. 4, p. 355–360, 2010.
ZHAO, K. et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. **Nature Communications**, v. 2, n. 1, p. 467, 2011.
ZHONG, H. et al. Uncovering the genetic mechanisms regulating panicle architecture in rice with GPWAS and GWAS. **BMC Genomics**, v. 22, n. 1, p. 86, 2021.