

Genome-enabled Breeding Strategies for Nitrogen Use Efficiency in Sweet Corn

Marco Antonio Peixoto

Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA

Christina Finegan

Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA; and Bayes Crop Science, Chesterfield, MO, USA

Queen Offord and Kristen Leach

Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA

Ying Hu

Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA; and Jilin Agricultural University, Changchun, China

William Tracy

Department of Plant and Agroecosystem Sciences, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

Marcio Resende Jr

Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA

Keywords. artificial intelligence, breeding program simulations, genomic selection, genome-wide association, indirect selection

Abstract. Sweet corn is one of the most widely cultivated vegetable crops in the United States. Nitrogen plays a critical role in promoting plant growth and development and is essential for maximizing grain yield and accelerating genetic progress. In this study, we aimed to evaluate the integration of modern breeding tools into the sweet corn breeding program to improve nitrogen use efficiency (NUE) traits. A total of 693 inbred lines and 108 hybrids were assessed, with NUE traits measured across three developmental stages (leaf nitrogen at R1, R3, and R6). All inbred lines were whole-genome resequenced. We performed a genome-wide association study (GWAS) and implemented a set of genomic models to predict inbred and hybrid performance. We simulated two traits reflecting the genetic architecture of NUE traits in a sweet corn pipeline using coalescent theory to compare genomic selection strategies against a benchmark phenotypic selection scenario. Our results revealed significant genetic variation among inbred lines and hybrids for most NUE traits, indicating a complex genetic architecture. The GWAS identified candidate genes potentially associated with NUE. Among the traits evaluated in inbred lines, R1.LN showed the highest prediction accuracies (0.36–0.38), followed by R3.LN (0.29–0.33) and R6.LN (0.03–0.06). However, genomic prediction accuracy in the hybrid population was lower (R3.LN: 0.14). Notably, using molecular markers to optimize crosses emerged as the most effective strategy for the long-term improvement of NUE traits. In summary, genomic tools can enhance NUE traits, thus highlighting their potential to improve NUE not only in sweet corn but also in other vegetables and move toward more sustainable production.

Sweet corn (*Zea mays* L.) is an important vegetable crop worldwide. In the United States, it has consistently ranked among the top vegetables in both harvested area and total production in recent years (US Department of Agriculture 2025). To sustain high productivity, farmers commonly rely on synthetic nitrogen fertilizers. However, nitrogen fertilizers are a significant environmental pollutant and require substantial energy to produce (Wani et al. 2021; Yu et al. 2022).

Breeding for nitrogen use efficiency (NUE) traits is important in scenarios in which breeders seek practices that promote more sustainable agriculture while reducing environmental impact. The NUE is defined as the ratio of grain yield to available nitrogen (Moose and Below 2009). In other words, high values of NUE represent plants that are more efficient in using nitrogen (uptake and utilization), which could reduce the amount of synthetic nitrogen needed. Despite its

importance, breeding for such traits in sweet corn has experienced slow progress. As is known, increases in field corn yield have been associated with increases in NUE (Govindasamy et al. 2023). However, NUE is not a trait that is directly selected for in sweet corn breeding programs. Future research is needed to understand the impact of explicitly selecting for NUE in a breeding program; however, increasing our understanding of the genetic control of this trait is a first step toward this goal.

The NUE can be broken down into two components: uptake efficiency, which refers to the amount of available nitrogen that the plant can absorb, and utilization efficiency, which refers to how efficiently the plant converts the additional nitrogen taken up into yield (Gheith et al. 2022). On average, plants take up two-thirds of the total nitrogen by the time they shift from vegetative to reproductive growth (stage R1). Thereafter, some uptake continues (postsilking uptake) while remobilization begins and nitrogen is reallocated from the shoots to reproductive regions (Coque and Gallais 2007). Selecting for NUE (and its components) is associated with few challenges. Phenotyping for NUE is labor-intensive and time-consuming, and it often requires destructive sampling (Sanchez et al. 2023). Moreover, NUE is a complex trait governed by multiple physiological processes (Gheith et al. 2022). In sweet corn populations, the genetic architecture underlying this trait remains poorly understood.

Because of the importance and complexity of breeding for NUE traits in sweet corn, molecular markers represent a tool to enhance breeding efforts. For instance, markers can be used to identify genomic regions associated with NUE traits through a genome-wide association study (GWAS) (Sanchez et al. 2023). In the GWAS, a panel of markers is used in a regression-like analysis with the phenotypic response (the trait of interest) to determine whether changes in the markers are associated with changes in the phenotype. In addition, they can be incorporated in genomic selection (GS) models, where historical phenotypes together with molecular markers in the same individuals are used in a model to predict the performance of individuals in a new set of only genotyped individuals. In the context of a breeding program, GS models can be used to increase selective accuracy (i.e., enhance the certainty of which individuals to select or to discard), reduce phenotyping costs, shorten the breeding cycle, and guide crosses (Crossa et al. 2021; Marinho et al. 2022).

Our objectives were the following: provide insights into the genetic architecture (heritabilities, variance components, and correlations) of NUE traits in a sweet corn population; identify genomic regions associated with NUE traits and potential candidate genes for those traits; apply genomic models to predict the performance of inbreds and hybrids; and evaluate the long-term efficiency of genomic selection in multitrait framework through stochastic simulations to improve NUE traits.

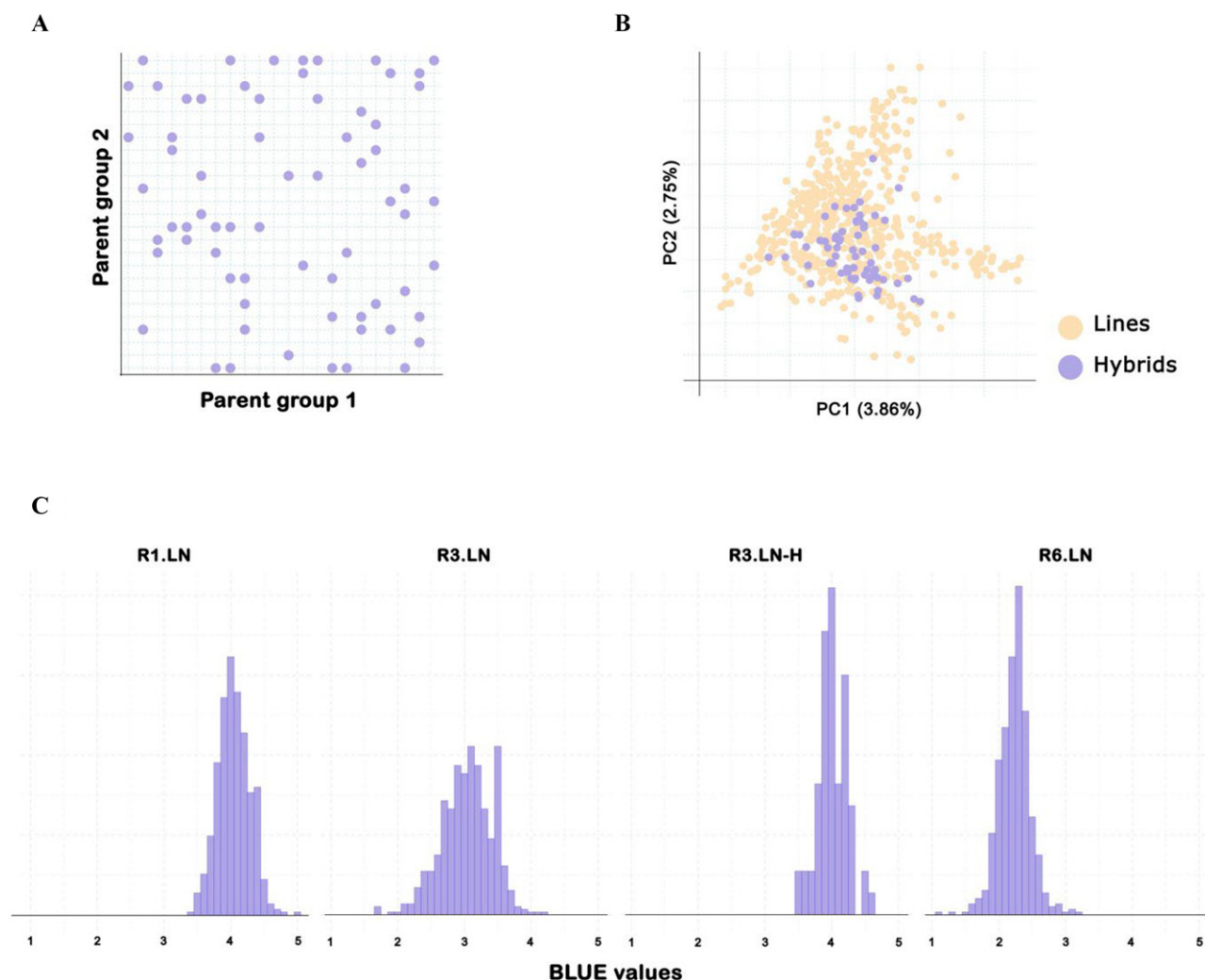


Fig. 1. (A) The mating plan used to create the hybrids. Each dotted line represents a parental line, and each dot represents a single hybrid. (B) Genetic diversity within the individuals were analyzed by conducting a principal component analysis based on the markers for the lines and hybrids. (C) Distribution of nitrogen use efficiency traits. The blue values represent the best linear unbiased estimation for each trait. R1.LN, R3.LN, and R6.LN represent leaf nitrogen at stages R1, R3, and R6, respectively. R3.LN-H represents the trait measured in the hybrid population.

Materials and Methods

We used a diverse sweet corn population genotyped with whole-genome resequencing and phenotypes of key NUE traits from the lines and hybrids derived from the population to build a GWAS model for candidate gene identification and a genomic model (GS) to predict the performance of the individuals. In addition, simulations were conducted to infer the long-term effects of GS to improve NUE traits in a sweet corn program.

Plant material and genotyping

A diversity panel of 693 sweet corn inbred lines containing tropical and temperate-adapted materials was tested in this study. A sweet corn population assembled by Baseggio et al. (2021) was expanded using germplasm acquired from the US Department of Agriculture Germplasm Resources Information Network (GRIN) system, the University of Wisconsin-Madison Sweet Corn Breeding Program, and the University of Florida Sweet Corn Breeding Program. A subset of 49 inbred lines from the panel were crossed to create 108 unique F_1 hybrids (Fig. 1) that were later used for the field evaluation. To generate the hybrids, we used an incomplete North Carolina II crossing scheme. This was a collaborative effort involving 24 elite inbred lines from the University of Florida Sweet Corn Breeding Program and 25 inbred lines from the University of Wisconsin-Madison Sweet Corn Breeding Program.

For genotyping, DNA was extracted and sequenced with NovaSeq 2 $\times$ 150 bp reads at a shallow depth (mean depth = 8.45 $\times$) to create a high-density panel of single-nucleotide polymorphism (SNP) markers distributed

throughout the genome. Reads were aligned to the *Ia453-sh2* reference genome (Hu et al. 2021). Variants (including SNPs and small insertion-deletion polymorphisms) were called using both GATK (Van der Auwera and O'Connor 2020) and Freebayes (Garrison and Marth 2012). The overlapping variants resulting from these two methods (47,160,177 variants) were selected and further filtered using GATK best practices (DePristo et al. 2011) and by missingness (>30%) and minor allele frequency (>1%), resulting in a final panel of 28,498,353 million variants (Colantonio et al. 2022).

Experimental design and traits measured

Inbred lines were grown at the University of Florida Plant Science Research and Education Unit in Citra, FL, USA, in 2019, while the hybrid population was generated from inbreds in 2019 and grown in Citra, FL, USA, in 2021. A resolvable incomplete block design with two replications and nine incomplete blocks was used for the inbred line evaluation, whereas a resolvable incomplete block design with two replicates and four incomplete blocks was used

Received for publication 12 Sep 2025. Accepted for publication 11 Oct 2025.

Published online 11 Nov 2025.

This work was supported by the National Institute of Food and Agriculture SCRI 2018-51181-28419, AFRI 2019-05410, and USDA-NIFA 2022-51181-38333.

M.R. is the corresponding author. E-mail: mresende@ufl.edu.

This is an open access article distributed under the CC BY-NC license (https://creativecommons.org/licenses/by-nc/4.0/).

for hybrids. Genotypes were planted in single-row plots with 12 plants per row. Plants were spaced at 7 inches apart, and rows were 36 inches apart. For each trial, seeds were treated with fungicide, and plots were irrigated and fertilized with nitrogen at 220 kg/ha.

The leaf nitrogen content was determined by a combustion analysis of pooled leaf samples collected from the leaf immediately above the uppermost ear on three representative plants per row. Sampling occurred at three key developmental stages: silking (R1, referred to as the R1.LN trait), 21 d after pollination or milk stage (R3, referred to as the R3.LN trait), and physiological maturity (R6, referred to as the R6.LN trait). This approach allowed us to track nitrogen content through plant development, including remobilization and contribution to yield. A combustion analysis was conducted by Waters Agricultural Laboratories, Inc. (Camilla, GA, USA). In the hybrid population, because of the high correlation between R1.LN and R3.LN, leaf nitrogen data were collected only at the R3 stage.

Genetic analysis

A two-stage analysis was implemented to evaluate the traits (Holland and Piepho 2024). In the first stage, the restricted maximum likelihood method (REML) and best linear unbiased prediction (BLUP) procedure (Henderson 1974; Patterson and Thompson 1971) were used along with the following linear mixed model:

$$y = Xb + Zg + Wp + e \quad [1]$$

where y is the vector of phenotypic data, b is the fixed effect of repetition and checks inside blocks (assumed as fixed) summed with the overall mean, g is the vector of genotypes (assumed as random), with $g \sim N(0, \sigma_g^2)$, where σ_g^2 is the genotypic variance, p is the vector of blocks (assumed as random), with $p \sim N(0, \sigma_p^2)$, where σ_p^2 is the block variance and e is the residual effect (random), with $e \sim N(0, \sigma_{res}^2)$, where σ_{res}^2 is the residual variance. Additionally, X , Z , and W were the incidence matrices for b , g , and p , respectively. Broad-sense heritability was calculated as the ratio of genotypic variance to phenotypic variance. For the second stage of the analysis, we assumed that the genotypic effect was fixed in the same model described, and best linear unbiased estimates (BLUEs) were estimated.

The variances of the random effects were tested using the likelihood ratio test (Rao 1973), as follows:

$$LRT = -2(\log L_r - \log L), \quad [2]$$

where $\log L_r$ and $\log L$ are the logarithm of the maximum residual likelihood function of the reduced and full model, respectively. For the likelihood ratio test, χ^2 statistics with 1 degree of freedom (df) and 5% probability of error type I were considered.

Accounting for spatial trends in the phenotypic data

Modeling of nongenetic effects was implemented to account for heterogeneity within the trial. The method allows for spatial autocorrelations among observations in a row-column fashion (Gilmour et al. 1997; Stefanova et al. 2009; Werner et al. 2024). Five models were implemented to expand the linear model in equation 1. The first model (model 1) did not consider any spatial pattern and was implemented as a benchmark comparison model. Model 1 was implemented by assuming spatially independent observations ($R = \sigma_{res}^2 I_r \otimes I_c$), where I_r is the identity matrix for rows ($r \times r$) and I_c represents the identity matrix for columns ($c \times c$).

Moving forward, we modeled the spatial or correlated error. This error captures local and global trends, such as scale fertility gradients, small changes in soil composition, small-scale disease/insect damage at the plot level, and other factors (Coelho et al. 2021; Gilmour et al. 1997; Stefanova et al. 2009; Werner et al. 2024), that may affect the experiment under evaluation creating dependency among spatially close plots units. For such, we shaped spatial correlation among rows and columns by including the autoregressive process of order 1 (AR1 structure) (Box et al. 2015). The residual was modeled as $R = \sigma_{\xi}^2 \Sigma_r(\rho_r) \otimes I_c$ and $R = \sigma_{\xi}^2 I_r \otimes \Sigma_c(\rho_c)$, for model 2 and model 3, respectively, where σ_{ξ}^2 represents the spatial dependent or correlated error, ρ_r and ρ_c represent the correlations for rows and columns, respectively, and Σ_r and Σ_c represent the first-order autoregressive correlation matrix for columns and rows. Model 4 accounted for correlations among rows and columns simultaneously, with the residual structure denoted by $R = \sigma_{\xi}^2 \Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)$.

Finally, the inclusion of a random error, or nugget effect, which captures any additional variation that is not a trend, was used (Gilmour et al. 1997). In the model, this was achieved by including an independent or uncorrelated term and accounting for the variance of the same component (σ_{η}^2) implemented in model 4, which became model 5. Such a structure captures any noise, measurement error, or even intrinsic variability within the plots (random, independent, or nonspatially correlated) and could not be captured by the dependent structure (Gilmour et al. 1997; Wilkinson et al. 1983).

The best model for each trait was identified using the Bayesian Information Criterion (BIC) (Schwarz 1978). Models (from model 1 to model 5) with the lowest BIC for each trait were assumed as the best models. Only the experiment from the lines field was modeled. All analyses conducted during the second stage used the genetic values generated by such a model.

Genome-wide association analysis

We hypothesized that genetic variants could be involved in the expression of NUE traits. To test this hypothesis, we conducted a genome-wide association analysis using two distinct methods. Initially, SNPs with minor

allele frequency (<0.01) were filtered out, resulting in a set of 16,755,210 variants (hereafter referred to as the 16M set). The first method used the R package GAPIT (Wang and Zhang 2021), with the kinship matrix calculated using the default algorithm in the FarmCPU model. The analysis was conducted using a maxLoop threshold of 10 and a QTN threshold of 10. The second method used EMMAX (Kang et al. 2008) software to fit a standard univariate linear mixed model. For both methods, a Bonferroni-corrected significance threshold of 0.05 was applied to identify candidate SNPs, which represented a threshold of $-\log_{10}(p) = 8.32$.

The genome sequence of the maize line B73 (RefGen_v5) (https://www.maizedb.org/) was used as the reference genome for candidate gene searching. The confidence interval of significant SNPs was determined based on the LD of the population (Colantonio et al. 2022) at 1 MB upstream and downstream of the significant SNPs. The confidence interval of each significant SNP was used to determine the search scope of candidate genes. First, the SCI-SPACE AI tool (PubGenius Inc. 2025) was used to identify genes with known functions related to nitrogen metabolism by using the prompt “candidate genes involved in nitrogen metabolism, uptake, utilization, and mobilization in maize,” which returned a description of genes with known functions related to nitrogen metabolism alongside cited articles. Any identified genes found in the artificial intelligence (AI)-provided list were preferred as candidate genes. Then, studies of NUE-related traits using the GWAS (He et al. 2020), transcriptomics (Singh et al. 2023; Zhang et al. 2024), network and regulatory analysis (Plett et al. 2017), and combined omics (Gong et al. 2020; Liu et al. 2012) suggested by the AI tool were used to assist with screening candidate genes.

Genomic prediction

We hypothesized that molecular markers can be used to predict the performance of individuals for NUE traits. To test this hypothesis, we implemented genomic selection models and tested their performance in cross-validation schemes. Genomic selection models were implemented in single-trait and multi-trait approaches. First, a genomic BLUP (GBLUP) model was used (VanRaden 2008), which was represented as follows:

$$y = \mu + Za + e \quad [3]$$

where y is the matrix of BLUEs estimated in the first stage of the analysis, μ is the overall mean for each trait, Z is the incidence matrix connecting the observations with the response variable, a is the vector of additive genetic effects (assumed as random), where $a \sim N(0, G \otimes \Sigma_a)$, with G as the additive relationship matrix (VanRaden 2008) and Σ_a is the variance-covariance matrix for the genetic variance, and e is the residual effect (random), where $e \sim N(0, I \otimes \Sigma_{res})$, with Σ_{res} as the residual variance-covariance matrix of residual variances, and I is an identity matrix. Then, the

GBLUP was derived, and it was implemented in its univariate form.

The second set of models used included marker-based models, where marker effects were estimated rather than individual predictions, as in the GBLUP model. Implementation involved two special models: a univariate BayesB model (Meuwissen et al. 2001) and a similar implementation in a multitrait framework called the spike-slab model (Habier et al. 2011; Pérez-Rodríguez and de los Campos 2022). Then, the model was implemented as follows:

$$y = \mu + X\beta + e \quad [4]$$

where y is the matrix of BLUEs from the first-stage analysis, μ is a vector of trait-specific overall means, X is the matrix for the predictors (i.e., the marker matrix for each individual in y), β is a matrix with the effect of each SNP (markers by traits), where $\beta \sim N(0, \sigma_\beta^2)$, with σ_β^2 as the variance of the non-zero markers (π), which follow an inverse χ^2 distribution, and e as the residual (random), where $e \sim N(0, \sigma_{res}^2)$, with σ_{res}^2 is the residual variance.

In a BayesB model, a large proportion of the SNPs are assumed to have zero effects on the target trait ($1 - \pi$), while a small proportion (π) is considered with nonzero effects. The prior distribution of the effect of each SNP is a mixture of a scaled-t distribution with probability π and a distribution of point mass at zero with probability $1 - \pi$. In the spikes-slab model, $p(\pi^*_1, \pi^*_2, \dots, \pi^*_t) \propto \prod_{k=1}^t \pi^{\alpha_1}_{*k} (1 - \pi^*_{*k})^{\alpha_2}$, where α_1 and α_2 are the prior shape1 and shape2 parameters for the k^{th} trait. An unstructured variance-covariance matrix was chosen for the effect in this model (Ω), with the prior a $[P(\Omega)]$ assumed to be an inverse Wishart distribution.

Using BGLR software, the residual variance is assigned to a scaled inverted χ^2 prior, where $\sigma_e^2 \sim \chi^2(\sigma_e^2 | d_e, S_e)$, with d_e and S_e being the df and the scaling parameter, respectively. Additionally, for R_0 , an inverse Wishart prior was assumed. The overall mean (μ) is assigned to a flat prior.

A subset of 200,000 SNPs was randomly selected from the 16M set. The SNPs were filtered out by minor allele frequency (<0.05) and missingness (>0.80), yielding a total of 101,348 SNPs. The additive realized relationship matrix (VanRaden 2008) was generated with the aid of the R package AGHmatrix (Amadeu et al. 2023) and used for model prediction and cross optimization. The missing values of the markers were replaced by the mean of the markers, which is the default in AGHmatrix. For the hybrid combinations, we recreated the value of the markers of the hybrid by summing the values of the parent markers (Peixoto et al. 2024c). In addition, to summarize and visualize the genetic diversity of the populations (lines and hybrids), we plotted a principal component analysis using the *prcomp()* function from base R.

Cross-validation schemes were used to evaluate the predictive performance of genomic models for unobserved genotypes. Two

schemes were implemented. First, cross-validation scheme one (CV1) involved five-fold cross validation applied to the population of inbred lines and repeated 10 times to ensure robustness. The second scheme (CV0) was designed to reflect a more realistic breeding program scenario in which models are expected to predict the performance of future hybrids in untested conditions. Using this approach, all inbred lines, including the parents, were used to predict hybrid performance. The prediction accuracy of the genomic models under each scenario was assessed by calculating the correlation between the estimated breeding values and the BLUEs obtained in the first-stage analysis for each trait.

Cross prediction for NUE traits

We synthetically generated all possible pairwise crosses among the individuals in the line panel to develop a mating plan targeting the improvement of NUE traits. A total of 39,778 crosses were predicted from all combinations of the 546 phenotyped lines. Marker data and estimated SNP effects were used to predict the midparent value (MPV) for each F1 combination for both traits—R1.LN and R3.LN—using the *getMPV()* function from the *SimpleMating* package (Peixoto et al. 2025). Equal weights (1:1) were assigned to both traits to ensure balanced selection pressure.

Subsequently, the top 100 crosses were selected to form a mating plan using the *selectCrosses()* function under constraints that limited each parent to a maximum of two crosses and imposed a co-ancestry threshold of -0.02 to minimize relatedness. The SNP effects used for prediction were obtained from the BGLR package via the *Multitrait()* function by applying a spike-slab model (Pérez-Rodríguez and de los Campos 2022).

Breeding program simulations

We hypothesized that implementing a genomic selection model for NUE traits would be the most effective strategy for achieving long-term genetic gains in the sweet corn breeding program. Additionally, as previously described, the use of cross prediction and optimization could enhance the breeding program by improving the selection of superior crosses for NUE traits. To evaluate this hypothesis, we simulated the sweet corn breeding pipeline from the University of Florida (for further details, see Peixoto et al. 2024b). In the simulations, various scenarios were implemented, and genetic progress was monitored across selection cycles.

Briefly, the University of Florida sweet corn breeding program follows a doubled haploid pipeline with three rounds of hybrid evaluation in the target environment. The program operates across two environments: the target environment, where hybrid testing occurs over three consecutive testcross rounds, and an off-season nursery used for creating the crosses, hybrids, and DH lines. The program leverages two pseudo-heterotic groups: one composed of University of Florida-derived lines and the other consisting of proprietary material from a

commercial partner. Crosses between these groups are evaluated as hybrids in the target environment.

Stochastic simulations of the sweet corn breeding program. We used the AlphaSimR package (Gaynor et al. 2021) to simulate 20 years of the sweet corn breeding program at the University of Florida. The breeding pipeline, including population sizes and selection decisions, followed the doubled haploid scheme thoroughly described by Peixoto et al. (2024a). To initiate the simulation, a base genome of 100 individuals was generated using the Markovian Coalescent Simulator (Chen et al. 2009). The *Maize* option was specified in the species argument to ensure realistic haplotype structures and genome characteristics representative of maize.

Two traits were simulated to reflect the nitrogen content at silking (R1.LN) and 21 d after pollination (R3.LN) by incorporating additive, dominance, and genotype-by-environment ($G \times E$) effects. The traits were modeled with a genetic correlation of 0.48, mimicking the empirical relationship observed between R1.LN and R3.LN. The additive means and variances were set to 4.04/3.04 and 0.5/0.5, respectively, for R1.LN and R3.LN. Residual variances were set to 0.9 and 0.7, respectively. The $G \times E$ variance for each trait was set to three-times the residual variance. A dominance degree of 0.93 and dominance variance of 0.2 were used for both traits. For those parameters that could not be estimated in the base population ($G \times E$ variance and dominance degree), we used previous studies of sweet corn to incorporate them into the simulated pipeline (for more details, please see Peixoto et al. 2024a, 2024b).

Breeding scenarios. Three breeding scenarios were simulated and compared. All were designed to reflect feasible strategies for improving NUE within the current sweet corn breeding pipeline.

Phenotypic selection (conventional)

This baseline scenario represents the existing breeding approach, which relies on truncated phenotypic selection and doubled haploid production. Parents are selected based on phenotypic values and randomly mated to generate 50 crosses at the start of each cycle. This scenario includes a 15 year burn-in phase, after which all subsequent scenarios begin from the same genetic background and continue for an additional 20 years under the conventional (Conv) strategy.

Genomic selection

This scenario integrates genomic selection to replace phenotypic selection. Individuals are selected at each stage based on estimated breeding values using a truncated selection strategy. Genomic prediction is applied to identify superior DH lines for use as parents and to advance lines throughout the testcross stages (see Peixoto et al. 2024a for details). As in the Conv scenario, 50 crosses are

produced per cycle by randomly mating selected parents.

Optimal cross selection

Building on the GS framework, this scenario adds an optimized mating plan using genomic predictions. All DH lines are considered potential parents. Crosses are predicted using midparent values computed via the `getMPV()` function in the `SimpleMating` R package, followed by optimization with the `selectCrosses()` function. Constraints included a maximum of two crosses per parent and a minimum pairwise co-ancestry threshold of -0.02 to limit relatedness and manage inbreeding.

The simulations and scenarios described were designed to mimic the breeding process for the R1.LN and R3.LN traits in the sweet corn pipeline. These traits are not only costly to measure but also moderately correlated ($r = 0.48$). Because R1.LN data were not available for the hybrid population, we simulated all three scenarios using selection based solely on the R3.LN trait. This allowed us to evaluate the indirect response in R1.LN and assess the extent of genetic gain achievable by phenotyping only at a single developmental stage for NUE traits.

Breeding value estimation. The genomic model for the GS and optimal cross selection (OCS) scenarios was implemented in BGLR using the `BGLR()` function and the BayesB model to estimate marker effects. The training population was updated using a 4 year sliding window approach and the most recent 4 years of genotypic and phenotypic data. These individuals from the training population came from the parents' performance and were assessed by a general combining ability model.

Scenarios comparison. The parameters, additive population mean and variance, were tracked to compare the performance of the scenarios at each cycle. The measurement was conducted in the parental population in each cycle.

Software

The mixed models (REML/BLUP) for BLUEs estimation, heritability calculation, and spatial model correction were implemented in the R package ASREML (Butler et al. 2017). The genomic models were implemented in BGLR (Pérez-Rodríguez and de los Campos 2022, 2014), and the burn-in and number of iterations were equal to 1000 and 10,000, respectively (for breeding program simulations and CV1 and CV0 schemes). The datasets and scripts to reproduce all analyses can be found online (https://github.com/Resende-Lab/NUE-Traits_SweetCorn).

Results

Phenotypic and genotypic variability

The phenotypic distributions of the measured traits indicated that phenotypic ranges were consistently broader among inbred lines compared with hybrids (Fig. 1C). The R6.LN trait was the only one not showing statistical significance for the genotypic effect, which

Table 1. Bayesian Information Criterion (BIC) values for all spatial models implemented.

Trait	BIC					
	M1	M2	M3	M4	M5	H2
R1.LN	-748.588	-754.61	-742.385	-748.525	-749.08	0.35 (0.03) *
R3.LN	-293.81	-293.509	-289.678	-289.596	-293.55	0.41 (0.03) *
R6.LN	-272.149	-300.92	-266.074	-295.069	-319.1	0.03 (0.03) ^{ns}

* Significant according to the likelihood ratio test with 5% probability.

^{ns} Nonsignificant according to the likelihood ratio test with 5% probability.

R1.LN = leaf nitrogen content in the R1 stage; R3.LN = leaf nitrogen content in the R3 stage; R6.LN = leaf nitrogen content in the R6 stage. H2 = broad-sense heritability for the best model based on the BIC and standard deviation (*SD*; in parentheses).

M1 = model 1; M2 = model 2; M3 = model 3; M4 = model 4; M5 = model 5.

meant that the performance of the lines was the same for all individuals for that trait. The broad-sense heritability varied across traits and populations (range, 0.03 – 0.41), with higher values generally observed in the inbred lines (e.g., R3.LN: 0.41 in lines vs. 0.18 in hybrids). This is evidence that the trait is moderately heritable, which, for breeding, opens a window to select and make genetic progress. Among the trait correlations, only the R1.LN–R3.LN pair showed a statistically significant relationship with a moderate correlation of 0.48 .

Regarding the phenotypic assessment models, model 1, which did not account for spatial correlation, provided the best fit for R3.LN in the inbred line population. For R1.LN, model fit improved with the inclusion of an autoregressive structure for rows (model 2). For the R6.LN trait, the most complex model tested (model 5), achieved the lowest BIC; therefore, it was selected as the best-fitting model (Table 1 and Supplemental Material A Fig. 1A).

Genomic regions were found to be associated with NUE traits

A genome-wide association analysis of the inbred diversity panel was performed to identify genomic variants associated with NUE trait expression. The R6.LN trait was excluded because of the lack of genotypic variance captured (Table 1). Based on the GWAS results of both methods [FarmCPU and mixed linear model (MLM)], a total of 17 significant SNPs reached the Bonferroni threshold across chromosomes (Fig. 2).

Using the MLM, there was no significant SNP from R1.LN, and we identified only one significant SNP for R3.LN. The significant SNP for R1.LN is located on chromosome 10. Using the model FarmCPU, we identified 16 significant SNPs, 11 SNPs from R1.LN, and five SNPs from R3.LN. The significant SNPs for R1.LN were distributed across chromosomes 1, 3, 4, 5, 8, and 9, while significant SNPs that were found for the R3.LN trait were distributed across chromosomes 2, 4, 6, 7, and 8.

In addition, the quantile-quantile plot, which assesses how well the SNP P value distribution fits the null hypothesis of no association, showed that most SNPs aligned closely with the expected distribution along the diagonal line, with only a few deviating significantly (Supplemental Material Fig. 1B). This suggests that the models effectively corrected

for population structure using the kinship matrix while also identifying potential true associations.

Candidate genes were found in significant GWAS regions

In the FarmCPU model, candidate genes were identified for three of the 11 significant SNPs from R1.LN and three of the five significant SNPs from R3.LN (Fig. 2, Table 2, and Supplemental Material B). For the R1.LN trait, in the confidence interval of chromosome number 1 (position 222056270), one gene model was found (Zm00001eb042130). The gene encodes an aminopeptidase involved in the peptide catabolic process. Two significant candidate genes, Zm00001eb002490 and Zm00001eb339390, were identified as candidates within 234 kb of SNP chromosome number 1 (position 7261188) and 276 kb of chromosome number 8 (position 29802249), respectively. These genes encode protein NLP7 and protein NLP4, respectively, which both function as DNA-binding transcription factors involved in controlling nitrate-responsive gene networks. A significant candidate gene, Zm00001eb214960, located 169 kb of SNP chromosome 5 (position 7874493), that encodes an ammonium transporter 10, which is involved in transmembrane ammonium ion transport, was identified.

For R3.LN traits, we selected a candidate gene within 80 kb of the SNP in chromosome 4 (position 192007595). The gene model, Zm00001eb196180, encodes a putative peptide/nitrate transporter (NRT) involved in the transmembrane transport of nitrate. We also found another nitrate transporter gene, Zm00001eb287950, encoding protein NRT1/PTR family 3.1, within 336 kb of the SNP in chromosome 6 (position 157366906). We selected candidate gene Zm00001eb326420, within 167 kb of the SNP in chromosome 7 (position 171898601), which encodes a putative indole-3-acetic acid-amido synthetase GH3.1/GH3.8 and is involved in amino acid ligase activity.

Using the MLM, only one significant candidate gene was associated with the R3.LN trait. The gene model, Zm00001eb406990, located within 108 kb of the SNP of chromosome 10 (position 5917216), encodes a cationic amino acid transporter involved in the transmembrane transport of amino acids. Overall, candidate genes included multiple nitrate regulatory genes, nitrate transporters, and GABA transaminases.

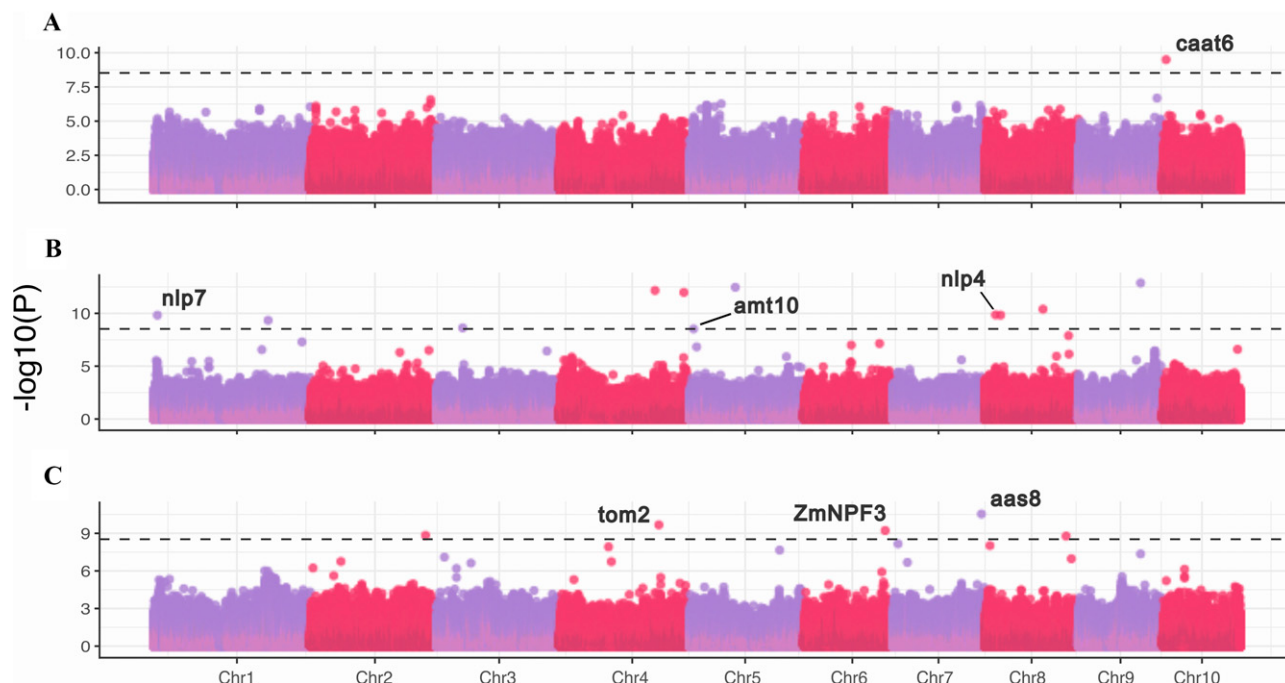


Fig. 2. Manhattan plot with the single nucleotide polymorphisms (SNPs) associated with nitrogen use efficiency (NUE) traits. Some candidate genes were indicated for each SNP. (A) Plot of the R3.LN trait from mixed linear model (MLM). (B) Plot of the R1.LN trait from the FarmCPU model and (C) plot of the R3.LN trait from FarmCPU model. The dotted line represents the Bonferroni threshold, which is the same across all three plots (same number of SNPs used). Note that the scale of the x-axis is different in plots (A), (B), and (C).

No significant SNPs were shared between R1.LN and R3.LN. The complete list of genes and their putative descriptions are provided in Supplementary Material B.

Genomic selection models performance

For the genomic selection models, we evaluated the following key aspects: the impact of different statistical methods on prediction accuracy and the use of cross-validation schemes for validating the genomic model for selection efficiency in a breeding program targeting NUE traits. Among the traits evaluated in the inbred line population, R1.LN showed the highest prediction accuracies (0.36–0.38),

followed by R3.LN (0.29–0.33) and R6.LN (0.03–0.06) (Fig. 3). These results are particularly interesting because this model can be directly used as a training population in a breeding program for either selecting top-performing individuals or discarding undesired performance-based individuals.

Although all statistical models showed similar trends across traits, multitrait models consistently outperformed others, with the MT_GBLUP model yielding the best overall correlation. In addition, the multi-trait model showed a smaller spread for the predictions for all traits compared with single-trait models. Additionally, R6.LN showed low prediction

accuracy. This result may be attributable to the fact that there is very little variability for this trait (Fig. 1).

To assess how well our models predict the performance of untested future generations (hybrids) and untested environments, models were calibrated using the inbred lines (which included the parents) to predict hybrid performance for the R3.LN trait. The predictive accuracy of the models for hybrids was lower than the accuracy observed within the inbred population in which they were trained (Fig. 3B). Among the models tested, GBLUP slightly outperformed BayesB in terms of prediction accuracy (0.14 vs. 0.13).

Table 2. Candidate genes identified from the single nucleotide polymorphisms (SNPs) associated with the nitrogen use efficiency (NUE) traits.

Trait	Model	Chromosome	SNP Pos	Gene ID/MaizeGDB	Gene description	Annotation
R1.LN	FarmCPU	Chr1	222056270	<u>Zm00001eb042130</u>	Aminopeptidase	Peptide binding
	FarmCPU	Chr1	7261188	<u>Zm00001eb002490</u>	Protein NLP2	DNA binding
	FarmCPU	Chr5	7874493	<u>Zm00001eb214890</u>	Ammonium transporter	Ammonium transmembrane transporter activity
	FarmCPU	Chr8	29802249	<u>Zm00001eb339390</u>	RWP-RK domain-containing protein; protein NLP6	DNA binding
R3.LN	FarmCPU	Chr4	192007595	<u>Zm00001eb196180</u>	Major facilitator superfamily (MFS) profile domain-containing protein; carbohydrate transporter/sugar porter/transporter	Transmembrane transporter activity
	FarmCPU	Chr6	157366906	<u>Zm00001eb287950</u>	Protein NRT1/PTR family	Transmembrane transporter activity
	FarmCPU	Chr7	171898601	<u>Zm00001eb326420</u>	Putative indole-3-acetic acid-amido synthetase GH3.1; indole-3-acetic acid-amido synthetase GH3.8	Acid-amino acid ligase activity
	MLM	Chr10	5917216	<u>Zm00001eb406990</u>	Cationic amino acid transporter 6 chloroplastic; cationic amino acid transporter C-terminal domain-containing protein	Amino acid transmembrane transporter activity

The gene identification (ID) refers to the B73 genome (version 5). R1.LN and R3.LN represent the leaf nitrogen measured at stages R1 and R3, respectively.

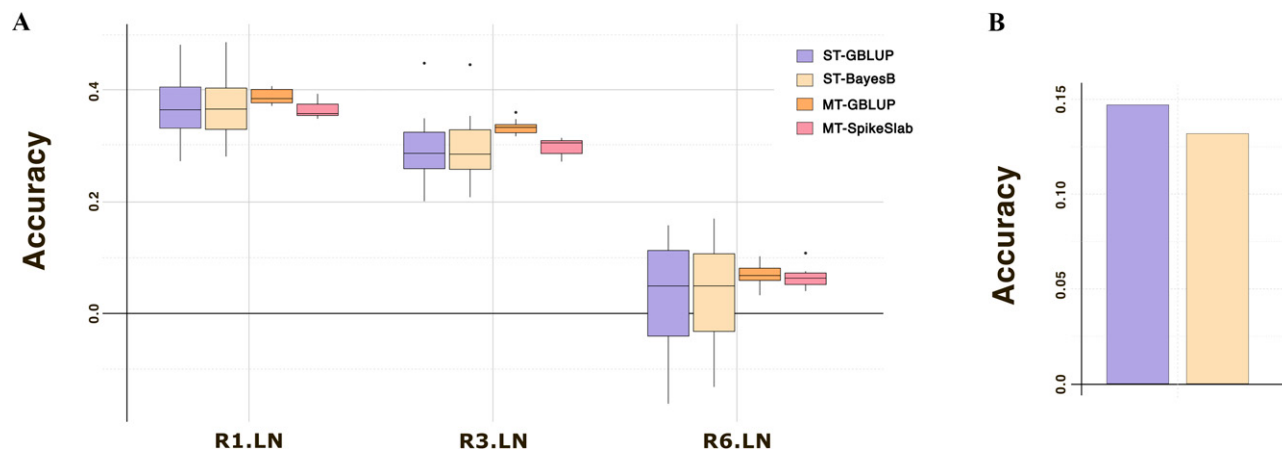


Fig. 3. Prediction accuracy of cross-validation scheme one (CV1) for the R1.LN, R3.LN, and R6.LN traits (A) and cross-validation scheme zero (CV0) for the R3.LN trait (B). The accuracy reported is the Pearson correlation between best linear unbiased estimates (BLUEs) and estimated breeding values from the genomic model. Note that the x-axis in (A) and that in (B) are on different scales. GBLUP = genomic best linear unbiased prediction; MT_GBLUP = multi-trait GBLUP model; MT_SpikeSlab = multi-trait spike-slab model; ST_BayesB = single-trait BayesB model; ST_GBLUP = single GBLUP model.

Cross prediction

The list of the top 100 predicted crosses with the highest performance is provided in Supplemental Material C. These crosses involved 72 unique parents. The average predicted performance for the index for the selected crosses was 2.69, with an estimated inbreeding coefficient of 0.022 for the resulting generation. A correlation of 0.13 was observed between the predicted values of these crosses and the phenotypic values of the corresponding F1 hybrids evaluated in the field (Fig. 4).

Breeding strategies for breeding NUE traits

The set of simulations represented breeding strategies that could be realistically

implemented in the sweet corn breeding program to improve NUE traits (Fig. 5 and Table 3). Among them, the most effective was the application of OCS, which led to a 20% higher genetic gain compared with the benchmark Conv scenario.

This gain was achieved through the optimization of crosses using the SimpleMating algorithm to select the best combinations throughout the pipeline. The target trait (mimicking R3.LN) exhibited the highest performance, while trait 2 (representing indirect selection in R1.LN) showed a lower, yet positive, genetic gain after 20 years of selection. As expected, genomic selection increased the genetic gain for both traits, outperforming the Conv scenario. These results provide strong evidence that genomic tools can effectively

accelerate genetic gain for NUE traits in the sweet corn breeding program.

Discussion

NUE traits presented variability and estimates of heritability suitable for breeding

The heritability estimates and variance components for traits within the diversity panel fall within expected ranges, indicating meaningful breeding potential, consistent with findings in field corn (Mastrodomenico et al. 2018). In terms of genetic architecture, we identified that NUE traits in our population had a complex genetic architecture, with low to moderate heritability values. Notably, an increase in the mean value of R3.LN was

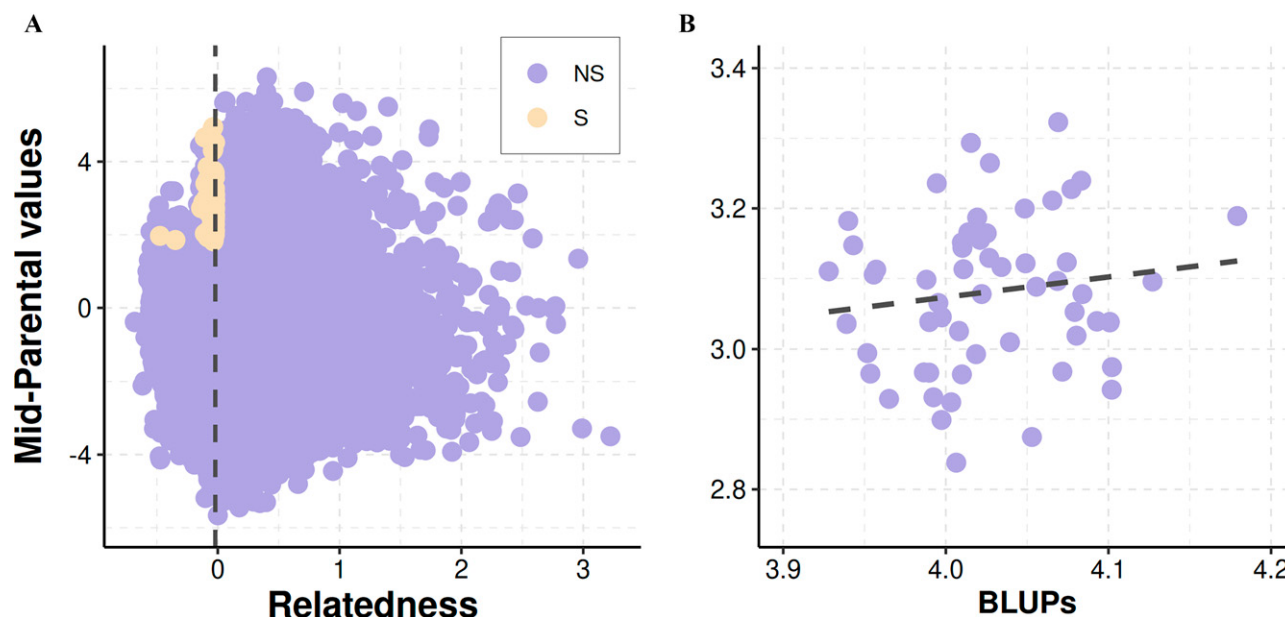


Fig. 4. F1 progenies predicted via SimpleMating. (A) Pairwise prediction of midparental values for all possible crosses among the lines' population plotted against the pairwise covariance coming from the additive relationship matrix. The dotted line represents the culling parameters. (B). Predicted crosses vs. the best linear unbiased estimates (BLUEs) values for the trait measured in the hybrid population. The dotted line represents a regression line. NS = nonselected; S = selected by the SimpleMating algorithm.

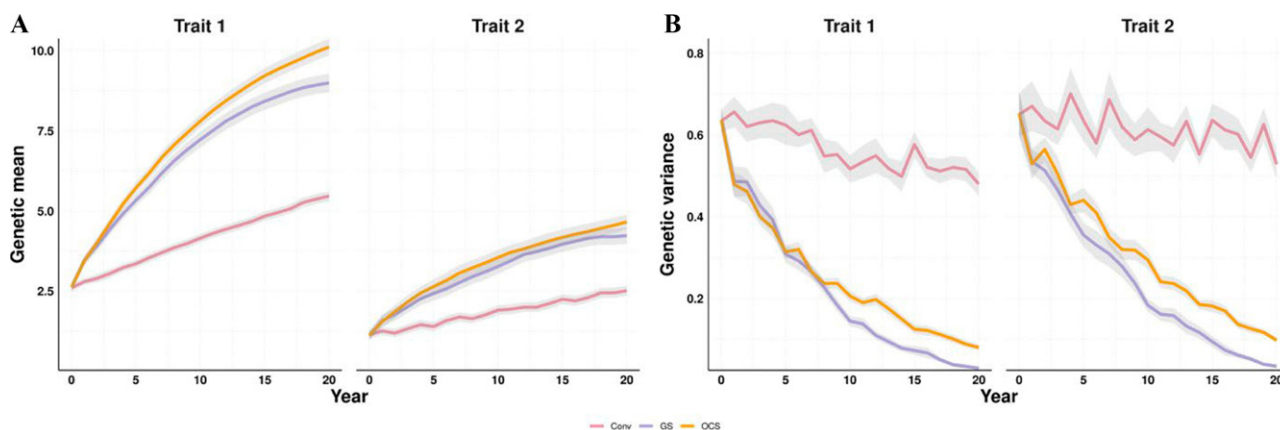


Fig. 5. Additive population mean and variance through 20 years of simulation for the three scenarios simulated. The shading around the curve represents the standard error (*SE*) for the 20 repetitions. (A) Mean and (B) variance. The OCS scenario (using the SimpleMating algorithm) yielded a smaller decrease in genetic diversity. Conv = truncated phenotypic selection; GS = truncated genomic selection; MPV = midparent value; OCS = optimum cross selection with the MPV-based performance.

observed in hybrids derived from the diversity panel, likely reflecting heterosis (Labroo et al. 2021), because these hybrids were generated from a subset of the inbred lines. Within the inbred population, R1.LN and R3.LN showed a moderate positive correlation (0.48), suggesting a degree of pleiotropy. However, the fact that this correlation is not perfect implies the existence of partially independent genetic mechanisms underlying the two traits. This finding underscores the importance of considering potential indirect responses when selecting for one trait because it may impact others, and it is relevant for designing effective breeding strategies (Covarrubias-Pazarán et al. 2022; Marulanda et al. 2021).

The application of spatial models for genotypic evaluation appeared to capture certain patterns of field variability affecting the traits under study. Interestingly, although all traits were measured within the same trial, the spatial correction patterns differed. This discrepancy likely reflects environmental influences that vary across different stages of plant development. For R6.LN, in particular, a row-column effect and a nugget effect were included in the spatial model (Supplemental Material A Fig. B). The nugget effect helps account for intrinsic measurement errors or small-scale environmental variation (Werner et al. 2024). While the inclusion of this component improved overall model performance, the genotypic effect for R6.LN was not statistically significant. This suggests an absence of genetic variation for this trait among the evaluated genotypes, implying that selection would yield no genetic gain. This conclusion is further supported by the very low heritability estimate for R6.LN

($H^2 = 0.03$). In summary, spatial models can enhance the precision of genotypic evaluations and, thus, increase the efficiency of breeding programs targeting NUE. However, their effectiveness relies on the presence of genetic variation in the trait of interest.

AI screening can identify possible NUE candidate genes from genome-wide association analyses

We used GWAS to identify candidate genes associated with NUE traits, specifically R1.LN and R3.LN. Because of the highly quantitative nature of these traits, we used the FarmCPU method based on the hypothesis that it would outperform the traditional MLM, particularly for traits controlled by many small-effect SNPs. This hypothesis was supported by the following results: FarmCPU identified candidate genes for both traits, whereas the MLM detected associations for only one trait. In addition, the variance explained by the significant markers had a small value per se, which was also an indication of the quantitative nature of these traits.

Gene mining and annotation are the most laborious steps of running a GWAS pipeline. However, with the advances of AI in several fields, some tools could be used to speed this process. After identifying the genes near the significant SNPs for both traits, we used the SCISPACE AI tool to perform an initial screening to identify genes related to NUE traits in sweet corn. Among those, we could identify the following four relevant categories of genes: nitrogen uptake genes, including the high-affinity nitrate transporters (NRT2.2 and NRT2.5) and the low-affinity nitrate

transporters (ZmNPF7.9 or NRT1.5) essential for seed development, facilitating nitrate transport from maternal tissues to the endosperm (Wei et al. 2021); nitrogen metabolism and assimilation genes, including glutamine synthetase and asparagine synthetase, which convert inorganic nitrogen into organic forms for plant use (Singh et al. 2023), and ammonium transporter 1, which plays a role in ammonium uptake and transport; nitrogen utilization and efficiency, including the transmembrane amino acid transporter family protein associated with nitrogen compound metabolic processes and NUE, and the MADS26 transcription factor, which, when overexpressed, enhances nitrate utilization (Zhang et al. 2024); and nitrogen remobilization genes, including ZmASR6 and the ATP-dependent Clp protease gene, which have been linked to nitrogen remobilization efficiency, particularly during leaf senescence, and are part of a complex regulatory network involving hormone signaling (Gong et al. 2020).

In this context, we highlighted that the use of the AI tool was instrumental in accelerating the candidate gene identification process, offering a promising avenue to streamline candidate gene screening within breeding programs. The GWAS is known to include true positive and false positive quantitative trait loci (Fernando et al. 2004; Hayes 2013). Future functional studies are essential to validating the efficiency and accuracy of the SCISPACE tool following the GWAS. While validating the candidate genes identified in this study will be strategic for their introgression into elite germplasm to enhance NUE performance, our work represents an initial step toward that goal.

Genomic selection model performance enables effective selection and deployment in sweet corn

Breeding for NUE traits in the sweet corn program can greatly benefit from the implementation of genomic selection across multiple stages of the pipeline. First, genomic prediction can be used to identify the best

Table 3. Genetic mean, variance, and standard error (*SE*; in parentheses) for trait 1 (targeted) and trait 2 (indirect selection). Trait 1 and trait 2 were mimicking R1.LN and R3.LN traits.

Scenario	Genetic mean		Genetic variance	
	Trait 1	Trait 2	Trait 1	Trait 2
Conv	5.46 (0.158)	2.51 (0.152)	0.47 (0.028)	0.52 (0.035)
GS	8.99 (0.286)	4.22 (0.256)	0.02 (0.004)	0.03 (0.004)
OCS	10.11 (0.265)	4.65 (0.232)	0.07 (0.007)	0.09 (0.006)

Conv = conventional; GS = genomic selection; OCS = optimal cross selection.

inbred lines to advance to hybrid development (Graciano et al. 2025; Peixoto et al. 2024a). At this stage, a model capturing only additive genetic effects may be sufficient to select the top-performing individuals. Second, the same model can be used to estimate marker effects and, from a pool of potential parental candidates, facilitate the design of crosses that are more likely to enhance NUE traits (Peixoto et al. 2024a). Ultimately, because the primary goal of the sweet corn breeding program is the development of high-performing hybrids, genomic models can also be applied to predict hybrid performance (Peixoto et al. 2024c; Zystro et al. 2021a, 2021b). This enables the selection of the most promising combinations to advance for field evaluation.

As a vegetable crop, sweet corn presents multiple breeding targets, making the selection of superior commercial genotypes a complex task. While this complexity can hinder genetic gains, it also presents an opportunity for quantitative geneticists to apply multi-trait genomic selection models that leverage correlations among traits (Calus et al. 2013; Cui et al. 2020; Sandhu et al. 2022). In our study, the use of a multi-trait GBLUP model led to increased predictive performance and genetic gain compared with its single-trait counterpart, resulting in improvements of 5% for R1.LN, 12.5% for R3.LN, and 120% for R6.LN. This added benefit of genomic information in multi-trait models is most likely attributable to the moderate correlation between traits.

Our results highlight that the genomic models for selecting lines with outstanding performance in NUE traits and for guiding cross prediction demonstrated good predictive accuracy and could be implemented immediately in the breeding pipeline. However, the performance of the model when predicting hybrid performance, particularly the R3.LN trait, was not as strong as that in the inbred line panel. The following several factors may explain this, despite the predictive values not being necessarily poor: only a subset of the inbred lines was used as parents, and not all of those hybrid parents were genotyped; the number of hybrids generated was limited, which could have introduced bias in the analyses when the experiment was small and/or poorly replicated; the experimental design for evaluating hybrids was suboptimal (although 108 hybrids were taken to the field, only 49 had phenotypes for the target trait); and parents were present in only one or a few crosses.

As reported by Peixoto et al. (2024b), we implemented a model with nonadditive effects (additive plus dominance effects). We hypothesized that this could improve prediction accuracy in the hybrid population. However, following the same pattern of such work, we found lower or similar prediction accuracy compared with that of the model with additive effects only (results not shown). Despite the decrease in accuracy when predicting untested hybrids, the predictive ability can still be valuable for breeding. Particularly,

the model can still be used to eliminate low-performing candidates, reduce phenotyping efforts, and ensure that the most outstanding candidates reach the field stage (Beyene et al. 2021).

Deployment of genomic tools in the sweet corn breeding program improves short-term and long-term genetic gains

The genomic model applied to the inbred line population demonstrated reliable predictive accuracy for both traits that were studied (R1.LN and R3.LN). Because of the moderate genetic correlation between these traits and their high phenotyping costs, the following key question arose: what would be the impact of applying genomic selection targeting only one trait on the long-term genetic gain for both? The results indicated a clear benefit of implementing genomic selection for NUE traits. Direct genetic gain was substantial for trait 1 (mimicking R3.LN); notably, a positive genetic response was also observed for trait 2 (mimicking R1.LN). This outcome emphasized the potential of indirect selection to improve correlated traits (Marulanda et al. 2021), and it is useful for defining breeding goals that target NUE traits in the sweet corn breeding program.

Therefore, the use of genomic selection for predicting crosses was highly recommended and not only improved trait performance over the long term but also helped maintain genetic variance when the SimpleMating algorithm was applied with appropriate constraints. Notably, the strategy adopted in this study did not incur additional costs to the breeding program because both prediction and optimization were based on existing marker effects and the genomic relationship matrix available in both scenarios. Beyond enabling a cost-neutral increase in genetic gain, the implementation of OCS transformed the breeding program into a more data-driven operation (Akdemir and Sánchez 2016; Gorjanc et al. 2018; Peixoto et al. 2025).

Breeding for multiple traits in sweet corn, although challenging, can be efficient

In a sweet corn breeding program, while some traits, such as plant height, ear width, and taper, are relatively easy and less laborious to measure (Gonzalez et al. 2022), others require more arduous and time-consuming techniques (e.g., NUE, phytoglycogen content, and disease resistance scores) (Mahon 2023). As we advance through the selection pipeline, it is crucial to understand how the improvement of one trait may impact others. This impact is known as the indirect response, and it arises from the genetic correlation between the main trait and secondary traits (Mrode 2014). A negative genetic correlation can lead to an undesirable tradeoff, whereby improving one trait causes a decline in another. To mitigate such outcomes, a selection index can be used. This approach assigns weights to each trait and combines them into a single index value, enabling balanced selection that accounts for the genetic

relationships among traits (Batista et al. 2021; Marulanda et al. 2021; Silva et al. 2021).

For NUE traits, indirect selection appears to be beneficial. First, because these traits are labor-intensive and expensive to measure, targeting only one trait (i.e., measuring it in the laboratory) can still yield a positive response in the correlated trait. Second, borrowing information across traits using a multi-trait genomic selection framework can enhance prediction performance. This approach is particularly advantageous for traits that are difficult or costly to measure, expressed in late phenological stages, or have low heritability because it can improve predictive accuracy through shared genetic signals (Calus et al. 2013; Lyra et al. 2017; Sandhu et al. 2022).

Conclusions

Through this work, we computed genetic parameters and suggested that NUE traits in sweet corn have complex genetic architecture but potential for breeding. Additionally, we provided GWAS hits and candidate genes that should be further validated in downstream validation analyses. We also drew attention to the potential of genomic selection to accelerate and increase genetic gains for those traits and opened a window for testing the genomic model in larger populations, especially hybrids across different breeding states. Finally, we showed that combining genomic selection with cross prediction and optimization can increase genetic gain for NUE traits in the long-term. Therefore, including genomic tools to breed for NUE traits is effective and opens a new window for future implementations. Ultimately, these results and strategies can be used in the University of Florida breeding program and any other program that targets NUE.

References cited

- Akdemir D, Sánchez JI. 2016. Efficient breeding by genomic mating. *Front Genet.* 7:210. <https://doi.org/10.3389/fgene.2016.00210>.
- Amadeu RR, García AAF, Muñoz PR, Ferrão LFV. 2023. AGHmatrix: Genetic relationship matrices in R. *Bioinformatics.* 39(7):btad445. <https://doi.org/10.1093/bioinformatics/btad445>.
- Baseggio M, Murray M, Wu D, Ziegler G, Kaczmar N, Chamness J, Hamilton JP, Buell CR, Vatamaniuk OK, Buckler ES, Smith ME, Baxter I, Tracy WF, Gore MA. 2021. Genome-wide association study suggests an independent genetic basis of zinc and cadmium concentrations in fresh sweet corn kernels. *G3 (Bethesda).* 11(8):jkab186. <https://doi.org/10.1093/g3journal/jkab186>.
- Batista LG, Gaynor RC, Margarido GRA, Byrne T, Amer P, Gorjanc G, Hickey JM. 2021. Long-term comparison between index selection and optimal independent culling in plant breeding programs with genomic prediction. *PLoS One.* 16(5):e0235554. <https://doi.org/10.1371/journal.pone.0235554>.
- Beyene Y, Gowda M, Pérez-Rodríguez P, Olsen M, Robbins KR, Burgueño J, Prasanna BM, Crossa J. 2021. Application of genomic selection at the early stage of breeding pipeline in

- Tropical Maize. *Front Plant Sci.* 12:685488. <https://doi.org/10.3389/fpls.2021.685488>.
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015. Time series analysis: forecasting and control. John Wiley & Sons, New York, NY, USA.
- Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R. 2017. ASReml-R reference manual version 4. VSN International Ltd., Hemel Hempstead, UK.
- Calus MPL, Berry DP, Banos G, de Haas Y, Veerkamp RF. 2013. Genomic selection: The option for new robustness traits? *Adv Animal Biosci.* 4(3):618–625. <https://doi.org/10.1017/S2040470013000186>.
- Chen GK, Marjoram P, Wall JD. 2009. Fast and flexible simulation of DNA sequence data. *Genome Res.* 19(1):136–142. <https://doi.org/10.1101/gr.083634.108>.
- Coelho IF, Peixoto MA, Marçal TDS, Bernardeli A, Alves RS, Lima RO, Reis EF, Dos, Bhering LL. 2021. Accounting for spatial trends in multi-environment diallel analysis in maize breeding. *PLoS One.* 16(10):e0258473. <https://doi.org/10.1371/journal.pone.0258473>.
- Colantonio VN. 2022. Applications of Artificial Intelligence for the Genetic Improvement of Crop Plants (PhD Diss). University of Florida, Gainesville, FL, USA.
- Coque M, Gallais A. 2007. Genetic variation for nitrogen remobilization and postsilking nitrogen uptake in maize recombinant inbred lines: Heritabilities and correlations among traits. *Crop Sci.* 47(5):1787–1796. <https://doi.org/10.2135/cropsci2007.02.0096>.
- Covarrubias-Pazarán G, Gebeyehu Z, Gemenet D, Werner C, Labroo M, Sirak S, Coaldrake P, Rabbi I, Kayondo SI, Parkes E, Kanju E, Mbanjo EGN, Agbona A, Kulakow P, Quinn M, Debaene J. 2022. Breeding schemes: What are they, how to formalize them, and how to improve them? *Front Plant Sci.* 12:791859. <https://doi.org/10.3389/fpls.2021.791859>.
- Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, Bentley AR. 2021. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. *Front Plant Sci.* 12:651480–651486. <https://doi.org/10.3389/fpls.2021.651480>.
- Cui Z, Dong H, Zhang A, Ruan Y, He Y, Zhang Z. 2020. Assessment of the potential for genomic selection to improve husk traits in maize. *G3 (Bethesda).* 10:3741–3749. <https://doi.org/10.1534/g3.120.401600>.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43(5):491–498. <https://doi.org/10.1038/ng.806>.
- Fernando RL, Nettleton D, Southey BR, Dekkers JCM, Rothschild MF, Soller M. 2004. Controlling the proportion of false positives in multiple dependent tests. *Genetics.* 166(1):611–619. <https://doi.org/10.1534/genetics.166.1.611>.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.
- Gaynor RC, Gorjanc G, Hickey JM. 2021. Alpha-SimR: An R package for breeding program simulations. *G3: Genes, Genomes, Genetics.* 11:1–21.
- Gheith EMS, El-Badry OZ, Lamlom SF, Ali HM, Siddiqui MH, Ghareeb RY, El-Sheikh MH, Jebri J, Abdelsalam NR, Kandil EE. 2022. Maize (*Zea mays* L.) productivity and nitrogen use efficiency in response to nitrogen application levels and time. *Front Plant Sci.* 13:941343. <https://doi.org/10.3389/fpls.2022.941343>.
- Gilmour AR, Cullis BR, Verbyla AP, Verbyla AP. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *J Agric Biol Environ Stat.* 2(3):269–293. <https://doi.org/10.2307/1400446>.
- Gong X, Li J, Ma H, Chen G, Dang K, Yang P, Wang M, Feng B. 2020. Nitrogen deficiency induced a decrease in grain yield related to photosynthetic characteristics, carbon–nitrogen balance and nitrogen use efficiency in proso millet (*Panicum miliaceum* L.). *Arch Agron Soil Sci.*
- Gonzalez JM, Ghosh N, Colantonio V, Pereira FdC, Pinto RA Jr, Wasson C, Leach KA, Resende MFR Jr. 2022. EarCV: An open-source, computer vision package for maize ear phenotyping. *Plant Phenome J.* 5(1):e20055. <https://doi.org/10.1002/ppj2.20055>.
- Gorjanc G, Gaynor RC, Hickey JM. 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet.* 131(9):1953–1966. <https://doi.org/10.1007/s00122-018-3125-3>.
- Govindasamy P, Muthusamy SK, Bagavathiannan M, Mowrer J, Jagannadham PTK, Maity A, Halli HM, Sujayanand GK, Vadivel R, Das TK, Raj R, Pooniya V, Babu S, Rathore SS, Muralikrishnan L, Tiwari G. 2023. Nitrogen use efficiency—a key to enhance crop productivity under a changing climate. *Front Plant Sci.* 14:1121073. <https://doi.org/10.3389/fpls.2023.1121073>.
- Graciano RP, Peixoto MA, Leach KA, Suzuki N, Gustin JL, Settles AM, Armstrong PR, Resende MFR. 2025. Integrating phenomic selection using single-kernel near-infrared spectroscopy and genomic selection for corn breeding improvement. *Theor Appl Genet.* 138(3):60. <https://doi.org/10.1007/s00122-025-04843-w>.
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics.* 12:186. <https://doi.org/10.1186/1471-2105-12-186>.
- Hayes B. 2013. Overview of statistical methods for genome-wide association studies (GWAS). *Genome-wide association studies and genomic prediction.* Methods Mol Biol. 1019:149–169.
- He K, Xu S, Zhang X, Li Y, Chang L, Wang Y, Shi Y, Cui T, Dong Y, Lan T, Liu X, Du Y, Zhang R, Liu J, Xue J. 2020. Mining of candidate genes for nitrogen use efficiency in maize based on genome-wide association study. *Mol Breeding.* 40(9):1–17. <https://doi.org/10.1007/s11032-020-01163-3>.
- Henderson CR. 1974. General flexibility of linear model techniques for sire evaluation. *J Dairy Sci.* 57(8):963–972. [https://doi.org/10.3168/jds.S0022-0302\(74\)84993-3](https://doi.org/10.3168/jds.S0022-0302(74)84993-3).
- Holland JB, Piepho H-P. 2024. Don't BLUP twice. *G3 (Bethesda).* 14(12):jkae250. <https://doi.org/10.1093/g3journal/jkae250>.
- Hu Y, Colantonio V, Müller BSF, Leach KA, Nanni A, Finegan C, Wang B, Baseggio M, Newton CJ, Juhl EM, Hislop L, Gonzalez JM, Rios EF, Hannah LC, Swarts K, Gore MA, Hennen-Bierwagen TA, Myers AM, Settles AM, Tracy WF, Resende MFR. 2021. Genome assembly and population genomic analysis provide insights into the evolution of modern sweet corn. *Nat Commun.* 12(1):1227. <https://doi.org/10.1038/s41467-021-21380-4>.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. 2008. Efficient control of population structure in model organism association mapping. *Genetics.* 178(3):1709–1723. <https://doi.org/10.1534/genetics.107.080101>.
- Labroo MR, Studer AJ, Rutkowski JE. 2021. Heterosis and hybrid crop breeding: A multidisciplinary review. *Front Genet.* 12:643761. <https://doi.org/10.3389/fgene.2021.643761>.
- Liu R, Zhang H, Zhao P, Zhang Z, Liang W, Tian Z, Zheng Y. 2012. Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. *Plant Mol Biol Rep.* 30(2):297–308. <https://doi.org/10.1007/s11105-011-0346-x>.
- Lyra DH, de Freitas Mendonça L, Galli G, Alves FC, Granato ISC, Fritsche-Neto R. 2017. Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol Breeding.* 37(6). <https://doi.org/10.1007/s11032-017-0681-1>.
- Mahon A. 2023. Leveraging Genomics and High-Throughput Phenotyping To Improve Selection of Phytochemicals in Sweet Corn Breeding. University of Florida, Gainesville, FL, USA.
- Marinho CD, Coelho IF, Peixoto MA, Carvalho GA Jr, Resende MFR Jr. 2022. Genomic selection as a tool for maize cultivars development. *Rev Bras Milho Sorgo.* 21:e1285.
- Marulanda JJ, Mi X, Utz HF, Melchinger AE, Würschum T, Longin CFH. 2021. Optimum breeding strategies using genomic and phenotypic selection for the simultaneous improvement of two traits. *Theor Appl Genet.* 134(12):4025–4042. <https://doi.org/10.1007/s00122-021-03945-5>.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4):1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>.
- Moose S, Below FE. 2009. Biotechnology approaches to improving maize nitrogen use efficiency, p 65–77. *Molecular genetic approaches to maize improvement.* Springer, New York, NY, USA.
- Mrode RA. 2014. Linear models for the prediction of animal breeding values. Cabi Digital Library.
- Patterson HD, Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika.* 58(3):545–554. <https://doi.org/10.1093/biomet/58.3.545>.
- Peixoto MA, Amadeu RR, Bhering LL, Ferrão LFV, Munoz PR, Resende MFR. 2025. SimpleMating: R-package for prediction and optimization of breeding crosses using genomic selection. *Plant Genome.* 18(1):e20533. <https://doi.org/10.1002/tpg2.20533>.
- Peixoto MA, Coelho IF, Leach KA, Lübberstedt T, Bhering LL, Resende MFR. 2024a. Use of simulation to optimize a sweet corn breeding program: Implementing genomic selection and doubled haploid technology. *G3 (Bethesda).* 14(8):jkae128. <https://doi.org/10.1093/g3journal/jkae128>.
- Peixoto MA, Leach KA, Jarquin D, Flannery P, Zystro J, Tracy WF, Bhering L, Resende MFR. 2024b. Utilizing genomic prediction to boost hybrid performance in a sweet corn breeding program. *Front Plant Sci.* 15:1293307. <https://doi.org/10.3389/fpls.2024.1293307>.
- Pérez Rodríguez P, de los Campos G. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 198(2):483–495. <https://doi.org/10.1534/genetics.114.164442>.
- Pérez-Rodríguez P, de los Campos G. 2022. Multi-trait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics.* 222(1):iyac112. <https://doi.org/10.1093/genetics/iyac112>.
- Plett D, Garnett T, Okamoto M. 2017. Molecular genetics to discover and improve nitrogen use efficiency in crop plants, p 93–122. *Plant*

- macronutrient use efficiency. Elsevier, Amsterdam, the Netherlands.
- PubGenius Inc. 2025. SciSpace AI tool. <https://typeset.io/>.
- Rao CR. 1973. Linear statistical inference and its applications. Wiley, New York, NY, USA.
- Sanchez DL, Santana AS, Morais PIC, Peterlini E, De La Fuente G, Castellano MJ, Blanco M, Lübberstedt T. 2023. Phenotypic and genome-wide association analyses for nitrogen use efficiency related traits in maize (*Zea mays* L.) exotic introgression lines. *Front Plant Sci.* 14:1270166. <https://doi.org/10.3389/fpls.2023.1270166>.
- Sandhu KS, Patil SS, Aoun M, Carter AH. 2022. Multi-trait multi-environment genomic prediction for end-use quality traits in winter wheat. *Front Genet.* 13:831020. <https://doi.org/10.3389/fgene.2022.831020>.
- Schwarz G. 1978. Estimating the dimension of a model. *Ann Stat.* 6:461–464.
- Silva LA, Peixoto MA, Peixoto LA, Romero JV, Bhering LL. 2021. Multi-trait genomic selection indexes applied to identification of superior genotypes. *Bragantia.* 80. <https://doi.org/10.1590/1678-4499.20200381>.
- Singh R, Saripalli G, Kumar A, Gautam T, Singh SK, Gahlaut V, Kumar S, Meher PK, Mishra RP, Singh VK, Sharma PK, Balyan HS, Gupta PK. 2023. QTL analysis for nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Euphytica.* 219(1):9. <https://doi.org/10.1007/s10681-022-03134-5>.
- Stefanova KT, Smith AB, Cullis BR. 2009. Enhanced diagnostics for the spatial analysis of field trials. *JABES.* 14(4):392–410. <https://doi.org/10.1198/jabes.2009.07098>.
- US Department of Agriculture. 2025. United States Department of Agriculture–National Agricultural Statistics Service QuickStats.
- Van der Auwera GA, O'Connor BD. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media, Sebastopol, CA, USA.
- VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci.* 91(11):4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Wang J, Zhang Z. 2021. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics.* 19(4):629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>.
- Wani SH, Vijayan R, Choudhary M, Kumar A, Zaid A, Singh V, Kumar P, Yasin JK. 2021. Nitrogen use efficiency (NUE): Elucidated mechanisms, mapped genes and gene networks in maize (*Zea mays* L.). *Physiol Mol Biol Plants.* 27(12):2875–2891. <https://doi.org/10.1007/s12298-021-01113-z>.
- Wei YM, Ren ZJ, Wang BH, Zhang L, Zhao YJ, Wu JW, Li LG, Zhang XS, Zhao XY. 2021. A nitrate transporter encoded by *ZmNPF7.9* is essential for maize seed development. *Plant Sci.* 308:110901. <https://doi.org/10.1016/j.plantsci.2021.110901>.
- Werner CR, Gemenet DC, Tolhurst DJ. 2024. FieldSimR: An R package for simulating plot data in multi-environment field trials. *Front Plant Sci.* 15:1330574. <https://doi.org/10.3389/fpls.2024.1330574>.
- Wilkinson GN, Eckert SR, Hancock TW, Mayo O. 1983. Nearest neighbour (NN) analysis of field experiments. *J Royal Stat Soc: Series B (Methodological).* 45(2):151–178. <https://doi.org/10.1111/j.2517-6161.1983.tb01240.x>.
- Yu X, Keitel C, Zhang Y, Wangeci AN, Dijkstra FA. 2022. Global meta-analysis of nitrogen fertilizer use efficiency in rice, wheat and maize. *Agric Ecosyst Environ.* 338:108089. <https://doi.org/10.1016/j.agee.2022.108089>.
- Zhang Z, Peng C, Xu W, Li Y, Qi X, Zhao M. 2024. Genome-wide association study of agronomic traits related to nitrogen use efficiency in Henan wheat. *BMC Genomics.* 25(1):7. <https://doi.org/10.1186/s12864-023-09922-0>.
- Zystro J, Peters T, Miller K, Tracy WF. 2021a. Classical and genomic prediction of hybrid sweet corn performance in organic environments. *Crop Sci.* 61(3):1698–1708. <https://doi.org/10.1002/csc2.20400>.
- Zystro J, Peters TE, Miller KM, Tracy WF. 2021b. Inbred and hybrid sweet corn genotype performance in diverse organic environments. *Crop Sci.* 61(4):2280–2293. <https://doi.org/10.1002/csc2.20457>.