SOFTWARE RELEASE

GenomicLand: A user-friendly tool for genomic prediction and association studies

Camila F. Azevedo[1]✉, Moyses Nascimento[1], Vitor C. Fontes[1], Marcos Deon V. Resende[2] and Cosme D. Cruz[3]

Crop Breeding and Applied Biotechnology
26(3): e5501263, 2026
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332026v26n3s33

Abstract: *GenomicLand is a free, open-source platform for genomic prediction and genome-wide association studies. Developed in R software using Shiny, it supports diploid and polyploid datasets and integrates statistical learning, Bayesian, machine learning, and neural network methods. GenomicLand provides an accessible environment for researchers, breeders, and students.*

Keywords: *Mixed models, statistical learning, machine learning, Bayesian inference, plant breeding*

INTRODUCTION

GenomicLand is a free software platform developed for genomic prediction and genome-wide association studies (GWAS). The first version was released in 2019 (Azevedo et al. 2019), combining R for analytical methods and Python for the graphical user interface. The current version has been fully migrated to R using the Shiny framework, improving accessibility, integration, and usability. *GenomicLand* provides a unified graphical interface compatible with Windows and Linux operating systems and supports phenotypic and genomic data from diploid and polyploid species.

Genome-wide selection (GWS) and GWAS are key tools in modern breeding programs. These approaches improve the accuracy of genetic value prediction (Meuwissen et al. 2001, Wellmann and Bennewitz 2012), accelerate the identification of superior genotypes (Vandenplas et al. 2018), increase genetic gain, and reduce generation intervals (Resende et al 2012). Additionally, they support parental selection by identifying individuals with desirable genetic profiles for future crosses. GWAS also contributes to understanding the genetic architecture of complex traits by identifying genomic regions associated with phenotypic variation (Uffelmann et al. 2021). However, implementing these methodologies often requires programming expertise and the integration of multiple tools, which may limit their accessibility to a broader user community.

The availability of free and user-friendly software is essential for expanding access to advanced genomic analyses. *GenomicLand* addresses this gap by integrating statistical learning, mixed models, Bayesian inference, machine learning, and artificial intelligence methods into a single analytical environment, enabling genomic analyses without requiring advanced programming skills.

***Corresponding author:**
E-mail: camila.azevedo@ufv.br

Scientific Editor:
Luiz Antônio dos Santos Dias

Received: 20 January 2026
Accepted: 25 May 2026
Published: 16 June 2026

[1] Universidade Federal de Viçosa (UFV), Departamento de Estatística, Avenida PH Rolfs, s/n, 36570-000, Viçosa, MG, Brazil
[2] UFV, Engenharia Florestal, Avenida PH Rolfs, s/n, 36570-000, Viçosa, MG, Brazil
[3] UFV, Departamento de Biologia Geral, Avenida PH Rolfs, s/n, 36570-000, Viçosa, MG, Brasil

The current version expands its capabilities by supporting additive and non-additive models, diploid and polyploid species, genomic prediction, GWAS, genotype-by-environment interaction analyses, and multivariate approaches for multi-trait and multi-environment data.

SOFTWARE ARCHITECTURE

GenomicLand was developed entirely in the R programming language using the Shiny framework. The software architecture adopts a modular architecture that integrates data processing, statistical modeling, and visualization within a unified analytical environment (Figure 1A).

The system architecture comprises three main layers: i) a user interface layer, which allows users to upload datasets, configure model parameters, and visualize analytical results through a graphical interface; ii) a computational layer, responsible for executing statistical and machine learning analyses using established R packages and custom routines; and iii) a visualization layer, which generates graphical outputs such as heatmaps, GWAS Manhattan plots, linkage disequilibrium (LD) decay curves, and model diagnostic plots.

GenomicLand is available for Windows and Linux operating systems; it is freely accessible at https://www.licae.ufv.br/genomicland/. The software supports genotype and phenotype datasets in standard tabular formats (e.g., CSV, TXT, and XLSX files) and provides example datasets to facilitate testing and reproducibility.

This architecture facilitates the seamless integration of multiple analytical approaches, thereby reducing the need for external tools and simplifying complex genomic analyses for users with varying levels of programming experience. This integrated design distinguishes *GenomicLand* from single-purpose tools by providing an integrated environment for end-to-end genomic analyses.

INPUT DATA STRUCTURE

GenomicLand requires genotype and phenotype data in standardized tabular formats to ensure compatibility across all analytical modules.

Genotype data

The genotype file is organized with genetic markers (single nucleotide polymorphisms, SNPs) in rows and individuals in columns. The first columns contain marker information, including the SNP identifier, chromosome, and genomic position, expressed either as a physical (base pairs) or genetic (centiMorgans) position. The remaining columns correspond to individual genotypes.

Phenotype data

The phenotype file is organized with individuals in rows. The first column contains individual identifiers matching those in the genotype file, whereas subsequent columns contain phenotypic traits and optional fixed and random effects.

This standardized structure ensures compatibility across analytical modules and facilitates genomic analyses within the software.

AVAILABLE PROCEDURES

GenomicLand comprises a comprehensive set of analytical modules for data processing, GWS, and GWAS (Figure 1A). These modules are integrated within a unified analytical environment and can be accessed through the main navigation menu, enabling users to perform exploratory analyses, model fitting, prediction, and statistical inference within a structured workflow (Figure 1B).

Data preprocessing

The data preprocessing module provides tools for genotype data preparation, quality assessment, and exploratory analyses prior to genomic analyses (Figure 1C), including:

1. *Allelic dosage:* Converts nucleotide genotypes into numerical allele dosage values ranging from 0 to the ploidy level. The most frequent nucleotide is designated as the reference (major) allele, whereas the less frequent nucleotide is considered the alternative (minor) allele. This encoding follows standard practices in genomic prediction and facilitates downstream statistical modeling.

2. *Quality control:* Enables the removal of markers with low minor allele frequency (MAF) and low call rate and individuals with excessive proportions of missing data.

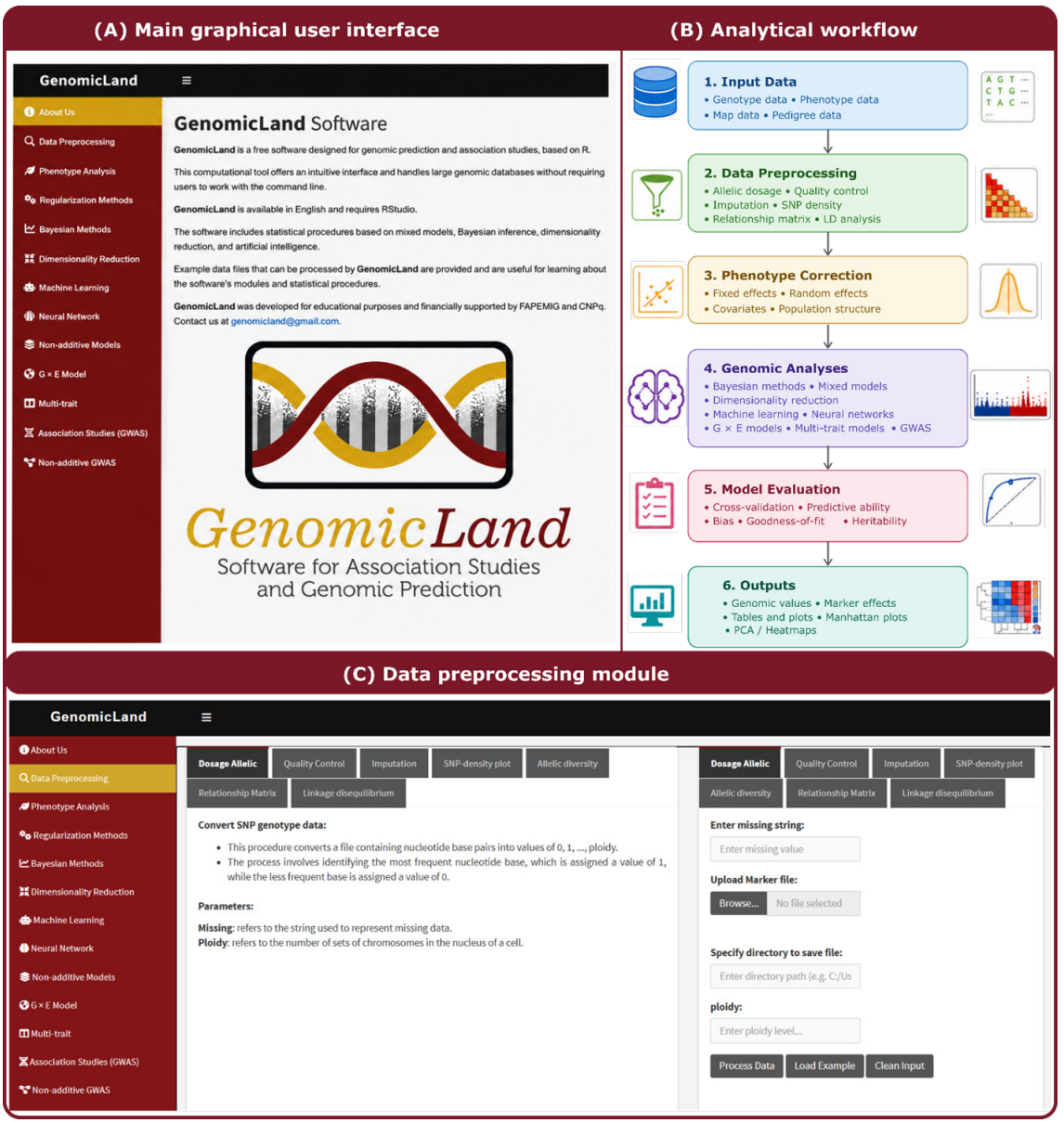

Figure 1. GenomicLand graphical interface, analytical workflow, and data preprocessing module. (A) Main graphical user interface showing the analytical modules available in the software. (B) Analytical workflow from data input and preprocessing to genomic analyses and output generation. (C) Data preprocessing module, including quality control, imputation, allelic diversity, relationship matrix construction, linkage disequilibrium analysis, and SNP density visualization.

3. *Imputation:* Missing allelic dosage values can be imputed using the marker incidence mean, mode, or a random imputation approach.

4. *SNP density plot:* Visualizes SNP distribution across the genome and highlights regions with varying marker density. Graphical outputs are generated using *CMplot* (LiLin-Yin 2024).

5. *Allelic diversity:* Estimates allele frequencies, polymorphic information content (PIC), observed heterozygosity, and expected heterozygosity.

6. *Relationship matrix (G):* Computes genomic relationship matrices for diploid and polyploid species. The module also generates heatmaps and principal component analysis (PCA) plots to support the evaluation of population structure and genomic relationships among individuals.

7. Linkage disequilibrium (LD): Estimates LD decay using marker data and genetic or physical distances. LD is quantified by the squared correlation coefficient (r^2) between marker pairs and summarized through LD decay curves.

Interface and input configuration

As illustrated in Figure 1C, the data preprocessing module provides an integrated interface for defining input parameters and executing preprocessing procedures. Required inputs include the genotype file, missing-data coding scheme, ploidy level, output directory, and filtering thresholds.

Three standard actions are available throughout the software: *Process Data*, *Load Example*, and *Clean Input*. Users can execute analyses using their own datasets or explore the software functionality through internal example datasets. Each module includes specific input parameters and generates tailored outputs according to the selected procedure. These procedures contribute to data quality and reliability, supporting accurate and robust downstream genomic analyses.

Phenotype analysis

Phenotype correction: Phenotypic data can be adjusted for fixed and random effects, including categorical factors, numeric covariates, and population structure, via principal components (Azevedo et al. 2017). Depending on the analytical design, adjustments can be performed using linear regression or mixed models to obtain genotype-level best linear unbiased estimators (BLUEs; Holland and Piepho 2024) for downstream genomic analyses (Figure 2). All analyses are implemented using the R package *sommer* (Covarrubias-Pazaran 2016).

Genomic prediction methods

The genomic prediction modules provide outputs including genomic estimated values, heritability estimates, marker effects, goodness-of-fit statistics, and predictive ability and bias assessed through K-fold cross-validation.

Regularization and mixed models

This module implements genomic prediction methods based on regression coefficient shrinkage and linear mixed models (Figure 2). Available methods include least absolute shrinkage and selection operator (LASSO), pedigree-based best linear unbiased prediction (A-BLUP), genomic BLUP (G-BLUP), single-step BLUP (H-BLUP), and heterogeneous G-BLUP.

1. *Least absolute shrinkage and selection operator (LASSO):* LASSO performs simultaneous variable selection and regularization through a shrinkage parameter (λ), optimized by K-fold cross-validation, producing sparse models by shrinking some coefficients to zero (Tibshirani 1996).

2. *Mixed model methods:* This module implements A-BLUP, G-BLUP, and H-BLUP based on pedigree (A), genomic (G), or combined (H) relationship matrices (Henderson 1975, VanRaden 2008, Legarra et al. 2014). These methods fit linear mixed models that accommodate fixed and random effects, including categorical factors and continuous covariates. Variance components are estimated by restricted maximum likelihood, and breeding values are obtained through BLUP. The module supports heterogeneous G-BLUP, which incorporates marker-specific weights when constructing the G matrix (Su et al. 2014). All models are fit using the R package *sommer* (Covarrubias-Pazaran 2016), and relationship matrices are computed using *AGHmatrix* (Amadeu et al. 2023).

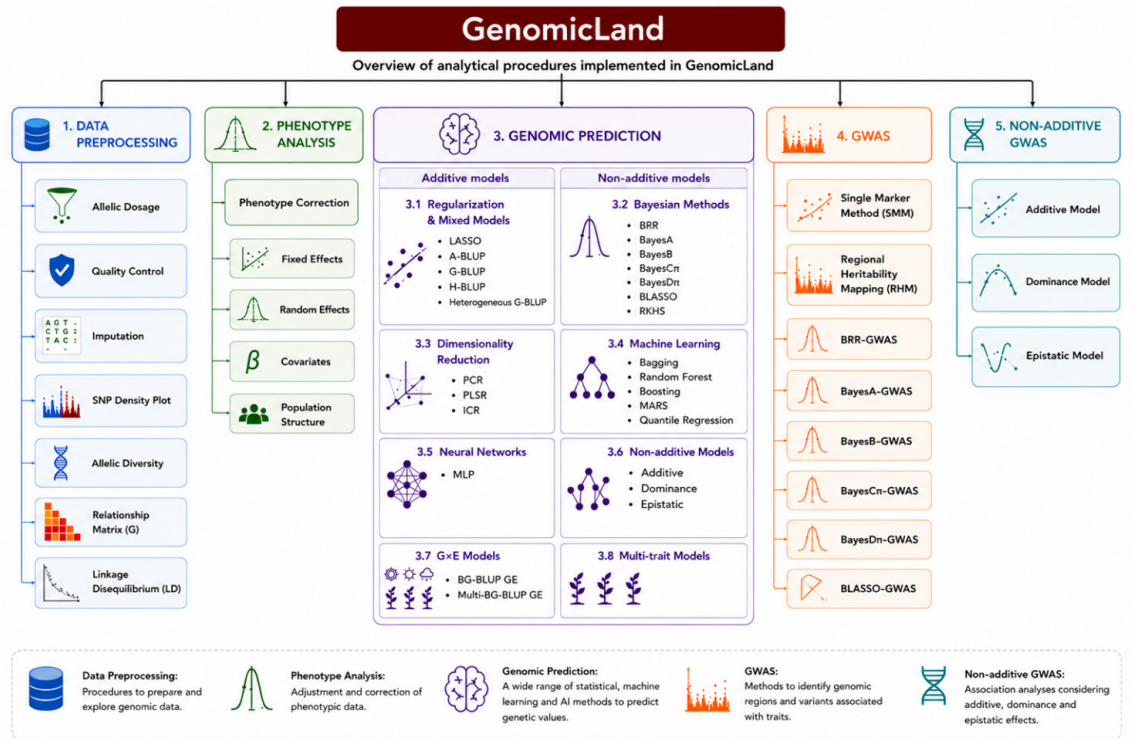

Figure 2. Overview of the analytical procedures implemented in GenomicLand. The figure summarizes the main analytical modules available in GenomicLand for genomic prediction and association studies. The software is organized into five major components: (1) Data preprocessing, including allelic dosage conversion, quality control, missing data imputation, SNP density visualization, allelic diversity analysis, genomic relationship matrix construction, and linkage disequilibrium (LD) analysis; (2) Phenotype analysis, including phenotype correction through fixed and random effects, covariates, and population structure adjustment; (3) Genomic prediction, encompassing regularization and mixed models, Bayesian methods, dimensionality reduction approaches, machine learning algorithms, neural networks, non-additive models, genotype-by-environment (G×E) models, and multi-trait analyses; (4) Genome-wide association studies (GWAS), including single-marker, regional heritability mapping, and Bayesian GWAS approaches; and (5) Non-additive GWAS, comprising additive, dominance, and epistatic association models.

Bayesian methods

This module implements marker-based Bayesian genomic prediction models that differ in their prior assumptions regarding marker effects and kernel-based approaches for estimating genomic values (Pérez and de los Campos 2014).

1. *Bayesian ridge regression (BRR):* Assumes homogeneous shrinkage, such that all marker effects are shrunk toward zero equally.

2. *BayesA:* Assumes heterogeneous shrinkage, allowing each marker to have its own variance and effect-specific shrinkage.

3. *BayesB:* Assumes that a user-defined proportion (π) of markers has no effect, whereas the remaining markers are modeled using heterogeneous shrinkage.

4. *BayesCπ:* Combines a point mass at zero with homogeneous shrinkage, while estimating the proportion of markers with non-zero effects from the data.

5. *BayesDπ:* Extends BayesCπ by assuming heterogeneous shrinkage for markers with non-zero effects.

6. *Bayesian least absolute shrinkage and selection operator (BLASSO):* Applies enhanced heterogeneous shrinkage to marker effects, favoring sparse solutions.

7. Reproducing kernel Hilbert space (RKHS): A non-parametric Bayesian approach that uses kernel matrices to model complex and potentially non-linear relationships between markers and phenotypes, allowing the capture of epistatic and other non-additive effects.

All Bayesian models are fit using the R package *BGLR* (Pérez and de los Campos 2014).

Dimensionality reduction methods

These methods transform high-dimensional predictors (e.g., marker incidence matrices) into a reduced set of latent components, improving computational efficiency and reducing collinearity (Azevedo et al. 2013). The optimal number of components is determined by K-fold cross-validation.

1. *Principal component regression (PCR):* Constructs uncorrelated components that maximize the variance explained by the predictor variables.

2. *Partial least squares regression (PLSR):* Constructs components that maximize the covariance between predictors and the response variable.

3. *Independent component regression (ICR):* Constructs statistically independent components to extract informative latent signals for prediction.

PCR and PLSR models are fit using the R package *pls* (Liland et al. 2024). ICR models are fit using functions from the R package *caret* (Kuhn 2008), combined with a custom ICR implementation based on Azevedo et al. (2013).

Machine learning methods

This module implements machine learning approaches for genomic prediction that can capture complex and potentially non-linear relationships between genetic markers and phenotypes (González-Camacho et al. 2018, Sousa et al. 2020, Oliveira et al. 2021, Oliveira et al. 2024).

1. *Bootstrap aggregating (Bagging):* Ensemble method that combines predictions from multiple regression trees fitted to bootstrap samples of the training data. Implemented using the R package ipred (Peters and Hothorn 2024).

2. *Random Forest:* Extension of Bagging that incorporates random subsets of predictors at each tree split to improve predictive performance. Implemented using the R package randomForest (Liaw and Wiener 2002).

4. *Boosting:* Sequential ensemble learning method in which trees are fitted iteratively to reduce prediction errors from previous iterations. Implemented using the R package gbm (Ridgeway and Developers 2026).

5. *Multivariate adaptive regression splines (MARS):* Non-parametric regression approach that models non-linear effects and interactions through piecewise linear splines. Implemented using the R package earth (Milborrow et al. 2024).

6. *Quantile regression:* Estimates predictor effects at user-defined quantiles of the response distribution, providing robustness to outliers and heterogeneous variances. Implemented using the R package quantreg (Koenker 2025).

Neural network methods

This module implements neural network models for genomic prediction that are particularly suitable for capturing complex non-linear and interaction effects among genetic markers (Cruz and Nascimento 2018, Costa et al. 2022).

Multilayer perceptron (MLP): (MLP): Feedforward neural network with one or more hidden layers for modeling complex non-linear relationships between predictors and response variables. Users can define the number of hidden layers, neurons, and activation functions. Models are implemented using neuralnet (Fritsch et al. 2019), with variable importance assessed using NeuralNetTools (Beck 2018) and model evaluation performed through K-fold cross-validation using caret (Kuhn 2008).

Non-additive models

This module extends the G-BLUP framework to include dominance and epistatic interactions, including additive × additive, additive × dominance, and dominance × dominance effects (Wellmann and Bennewitz 2011, Vitezica et al.

2013). By accounting for both additive and non-additive sources of genetic variation, these models provide a more comprehensive representation of the genetic architecture of complex traits. The significance of non-additive effects is evaluated using likelihood ratio tests. A Bayesian version of G-BLUP is also available with the same interaction structures, enabling model comparison based on the deviance information criterion (Wang et al. 2004, Wellmann and Bennewitz 2012, Azevedo et al. 2015). The module additionally includes marker-based Bayesian models for the joint modeling of additive and dominance effects.

Genotype-by-environment (G×E) models

This module evaluates genotype performance across multiple environments by explicitly modeling genotype × environment interactions, enabling the identification of genotypes with stable performance or specific adaptability (Resende et al. 2014).

1. *BG-BLUP GE:* A Bayesian extension of the G-BLUP framework that incorporates environmental effects, genomic relationships, and G×E interactions within a linear mixed-model framework. The significance of G×E effects can be evaluated using likelihood ratio tests.

2. *Multi-BG-BLUP GE:* A multivariate Bayesian model for multi-environment data that jointly analyzes observations across environments, enabling the estimation of genetic correlations and potentially improving predictive performance by borrowing information across environments.

Multi-trait models

This module implements multivariate Bayesian linear mixed models for the joint analysis of multiple traits. The models accommodate fixed and random effects while estimating genetic correlations among traits. By leveraging information shared across correlated traits, these approaches can improve prediction accuracy and the estimation of genetic parameters.

Genome-wide association methods

This module implements GWAS methods based primarily on additive genetic effects, supporting analyses at both the individual-marker and genomic-region levels. These approaches enable the identification of genomic regions associated with phenotypic variation while accounting for population structure and genetic relatedness.

1. *Single marker method (SMM):* Individual SNPs are tested sequentially while controlling population structure through PCA and background polygenic effects using the genomic relationship matrix (G). The method adopts a leave-one-chromosome-out (LOCO) strategy, in which chromosome-specific genomic relationship matrices are constructed excluding the chromosome under evaluation, reducing confounding between marker and polygenic effects. Analyses are implemented using the R package GWASpoly (Rosyara et al. 2016).

2. *Regional heritability mapping (RHM):* Estimates the heritability of predefined genomic regions, capturing the cumulative additive effects of markers within each segment (Suela et al. 2022). Population structure is accounted for through polygenic effects modeled using the genomic relationship matrix (G), and region size is defined by the user.

3. *Alphabetic Bayesian methods:* Bayesian GWAS analyses are performed using genomic prediction models under different prior assumptions. Genomic regions are identified through the window posterior probability of association (WPPA), which quantifies the proportion of genetic variance explained by markers within a genomic window (Fernando et al. 2017, Lima et al. 2022). Significance thresholds are user-defined.

Non-additive GWAS

This module extends the GWAS approaches described above by incorporating dominance effects, enabling the partitioning of genetic variance into additive and dominance components (Rosyara et al. 2016, Azevedo et al. 2022). By explicitly modeling non-additive genetic effects, these approaches may improve the detection of marker–trait associations when dominance contributes substantially to phenotypic variation.

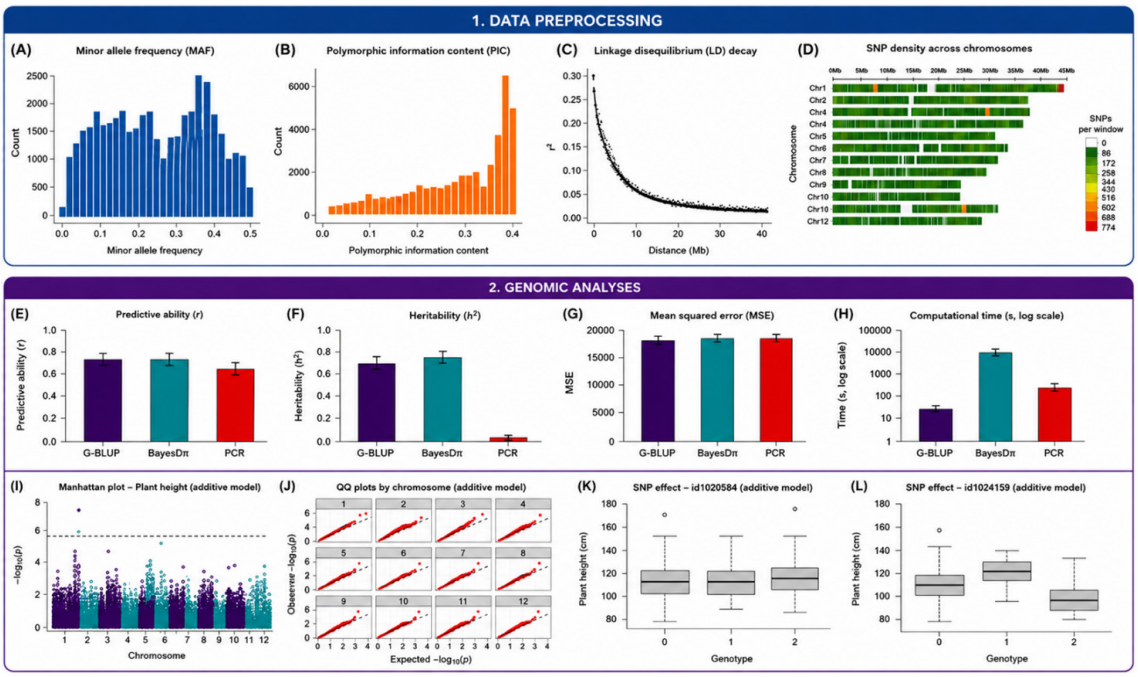

Figure 3. Representative outputs generated by GenomicLand using the Oryza sativa dataset. The figure summarizes the main analytical workflow available in GenomicLand, from data preprocessing to genomic prediction and genome-wide association analyses. (A) Distribution of minor allele frequency (MAF) across SNP markers. (B) Distribution of polymorphic information content (PIC). (C) Linkage disequilibrium (LD) decay represented by the decline of the squared correlation coefficient (r^2) as a function of physical distance (Mb). (D) SNP density across chromosomes. (E) Predictive ability, (F) heritability (h^2), (G) mean squared error (MSE), and (H) computational time obtained using different genomic prediction methods. (I) Manhattan plot of GWAS results for plant height under the additive model. (J) Chromosome-specific quantile–quantile (QQ) plots comparing observed and expected $-\log_{10}(p)$ values. (K–L) Phenotypic distributions according to genotype classes for representative SNPs associated with plant height.

Application example

To demonstrate the functionality of *GenomicLand*, a publicly available dataset of *Oryza sativa* L. (Asian rice) was analyzed. The dataset comprised 413 individuals and 36,901 SNP markers generated through the OryzaSNP and OMAP projects (Ammiraju et al. 2006, Zhao et al. 2011).

Plant height was selected to illustrate the analytical workflow. Representative outputs included marker quality and genomic structure summaries (Figure 3A-D), genomic prediction results (Figure 3E-H), and genome-wide association analyses (Figure 3I-L). Marker quality assessment revealed adequate allele frequency distributions, linkage disequilibrium patterns, and genome-wide marker coverage. Genomic prediction analyses enabled the comparison of alternative methods in terms of predictive ability, heritability, prediction error, and computational time. GWAS analyses identified putative genomic regions associated with plant height and provided graphical outputs, including Manhattan plots, QQ plots, and genotype-specific phenotype distributions.

GenomicLand was developed primarily for educational and research purposes. The analyses presented here were performed on a workstation equipped with an Intel Core i7-13700 processor and 64 GB of RAM. Computational requirements depend on dataset size and analytical complexity, particularly for Bayesian and machine learning methods. Future developments will focus on improving computational efficiency and expanding analytical capabilities for genomic datasets.

ACKNOWLEDGMENTS

The authors acknowledge the financial support provided by the Foundation for Research Support of the State of Minas Gerais (FAPEMIG) through projects APQ-02111-18 and APQ-01545-24 and by the Brazilian National Council for

Scientific and Technological Development (CNPq) through projects 309856/2023-0, 310755/2023-9, and 408833/2023-8. The authors also acknowledge, *in memoriam*, Dr. Fabyano Fonseca e Silva for his important contributions during the early development of this project, his role in shaping the software, and for naming it *GenomicLand*.

CREDIT STATEMENT

CFA; MN; VCF: Conceptualization; CFA; MN; MDVR; CDC: Methodology; CFA; MN; VCF: Software; CFA; MN: Formal analysis; CFA; MN; VCF: Visualization; CFA; MN: Investigation; CFA: Project administration; MDVR; CDC: Supervision; CFA; MN; VCF; MDVR; CDC: Validation; CFA: Funding acquisition; CFA; MN; VCF; MDVR; CDC: Resources; CFA; MN: Writing – original draft; CFA; MN; VCF; MDVR; CDC: Writing – review & editing.

REFERENCES

Amadeu RR, Garcia AAF, Munoz PR and Ferrão LFV (2023) AGHmatrix: genetic relationship matrices in R. **Bioinformatics 39**: btad445.

Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert G, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, SanMiguel P, Soderlund C, Jackson S and Wing RA (2006) 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 16**: 140-147.

Azevedo CF, Lima LP, Nascimento M and Nascimento ACC (2022) Bayesian methods for genomic association of chromosomic regions considering the additive–dominance model. **Crop Breeding and Applied Biotechnology 22**: e418422310.

Azevedo CF, Nascimento M, Fontes VCF, Silva FF, Resende MDV and Cruz CD (2019) GenomicLand: Software for genome-wide association studies and genomic prediction. **Acta Scientiarum. Agronomy 41**: e45361.

Azevedo CF, Resende MDV, Silva FF, Lopes OS and Guimarães SEF (2013) Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. **Pesquisa Agropecuária Brasileira 48**: 619-626.

Azevedo CF, Resende MDV, Silva FF, Nascimento M, Viana JMS and Valente MSF (2017) Population structure correction for genomic selection through eigenvector covariates. **Crop Breeding and Applied Biotechnology 17**: 350-358.

Azevedo CF, Resende MDV, Silva FF, Viana JMS, Valente MSF, Resende Junior MFR and Muñoz P (2015) Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genetics 16**: 1-13.

Beck MW (2018) NeuralNetTools: Visualization and analysis tools for neural networks. **Journal of Statistical Software 85**: 1-20.

Costa WG, Celeri MO, Barbosa IP, Silva GN, Azevedo CF, Borem A, Nascimento M and Cruz CD (2022) Genomic prediction through machine learning and neural networks for traits with epistasis. **Computational and Structural Biotechnology Journal 20**: 5490-5499.

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package *sommer*. **PLoS ONE 11**: e0156744.

Cruz CD and Nascimento M (2018) **Inteligência computacional aplicada ao melhoramento genético**. Editora UFV, Viçosa, 414p.

Fernando R, Toosi A, Wolc A, Garrick D and Dekkers J (2017) Application of whole-genome prediction methods for genome-wide association studies: a Bayesian approach. **Journal of Agricultural, Biological and Environmental Statistics 22**: 172-193.

Fritsch S, Guenther F and Wright M (2019) neuralnet: Training of neural networks. R package version 1.44.2.

González-Camacho JM, Ornella L, Pérez-Rodríguez P, Gianola D, Dreisigacker S and Crossa J (2018) Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. **The Plant Genome 11**: 1-15.

Henderson CR (1975) Best linear estimation and prediction under a selection model. **Biometrics 31**: 423-447.

Holland JB and Piepho HP (2024) Don't BLUP twice. **G3 Genes | Genomes | Genetics 14**: jkae250.

Koenker R (2025) quantreg: Quantile regression. R package version 6.1.

Kuhn M (2008) Building predictive models in R using the caret package. **Journal of Statistical Software 28**: 1-26.

Legarra A, Christensen OF, Aguilar I and Misztal I (2014) Single Step, a general approach for genomic selection. **Livestock Science 166**: 54-65.

Liaw A and Wiener M (2002) Classification and regression by randomForest. **R News 2**: 18-22.

Liland K, Mevik B and Wehrens R (2024) pls: Partial least squares and principal component regression. R package version 2.8-5.

LiLin-Yin (2024) CMplot: Circle manhattan plot. R package version 4.5.1.

Lima LP, Azevedo CF, Resende MDV, Nascimento M and Silva FF (2022) Evaluation of Bayesian methods of genomic association via chromosomic regions using simulated data. **Scientia Agricola 79**: 3.

Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics 157**: 1819-1829.

Milborrow S, Hastie T and Tibshirani R (2024) earth: Multivariate adaptive regression splines. R package version 5.3.4.

Oliveira CM, Costa WG, Nascimento ACC, Azevedo CF, Cruz CD, Sagae VS

and Nascimento M (2024) Multivariate adaptive regression splines enhance genomic prediction of non-additive traits. **Agronomy 14**: 2234.

Oliveira GF, Nascimento ACC, Nascimento M, Sant'Anna IC, Romero JV, Azevedo CF, Bhering LL and Caixeta ET (2021) Quantile regression in genomic selection for oligogenic traits in autogamous plants: A simulation study. **PLoS ONE 16**: e0243666.

Pérez P and de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. **Genetics 198**: 483-95.

Peters A and Hothorn T (2024) ipred: Improved predictors_. R package version 0.9-15.

Resende MDV, Resende Junior MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria AD, Pappas Junior GJ, Kilian A and Grattapaglia D (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist 194**: 116-128.

Resende MDV, Silva FF and Azevedo CF (2014) **Estatística matemática, biométrica e computacional**. Editora Suprema, Visconde do Rio Branco, 881 p.

Ridgeway G and Developers G (2026) gbm: Generalized boosted regression models. R package version 2.2.3.

Rosyara UR, Jong WS, Douches DS and Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. **Plant Genome 9**: 1-10.

Sousa IC, Nascimento M, Silva GN, Nascimento ACC, Cruz CD, Silva FF, Almeida DP, Pestana KN, Azevedo CF, Zambolim L and Caixeta ET (2020) Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. **Scientia Agricola 78**: e20200021.

Su G, Christensen OF, Janss L and Lund MS (2014) Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. **Journal of Dairy Science 97**: 6547-6559.

Suela MM, Azevedo CF, Nascimento M, Nascimento ACC and Resende MDV (2022) Regional heritability mapping and genome-wide association identify loci for rice traits. **Crop Science 62**: 839-858.

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistical Society, Series B (Methodological) 58**: 267-288.

Uffelmann E, Huang QQ, Munung NS, Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T and Posthuma D (2021) Genome-wide association studies. **Nature Reviews Methods Primers 1**: 59.

Vandenplas J, Calus MPL and Gorjanc G (2018) Genomic Prediction using individual-level data and summary statistics from multiple populations. **Genetics 210**: 53-69.

VanRaden PM (2008) Efficient methods to compute genomic predictions. **Journal of Dairy Science 91**: 4414-1423.

Vitezica ZG, Varona L and Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics 195**: 1223-1230.

Wang J, van Ginkel M, Trethowan R, Ye G, DeLacy IH, Podlich DW and Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. **Crop Science 44**: 2006-2018.

Wellmann R and Bennewitz J (2011) The contribution of dominance to the understanding of quantitative genetic variation. **Genetics Research 93**: 139-154.

Wellmann R and Bennewitz J (2012) Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research 94**: 21-37.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton JG, Islam AR, Reynolds A, Mezey J, Mcclung AM, Bustamante CD and Mcclung AM (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. **Nature communications 2**: 467.