Article

Factors Influencing Genomic Prediction Accuracies of Tropical Maize Resistance to Fall Armyworm and Weevils

Arfang Badji [1,*][iD]**, Lewis Machida** [2]**, Daniel Bomet Kwemoi** [3,*][iD]**, Frank Kumi** [4][iD]**, Dennis Okii** [1]**, Natasha Mwila** [1]**, Symphorien Agbahoungba** [5]**, Angele Ibanda** [1][iD]**, Astere Bararyenya** [1]**, Selma Ndapewa Nghituwamhata** [1][iD]**, Thomas Odong** [1]**, Peter Wasswa** [1]**, Michael Otim** [3]**, Mildred Ochwo-Ssemakula** [1][iD]**, Herbert Talwana** [1]**, Godfrey Asea** [3]**, Samuel Kyamanywa** [1] **and Patrick Rubaihayo** [1]

[1] Department of Agricultural Production, Makerere University, Kampala P.O. Box 7062, Uganda; dennisokii@gmail.com (D.O.); mwilanatasha@yahoo.co.uk (N.M.); angeltanzito@gmail.com (A.I.); barastere@gmail.com (A.B.); snghituwamhata@gmail.com (S.N.N.); thomas.l.odong@gmail.com (T.O.); wasswa@caes.mak.ac.ug (P.W.); mknossemakula@gmail.com (M.O.-S.); haltalwana@gmail.com (H.T.); Skyamanywa@gmail.com (S.K.); prubaihayo@gmail.com (P.R.)
[2] Alliance Bioversity-CIAT, Africa-Office, Kampala P.O. Box 24384, Uganda; l.machida@cgiar.org
[3] National Crops Resource Research Institute, Kampala P.O. Box 7084, Uganda; otim_michael@yahoo.com (M.O.); grasea9@gmail.com (G.A.)
[4] Department of Crop Science, University of Cape Coast, Cape Coast PMB, Ghana; frankkumifk@gmail.com
[5] Laboratory of Applied Ecology, University of Abomey-Calavi, Cotonou 01BP 526, Benin; agbahoungbasymphorien@gmail.com
[*] Correspondence: arbad2009@caes.mak.ac.ug (A.B.); kdbomet@gmail.com (D.B.K.)

Abstract: Genomic selection (GS) can accelerate variety improvement when training set (TS) size and its relationship with the breeding set (BS) are optimized for prediction accuracies (PAs) of genomic prediction (GP) models. Sixteen GP algorithms were run on phenotypic best linear unbiased predictors (BLUPs) and estimators (BLUEs) of resistance to both fall armyworm (FAW) and maize weevil (MW) in a tropical maize panel. For MW resistance, 37% of the panel was the TS, and the BS was the remainder, whilst for FAW, random-based training sets (RBTS) and pedigree-based training sets (PBTSs) were designed. PAs achieved with BLUPs varied from 0.66 to 0.82 for MW-resistance traits, and for FAW resistance, 0.694 to 0.714 for RBTS of 37%, and 0.843 to 0.844 for RBTS of 85%, and these were at least two-fold those from BLUEs. For PBTS, FAW resistance PAs were generally higher than those for RBTS, except for one dataset. GP models generally showed similar PAs across individual traits whilst the TS designation was determinant, since a positive correlation ($R = 0.92^{***}$) between TS size and PAs was observed for RBTS, and for the PBTS, it was negative ($R = 0.44^{**}$). This study pioneered the use of GS for maize resistance to insect pests in sub-Saharan Africa.

Keywords: prediction accuracy; mixed linear and Bayesian models; machine learning algorithms; training set size and composition; parametric and nonparametric models

1. Introduction

Insect damage on maize plants and stored grains potentially impedes food security in Africa [1–3]. The fall armyworm (FAW) and stem borers in the field, and the maize weevils (MWs) in storage facilities, are some of the most devastating insect pests on the continent. These insect pests cause yield losses ranging from 10–90% leading to loss of grain marketability, and consumer health concerns due to the possible contamination of the grain with mycotoxins, such as aflatoxins [3–6]. In Africa, tremendous efforts were made during the last two decades to build host plant resistance to insect pests in maize through traditional pedigree (phenotypic)-based selection (PS) with substantial desirable results. Several Africa-adapted maize lines were developed and successfully tested for resistance to MW damage on grains [7–12]. Some of the success stories are from the International Center for Maize and Wheat Improvement (CIMMYT) of Kenya through the Insect Resistant Maize

for Africa project (IRMA) that produced several storage pest and stem borer resistant maize lines [7,8,13–15]. On the other hand, the FAW is a new pest on the continent, first reported in 2016 in West and Central African countries [16], from where it spread throughout the African continent [17]. Hence, although efforts to develop FAW resistant varieties are underway at several institutions, including CIMMYT, published reports of FAW resistant varieties are not yet available [18,19].

The complex nature of insect resistance traits makes PS slow and expensive, and thus, difficult to implement, especially for resource-constrained breeding programs [20]. Application of traditional marker-assisted selection (MAS) is hampered by the necessity to first discover resistance-associated genomic regions through genetic linkage and genome-wide association mapping methods, both with several shortcomings, especially for complex traits [21–23]. In addition, genetic linkage and genome-wide association mapping studies have seldom been explored in African germplasm [8,24], which further impedes the application of MAS in the development of insect resistance maize germplasm in Africa. In a previous study, we discovered several quantitative trait nucleotides and genes that are putatively associated with FAW and MW resistance, confirming the quantitative nature of these traits, hence the difficulty in improving these traits through MAS [25]. An alternative to both PS and MAS is genomic selection (GS), which uses whole-genome markers to perform genomic prediction (GP) of breeding values of unphenotyped genotypes, from which one can select superior candidate genotypes for crossing to produce hybrids or to advance to the next generation [26]. GS was reported to achieve up to threefold annual genetic gain in maize improvement when compared to MAS, due to a more efficient accounting of trait-associated quantitative trait loci (QTL), faster selection cycles, and lower phenotyping costs [27–33].

Several statistical and machine learning GP models with various strengths and weaknesses have been developed to adapt to different contexts that are partly influenced by the genetic architecture of traits (number and effect size of QTL, proportions of additive and non-additive genetic effects) and reproductive classes of plants (allogamous vs. autogamous vs. clonally propagated) [34–36]. Therefore, to effectively implement GS in crop improvement programs, it is necessary to employ a holistic approach to determine the best GP strategy for particular breeding targets for given crop species [31,37]. Statistical models for GS vary in their prior assumptions and treatment of marker effects [31]. Parametric models focus on parameter estimates rather than prediction, while nonparametric algorithms give priority to prediction and have fewer assumptions [38]. Some parametric methods assume the SNP effects follow a normal distribution with equal variance for all loci, which seems unrealistic in practice.

Representative parametric methods are ridge regression best linear unbiased predictors (RR-BLUP) [39] and genomic BLUP (GBLUP) [40]. GBLUP was the first GP method to be developed, and it replaced the traditional pedigree-based relationship matrix with a genomic information-based matrix to improve prediction accuracies (PAs) [41]. Parametric methods BayesA [26] and weighted Bayesian shrinkage regression (wBSR) [42], on the other hand, consider a prior distribution of effect with a higher probability of moderate to large effects. Regarding parametric models such as BayesB [43] and BayesCπ [44], assumptions are made that consider the effects of some SNPs to be zero. The Bayesian least absolute shrinkage and selection operator (Bayesian LASSO) assumes that the effects of all markers follow a double exponential distribution [45], whilst the Bayesian sparse linear mixed model (BSLMM), a parametric method developed by Zhou et al. [46], combines the hypotheses of both GBLUP and Bayesian methods and achieves higher PAs than BayesCπ and BayesLASSO. Nonparametric or semi-parametric approaches such as random forest and reproducing kernel Hilbert space (RKHS) [47,48] are better suited for accounting for non-additive genetic effects (37,38), in contrast with parametric genomic prediction models [23,38,47,49]. Several studies compared the performances of GP models under different conditions. In a simulation study, Meuwissen et al. [26] found that while GBLUP achieved PAs of up to 73.2%, BayesA and BayesB comparatively provided additional increases of

around 9% and 16%, respectively. However, when a population is composed of close relatives and the target traits are controlled by several small effect genes, the different methods perform similarly [50–52]. On the contrary, BayesB and BayesCπ are better when dealing with distant relatives and traits affected by a small number of large-effect loci [23]. Kernel methods such as RKHS are robust in predicting non-additive effects and in solving complex multi-environment multi-trait models [53,54]. Compared to the above-mentioned parametric methods, deep learning techniques such as support vector regression (SVR), multilayer perceptron, and convolutional neural networks models performed poorly in some studies [55,56]. However, there are also instances where RKHS outperformed one or several of the parametric methods, for instance, GBLUP, rrBLUP, and Bayesian algorithms, in terms of several traits in several crops including maize [51,57–59]. These results were most likely because nonparametric GS models capture more adequately the non-additive genetic components which are an essential characteristic of complex traits [23,37,38] and hence could be good candidate tools for the prediction of FAW and MW-resistance traits which are controlled by both additive and non-additive gene action [21,23,31,41,60]. Therefore, since GP for maize resistance to insect pests such as FAW and MW is not yet well explored, it is pivotal to compare performances of several available prediction algorithms to inform better future GS programs. Therefore, the Genomic Prediction 0.2.6 plugin of the KDCompute 1.5.2. beta (https://kdcompute.igss-africa.org/kdcompute/home), an online database developed by Diversity Array Technologies (DArT, https://www.diversityarrays.com) for the analysis of DArT marker data, presents great interest for this purpose. It hosts a suite of parametric, semiparametric, and nonparametric GP methods that can be run simultaneously on genotype-phenotype datasets.

Additional factors that influence PAs are the different sizes of the training sets (TSs) and breeding sets (BS) and their genetic relationships, the number of markers used to estimate genomic estimated breeding values (GEBV) of lines, the population structure, and the extent of linkage disequilibrium [21,23,31,41,60]. Since phenotyping is the current bottleneck in plant breeding and one of the disadvantages of GP is the requirement of large TSs for high PAs to be achieved, determination of effective TS composition and size is critical for effective implementation of GS in crop improvement programs [21,61–64]. Additionally, the best TS determination will depend on the genetic architecture and the extent of population structure of the trait targeted for GP [63], two parameters that are substantially variable among plant breeding traits. Another factor that is a determinant of the predictive ability is the kinship between the TS and the BS (63). Several methods are used for TS optimization and these generally fall into two categories—namely, untargeted and targeted approaches. For the untargeted approach, the TS is determined independently of its genomic information, whereas the targeted method considers the genomic relationship between the TS and the BS as a means of maximizing PAs [65]. However, deciding on the best TS selection method is not straightforward and depends on context [66].

Furthermore, in maize, GPs were previously conducted using either genotypic best linear unbiased estimators (BLUEs) [67–69] or best linear unbiased predictors (BLUPs) [31, 41,70] as means of phenotypic correction [70]. BLUEs are obtained by treating the genotypic effect of a mixed linear model as fixed effects and provide an estimated mean for each individual of a population equal to its true value. On the other hand, BLUPs generated by considering the genotypic factors as random and allowing for the shrinkage of the means towards the population mean [71]. Whether to use BLUPs or BLUEs in GP is debatable. Phenotypic BLUEs allow avoiding double penalization which BLUPs suffer from. With phenotypic BLUPs, this double penalization is, however, compensated through maximization of the correlation between predicted and true line values, while phenotypic BLUEs do not rely on this shrinkage [70]. However, the shrinkage in the BLUP procedure accounts better for outliers and environmental variabilities [72], permitting better estimates of individual genetic effects than BLUEs [71], and therefore, it usually yields more accurate predictions of phenotypic performance [70,72,73]. Furthermore, BLUPs are better in handling unbalanced data, wherein, for example, the number of individuals is not the

same in different locations or in the different replications of an experiment [49,70]. On that basis, the current study was conducted to evaluate the efficacies of different parametric, semiparametric, and nonparametric methods from both statistical and machine learning GP models in generating prediction accuracies (PAs) for maize resistance to FAW and MW in a diverse panel using both genotypic BLUEs and BLUPs.

2. Material and Methods

2.1. Genetic Material and Field Experiments

The panel used in this study consisted of 358 maize lines with diverse genetic and geographic backgrounds, and they were sourced from the National Crop Resources Research Institute (NaCRRI/Namulonge, Uganda), the International Institute for Tropical Agriculture (IITA/Ibadan, Nigeria), and The International Maize and Wheat Improvement Center (CIMMYT/Nairobi, Kenya). The panel consisted of 71 inbred lines developed for various purposes at NaCRRI; 28 and five stem borer (SB)-resistant inbred lines from CIM-MYT [6,13,14] and IITA, respectively; 19 storage pest (SP)-resistant inbred lines [7,8]; and a doubled haploid (DH) panel of 235 lines developed at CIMMYT using six parents—three of which were stem borer-resistant, one was a storage pest-resistant inbred line (these were also included in the population), and two were CML elite lines (one, CML132 was included in the panel) (Supplementary Materials Table S1).

The panel was planted and evaluated in three environments, at Mubuku Irrigation Experimental Station in Kasese, western Uganda in 2017 (316 lines) during the second rainy season (2017B) and the National Crop Resources Research Institute (NaCRRI), Namulonge, central Uganda in 2018 (92 lines) and 2019 (252 lines) both during the first rainy seasons (2018A and 2019A, respectively). Detailed information on these locations is presented in Table 1.

Table 1. Geographical, climatic, and soil characteristics of the planting locations [74].

Locations	Geographical Coordinates	Altitude above Sea Level	Minimum Rainfall	Soil Characteristics
Kasese	0°16′10″ N 30°6′9″ E	1330 m asl	1000 mm	Sandy loam soils with a pH of 5.68
Namulonge	0°31′30″ N 32°36′54″ E	1160 m asl	1300 mm	Oxisols with a pH of 5.8

Each combination of location and season was considered an environment, resulting in a total of three environments. An augmented experimental design was adopted in all three environments using six checks in 2017B, two in 2018A, and four in 2019A replicated in all the blocks. The experiments in 2017B, 2018A, and 2019A consisted of twelve, five, and ten blocks, respectively, containing the replicated checks and unreplicated lines and the experiment in 2018A was replicated twice.

2.2. Genotyping, Quality Control, and Imputation for Genomic Prediction Analyses

Genotyping of the panel and SNP quality were described in our previous study [25]. In brief, maize leaves at the sixth-leaf stage of development were harvested from 341 of the 358 lines of the panel (5–10 plants per line) in 2017B and 2018A (for lines not captured in 2017B). The leaf samples were oven-dried overnight at 36 degrees Celsius and shipped to the Biosciences east and central Africa (BecA) Laboratory of the International Livestock Research Institute (ILRI, Nairobi, Kenya) for DNA extraction and genotyping. Diversity Array Technology (DArT) genotyping facilities (44) were used to successfully identify 34,509 SNPs from 341 of the 358 lines composing the panel; hence, only these lines were considered for the GP analyses. Duplicate SNPs were first removed using the R package DartR (45), leaving 28,919 unique SNPs (DRSNP) distributed across all the 10 chromosomes of the entire maize genome.

The DRSNP dataset was imputed before GP using KDCompute 1.5.2. beta (https://kdcompute.igss-africa.org/kdcompute/home), an online database developed by Diversity Array Technologies (DArT, https://www.diversityarrays.com) for the analysis of DArT marker data. KDCompute uses a suit of imputation methods to impute the SNP dataset and scores the imputation results by calculating simple matching coefficient (SMC). The method with the highest SMC is considered as optimal and used to impute the original genotypic dataset.

2.3. FAW and MW Resistance Phenotyping

After germination, plants were left unprotected to allow sufficient natural pressure of fall armyworm (FAW) population to build up. FAW damage scoring in all the three environments was carried out two months after planting when adequate natural FAW infestation levels had manifested, and scoring was based on a visual assessment using a scale of 1 (no or minor leaf damage) to 9 (all leaves highly damaged) [75], illustrated in Figure S1 [18].

Rearing of and bioassays for MW were performed as described in previous experiments carried out at NaCRRI [76,77]. Weevils were reared prior to the MW bioassay to obtain enough insects aged between 0 to 7 days for infestation. During rearing, standard conditions were provided for weevils to ensure proper acclimatization during the experiment. Rearing was carried out by preparing a weevil-maize grain culture of 300–400 unsexed insects and 1.5 kg of grains contained in 3000 cm^3 plastic jars incubated for 14 days in the laboratory at a temperature of 28 ± 2 °C and relative humidity of 70% ± 5%, to enhance oviposition. The lids of the jars were perforated and a gauze-wire mesh with a pore size smaller than one mm was fitted on each lid to allow proper ventilation while preventing the weevils from escaping.

After harvesting and shelling, grains of each line were bulked across environments. Then, samples of 30 g were weighed from each grain bulk, aiming to produce three replicates per line for the MW bioassay experiment. However, due to the lack of an adequate amount of grains for most of the lines of the panel, only 64, 123, and 132 lines could generate three, two, and one replicates, respectively, and were therefore considered for the MW bioassay experiment. Each of these samples was wrapped in polythene bags and kept at −20 °C for 14 days to eliminate any weevil infestation prior to the start of the experiment. After this disinfestation process, samples were left to thaw and transferred into 250 cm^3 glass jars and infested with 32 unsexed weevils. After 10-days of incubation to allow oviposition, all dead and living adult insects were removed. One month after infestation (MAI), each sample was removed from its jar, and the grains and the flour were isolated and their weights were recorded. The total number of holes inflicted by the weevils on the grains was counted along with the number of grains affected by such damages. Additionally, the numbers of dead and living weevils were recorded. After these measurements were collected, the grains were returned to their respective jars and all the measurements were repeated at two and three MAI. The collected data were used to infer, for each sample, the cumulative grain weight loss (GWL), the cumulative number of emerged adult weevil progenies (AP), and the final number of damage-affected kernels (AK).

2.4. Statistical Analyses of the Phenotypic Data

Both best linear unbiased estimators (BLUEs) and predictors (BLUPs) were generated using the general linear model with only phenotype option of the software Trait Association through Evolution and Linkage (TASSEL) [78] and the *ranef* function of the R package [79] lme4 [80], respectively. The mixed linear model for generating BLUEs (all factors considered as fixed) and BLUPs (all factors considered as random) for MW traits (GWL, AP, AK, NH, and FP) was as follows:

$$Y = \mu + Replication + Genotype + Error$$

The mixed model for generating BLUEs (all factors considered as fixed) and BLUPs (with all factors considered as random) for FAW damage scores across environments model was:

$$Y = \mu + Location + Block + Genotype + Location : Genotype + Error$$

where μ in the two equations is the intercept.

2.5. Strategies for TS and BS Determination

2.5.1. MW Resistance Traits

Due to inadequate amount of seeds, only 37% (126 out of 341 that had genotypic data) of lines from the panel had phenotypic data on grain weight loss (GWL), adult progeny emergence (AP), and number of affected kernels (AK). Therefore, to estimate GP accuracies for MW resistance, these 126 lines were used as the TS and the remaining 215 lines with only genotypic data constituted the breeding set (BS).

2.5.2. FAW Damage Resistance

The GP analyses for FAW resistance were carried out on the 341 lines of the panel that were genotyped and phenotyped for FAW damage resistance. To determine TS and BS sizes and compositions for the evaluation of maize resistance to FAW damage, two strategies, namely, random-based TS (RBTS) and pedigree-based TS (PBTS), were used.

2.5.3. Random-Based TS Determination

For the RBTS, 126 (37%) lines used for GPs of MW-resistance traits were used as the TS for FAW to predict the GEBVs of the remaining 215 lines first. To build the second TS for FAW, the 215 (63%) lines used earlier as BS were considered as a TS. Then to determine the third and fourth TSs for FAW, random selections of 75 and 85% of the lines in the entire panel were performed through the Excel formula "*=INDEX($A:$A,RANDBETWEEN(1,COUNTA ($A:$A)),1)*" and dragging until the adequate number of lines for each percentage determined above was obtained.

2.5.4. Pedigree-Based TS Determination

The four datasets determined based on the pedigrees of the lines in the panel (PBTS strategy) are presented in Table 2. For the first dataset (FAW.Ped1), the 235 (68.91%) CIMMYT doubled haploid (DH) lines were used as a TS and the remainder (106 lines) as a BS. Regarding the second dataset, the TS and BS were switched to consider the TS in FAW.Ped1 as BS, and BS in FAW.Ped1 as the TS. The third dataset, FAW.Ped3, had a TS composed of the 294 that were neither stem borer (SB) resistant nor storage pest (SP)-resistant lines from CIMMYT, whilst the 28 SB and 19 SP-resistant lines from CIMMYT constituted the BS. In the last dataset, FAW.Ped4, the 235 DH lines, the 28 SB and 19 SP-resistant lines from CIMMYT, and the five SB-resistant lines from IITA amounting to 287 (84.16%) genotypes were considered as the TS and the remaining 54 lines from NaCRRI lines were considered as the BS (Table 2).

Table 2. Compositions of the pedigree-based test sets (TSs) for fall armyworm (FAW) datasets.

FAW Datasets	FAW.Ped1	FAW.Ped2	FAW.Ped3	FAW.Ped4
TS composition	235 DH CIMMYT lines	106 Non-DH lines	294 Non-CIMMYT SB and SP resistant lines	287 DH and CIMMYT and IITA SB and SP lines
TS/Panel (%)	68.91	31.09	86.22	84.16

DH = doubled haploid; FAW, fall armyworm; FAW.Ped1 to 4, FAW datasets 1–4 with TS based on pedigree information of the lines in the panel; SB, stem borer; SP, storage pest; TS, training set; CIMMYT, International Center for Maize and Wheat Improvement; IITA, International Institute for Tropical Agriculture.

2.6. Genomic Prediction Algorithms

The GP analyses were performed using the BLUEs and BLUPs of the phenotypes and the 28,919 DRSNPs. Sixteen algorithms available in 10 GP methods were implemented using the Genomic Prediction 0.2.6 plugin of the KDCompute 1.5.2. beta. The 10 methods were directly translated from functions of five R packages designed for GP analyses:

2.6.1. Bayesian Models

Bayesian models have different prior distributions with a general model that can be as follows: $y = 1_n\mu + Z\mu + \varepsilon$, where y is the vector of observations, Z is the design matrix for random effects, and μ is the vector of random effects [31].

The BLR (Bayesian Linear Regression) algorithms from the BLR R Package [81] are used to fit the Bayesian ridge regression. The marker effects are assumed to have a Gaussian prior distribution with mean 0 and variance σ^2, where σ^2 is unknown and assumed to have scaled x^2 distribution. In the KDCompute genomic prediction 0.2.6 plugin, the Gibbs sampler is run with 4000 iterations and 1000 iterations for burn-in period as default parameters.

The Bayesian Generalized Linear Regression (BGLR) package fits various types of parametric and semi-parametric Bayesian regressions. The parametric Bayesian algorithms used from this package rely on different prior distributions that induce different types of shrinkages of the marker effects [82], including: Gaussian (Bayesian ridge regression, BRR [83]), scaled-t (BayesA [26]), double-exponential (Bayesian LASSO, BL [84]), and two component mixtures with a point of mass at zero and a distribution with a slab that can be either Gaussian (BayesC [44]) or scaled-t (BayesB [43]). In the KDCompute genomic prediction 0.2.6 plugin tool, defaults parameters for running the Gibbs sampler were used: 4000 iterations and 1000 iterations for burn-in period.

Reproducing kernel Hilbert space (RKHS) [47,48] is a semiparametric Bayesian method from the BGLR package implemented on the KDCompute genomic prediction 0.2.6 plugin. The RKHS methods employs a kernel function to convert the molecular markers as a between pairs of observations distances, thereby, generating a square matrix that fits in a linear model. This non-linear regression method is expected to capture dominance and epistasis effects more efficiently. This approach can be modelled as:

$$y = W\mu + K_h\alpha + \varepsilon,$$

where μ represents the fixed effects vector and ε is a vector of random residuals. The parameters α and ε are assumed to have independent prior distributions $\alpha \sim N(0, K_{h\sigma_\alpha^2})$ and $\varepsilon \sim N(0, I_{\sigma_\varepsilon^2})$, respectively, and the matrix K_h relies on a reproducing kernel function with a smoothing parameter h. The parameter h measures the genomic distances among genotypes that can be interpreted as a correlation matrix and it controls the rate of decay of the correlation among genotypes [51]. To perform this analysis, the same number of iterations and burn-in parameters as for the other Bayesian methods described above were set on the KDCompute genomic prediction 0.2.6 plugin.

2.6.2. Mixed Models

The Sommer (solving mixed model equations in R) package [85] was used to implement the *mmer* (mixed model equations in R) function on the KDCompute genomic prediction 0.2.6 plugin. The package solves mixed model equations proposed by Henderson [86]. It works incidence matrices and known variance covariance matrices for each random effect using four algorithms: efficient mixed model association (EMMA) [87], average information (AI) [88], expectation maximization (EM) [89], and the default Newton–Raphson (NR) [90].

The model by Sommer can be formulated as [85]: $y = X\beta + Z\mu + \varepsilon$ with variance $V(y) = V(Z\mu + \varepsilon) = ZGZ' + R$ Additionally, the mixed model equations for this model are:

$$\begin{bmatrix} X'R^{-1}X \ X'R^{-1}Z \\ Z'R^{-1}X \ Z'R^{-1}Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} = \begin{bmatrix} \beta \\ \mu \end{bmatrix}$$

where $G = K\sigma_\omega^2$ is the variance covariance matrix of the random effect μ, from a multivariate normal distribution $\mu \sim MVN(0, K\sigma_\mu^2)$, **K** is the additive or genomic relationship matrix (**A** or **A$_g$**) in the genomics context, **X** and **Z** are incidence matrices for fixed and random effects, respectively, and **R** is the matrix for residuals (here $I\sigma_e^2$). A mixed model with a single variance component other than the error (σ_e^2) can be used to estimate the genetic variance (σ_μ^2) along with genotype BLUPs to exploit the genetic relationships between individuals coded in **K(A)**. The genomic relationship matrix was constructed according to VanRaden where $K = ZZ'/2\sum p_i(1 - p_i)$ [91].

The ridge regression best linear unbiased predictor (rrBLUP) packages can either estimate marker effects by ridge regression, or alternatively, BLUPs can be calculated based on an additive relationship matrix or a Gaussian kernel. Additionally, using the rrBLUP package, **the mixed model solution (MMS)** that calculates the maximum-likelihood (ML) or restricted-ML (REML) solutions for mixed models to perform GP [92] was fitted in the KDCompute genomic prediction 0.2.6 plugin.

The mixed models fitted by rrBLUP can be formulated as:

$$y = X\beta + Z\mu + \varepsilon,$$

where β is a vector of fixed effects and μ is a vector of random effects with variance $Var[\mu] = K_{\sigma_\mu^2}$. The residual variance is $Var[\varepsilon] = I_{\sigma_\varepsilon^2}$. This class of mixed models, in which there is a single variance component other than the residual error, has a close relationship with ridge regression (ridge parameter $\lambda = \sigma_\varepsilon^2/\sigma_\mu^2$) (https://kdcompute.igss-africa.org/kdcompute/home).

2.6.3. Machine Learning Algorithms

The R package RandomForest that implements Breiman's random forest algorithm for classification and regression [93] was used on the KDCompute genomic prediction 0.2.6 plugin to fit the function missForest. Random forest is a non-linear machine learning algorithm that uses a two-layer randomization process to build decorrelated bootstrapped trees. As a first randomization layer, it builds multiple trees using a bootstrap sample of the marker data in the training. Then, a second randomization process is carried at the novel nodes to grow final trees. The random forest method selects at each node of each tree, a random subset of variables, and only those variables are used as candidates to find the best split for the node [94]. To predict the breeding value of a line in the TS, predictions over trees for which the given observation was not used to build the tree are averaged [51]. On the KDCompute 1.5.2. beta platform, both options for the *mtry*, square root and regression (*sqrt(p)* and *p/3*, respectively, where *p* is number of variables in x), for the classification of the number of variables randomly sampled as candidates at each split were implemented in this study. Additionally, the trees to grow (*ntree*) was set to 10, while *node size* (minimum size of terminal nodes) and *max nodes* (maximum number of terminal nodes trees in the forest can have) were set to 5 and 10, respectively. The 16 methods used in this study and their statistical characteristics are presented in Table 3.

Table 3. Genomic prediction methods used for the analysis of the different traits and datasets.

	GP Algorithms	Abbreviations	Method Type
1	Sommer with Average Information (AI)	mmer_AI	Parametric/Mixed model
2	Sommer with Expectation Maximization (EM)	mmer_EM	Parametric/Mixed model
3	Sommer with Efficient Mixed Model Association (EMMA)	mmer_EMMA	Parametric/Mixed model
4	Sommer with default Newton-Raphson (NR)	mmer-NR	Parametric/Mixed model
5	Ridge-regression Best linear unbiased Predictor	rrBLUP	Parametric/Mixed model
6	Mixed Model solution with Maximum Likelihood (ML)	mms_ML	Parametric/Mixed model
7	Mixed Model solution with Restricted Maximum Likelihood (REML)	mms_REML	Parametric/Mixed model
8	BayesB	BayesB	Parametric/Bayesian
9	BayesA	BayesA	Parametric/Bayesian
10	BayesC	BayesC	Parametric/Bayesian
11	Bayesian least absolute shrinkage and selection operator (LASSO)	BL	Parametric/Bayesian
12	Bayesian Ridge Regression	BRR	Parametric/Bayesian
13	Bayesian Linear Regression	BLR	Parametric/Bayesian
14	Reproducible kernel Hilbert space	RKHS	Semi-parametric/Bayesian
15	Random Forest with Square root	missForest_Sqt	Nonparametric/Machine Learning
16	Random Forest with Regression	missForest_Reg	Nonparametric/Machine Learning

2.7. Cross-Validations and PA Estimation

To calculate the predictive accuracies of each of the 17 methods, a cross-validation approach was performed using the data for the TS with 10 folds and five repetitions amounting to 50 replications. The PAs were estimated as the correlation coefficient (R^2) averaged across the 50 cross-validation replications between the observed phenotypic values and the predicted genomic-estimated breeding values (GEBV) (https://kdcompute. igss-africa.org/kdcompute/plugins).

3. Results

3.1. Higher PAs Achieved for FAW and MW-Resistance Traits with BLUPs when Compared to BLUEs across GP Algorithms

Both genotypic BLUEs and BLUPs for resistance to FAW and MW traits such as AK, AP, and GWL were used in GPs. In general, BLUPs produced better predictions than BLUEs by at least two orders of magnitude in terms of PAs (Figure 1). The PAs realized with BLUEs (Figure S2) varied from −0.246 for FAW (mms_ML) to 0.299 for AP (BayesB), while PAs for BLUPs ranged from 0.668 for AP (mmer_NR) to 0.823 for AP (missForest_Reg). The differences in terms of accuracies between BLUEs and BLUPs were high, despite the highly significant ($p < 0.001$) correlations between BLUEs and BLUPs for each trait ranging from 0.93 for FAW to close to 1 for AP, AK, and GWL (presented in Figure 1); therefore, only results for BLUPs will be presented hereafter.
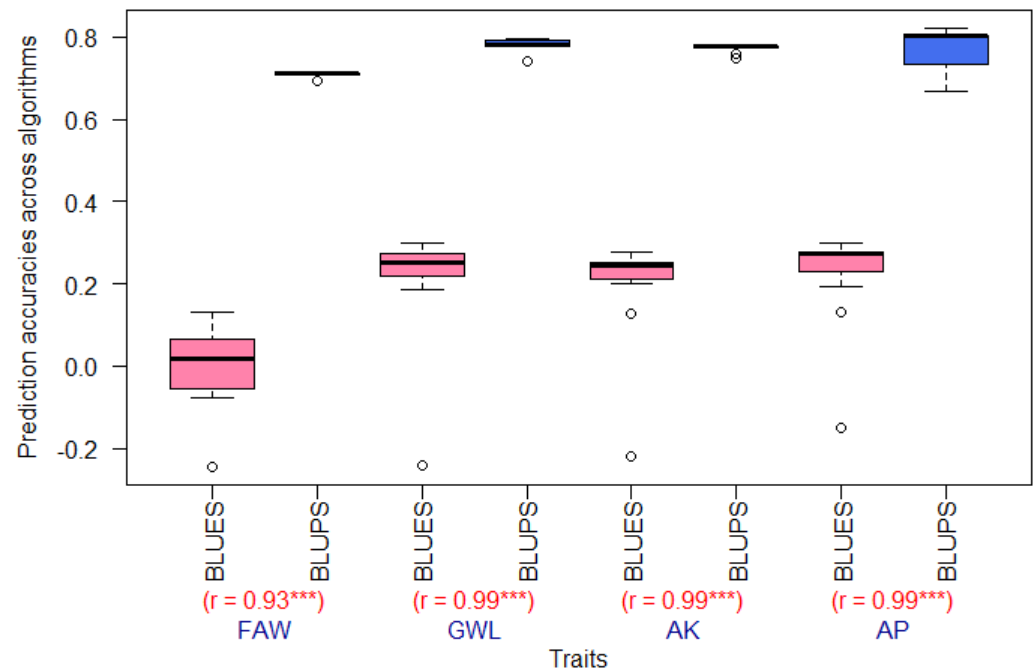
Figure 1. Boxplot of PAs (prediction accuracies) for best linear unbiased estimators (BLUEs) (in pink) and predictors (BLUPs) (in blue) of maize resistance to FAW and MW across prediction models and correlations (r) between BLUEs and BLUPs for each trait. FAW, fall armyworm; GWL, grain weight loss; AP, adult progeny emergence; AK, number of affected kernels. *** significant at $p < 0.001$.

3.2. PAs for MW Resistance Traits Using BLUPs

The PAs were generally high for the tested MW traits, mostly above 0.668 across the 12 GP models that were successfully run on the datasets (Figure 2); however, RKHS failed to work for AK. The highest PAs were achieved for AP with missForest_reg (0.823), followed by BRR (0.805), and RKHS (0.804), whilst mmer_NR algorithm had the lowest prediction accuracy of 0.667 (Figure 2). The PAs achieved for GWL ranged from 0.742 for missForest_Sqt to 0.795 for mmer_NR, while for AK, they varied from 0.749 for missForest_sqrt to 0.779 for BRR (Figure 2). In general, Bayesian models predicted better than both mixed model and machine learning methods, although the differences were small (Figure S3).
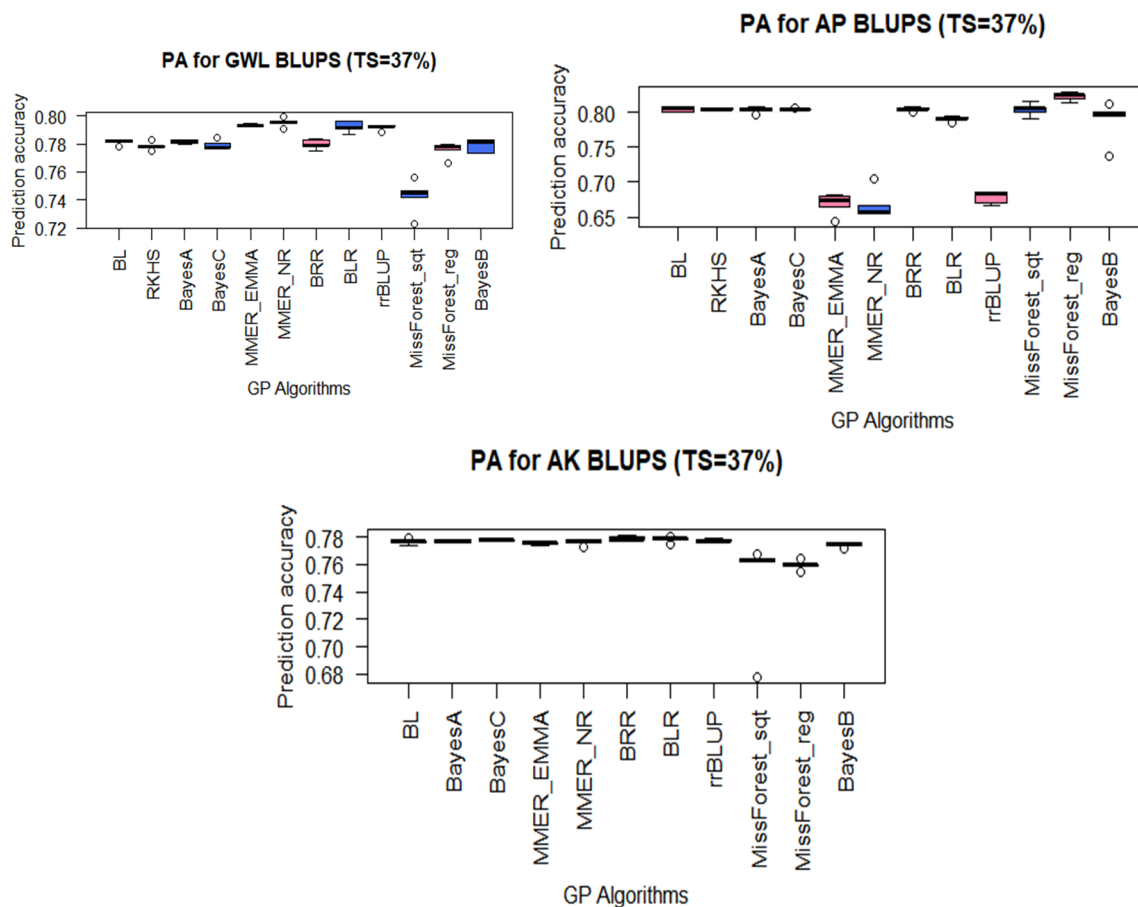
Figure 2. Boxplots of the genomic prediction accuracies of BLUPs for MW-resistance traits: GWL, grain weight loss; AP, adult progeny emergence; AK, number of affected kernels (See Table 3 for GP algorithms).

3.3. PA for FAW Resistance Using BLUPs

The different maize resistance to FAW datasets showed high predictive abilities with 10 of the 16 GP algorithms used in the study. For the RBTS approach, the PAs were lowest with the dataset that had a TS composed of 37% (lowest size) of the panel and highest with the largest TS (85% of the panel). Even with a TS of 37%, the PAs were still high, ranging from 0.694 to 0.714 for mms_ML and BLR methods, respectively (Figure 3). However, it should be noted that with equal TS sizes and same composition (37% of the panel), higher PAs were achieved for MW-resistance traits (GWL, AP, and AK) compared to FAW-resistance ones (Figure S3). The PA for the RBTS of 63% varied from 0.833 for BL method to 0.838 for the missForest_Sqt; thus, there was a small variation among different methods. Similarly, there was minimal variation among GP algorithms on the dataset with a 75% TS whose PAs varied from 0.838 for mms_REML to 0.843 for MissForest_Reg. The same trend was obtained on the dataset with a RBTS of 85% of the panel, with PAs ranging from 0.843 for the BRR model to 0.847 for the missForest_Reg method. Furthermore, there was a high and significant ($p < 2.2.10^{-16}$) positive correlation of 0.92 (Figure 4) between the PAs and TS sizes for FAW datasets for the RBTS denoting a steady improvement of the PAs as the TS size increased. However, the PAs for FAW resistance reached a plateau at TS size above 63% of the panel (Figure 5).
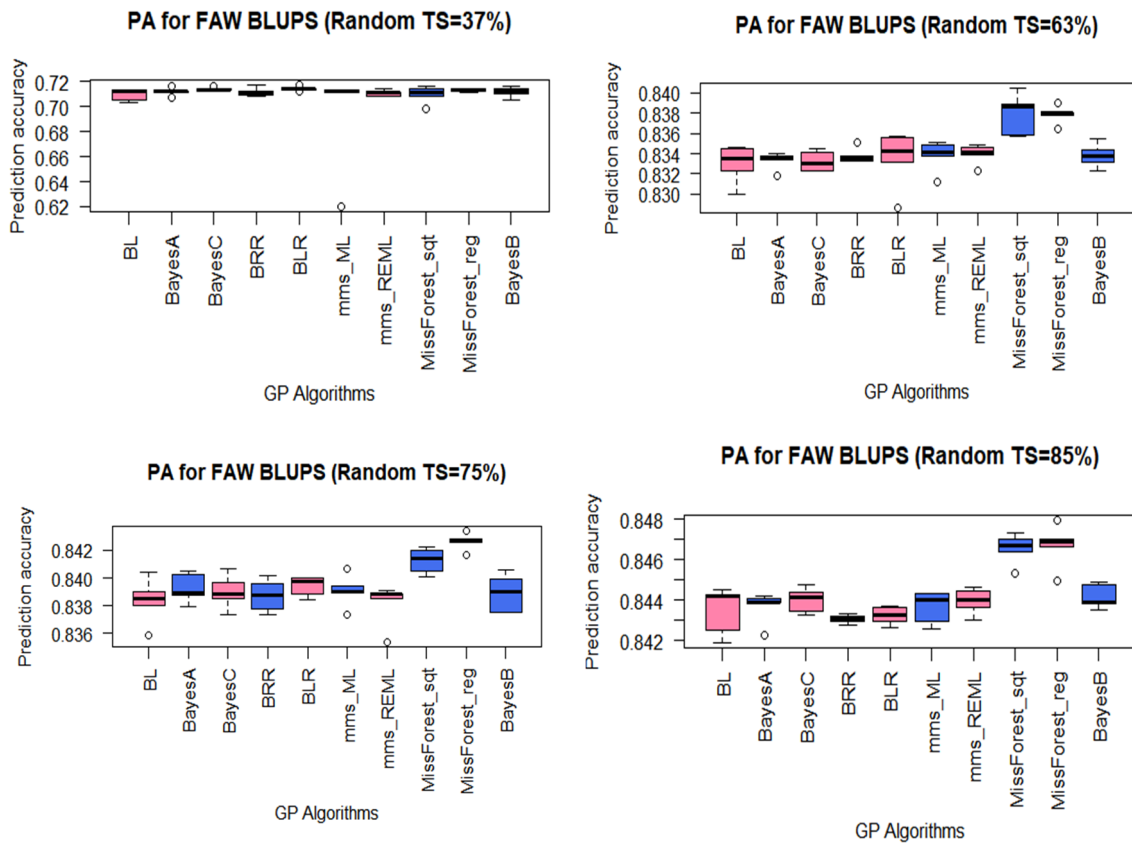
Figure 3. Boxplot of PAs for maize resistance to the fall armyworm (FAW) datasets with the RBTS approach with random selection of 37, 63, 75, and 87% of the entire panel (see Table 3 for GP algorithms).
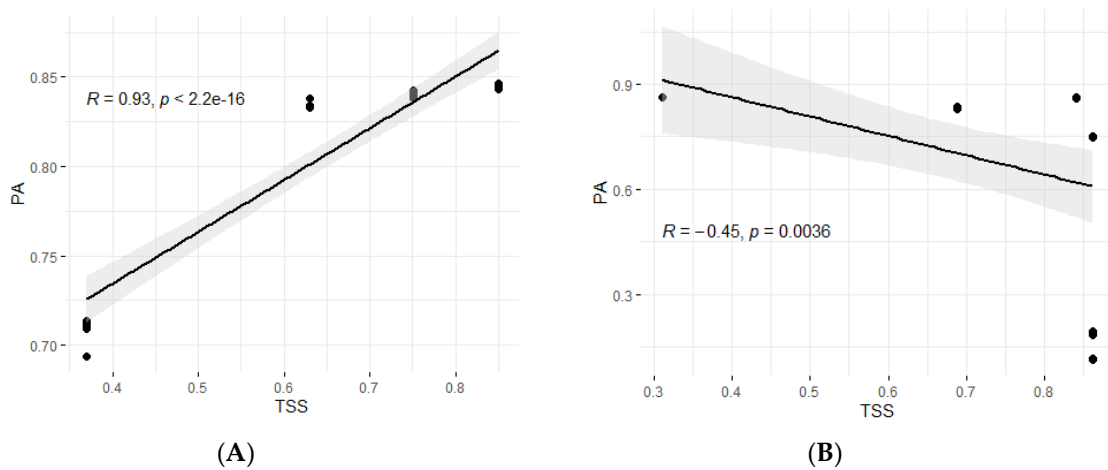

Figure 4. Pearson correlation between training set (TS) sizes and prediction accuracies (PAs) across the 10 genomic prediction algorithms conducted on RBTS (**A**) and PBTS (**B**) datasets for fall armyworm resistance (FAW) resistance.
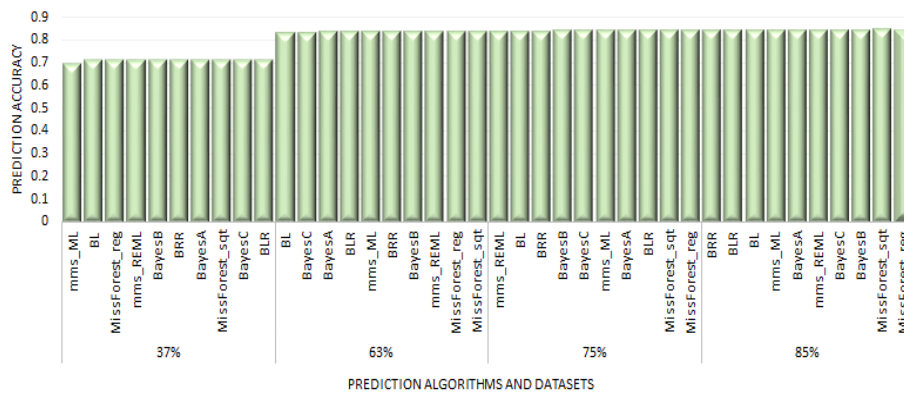
Figure 5. Prediction accuracies for FAW with RBTS across algorithms and training sets with different sizes in percent of the total panel.

Although the PAs did not vary much among GP algorithms, especially when the analyses involved larger TS sizes equal or bigger than 63% of the panel, the machine learning methods slightly outperformed other GP algorithms for all the traits, except for the TS of 37% where Bayesian methods such as BLR and BayesC showed a slight advantage over the machine learning methods (Figure S4). The PAs for FAW-resistance datasets with PBTS were generally high, mostly above 0.82 (Figure 6). For the first dataset (FAW.Ped1) with a TS of 68.91% of the panel (see Table 2), the PAs varied between 0.828 for BLR to 0.835 for missForest_Sqt. For FAW.Ped2 (TS = 31.09%), the PAs ranged from 0.862 for BayesC to 0.864 for mms_REML.
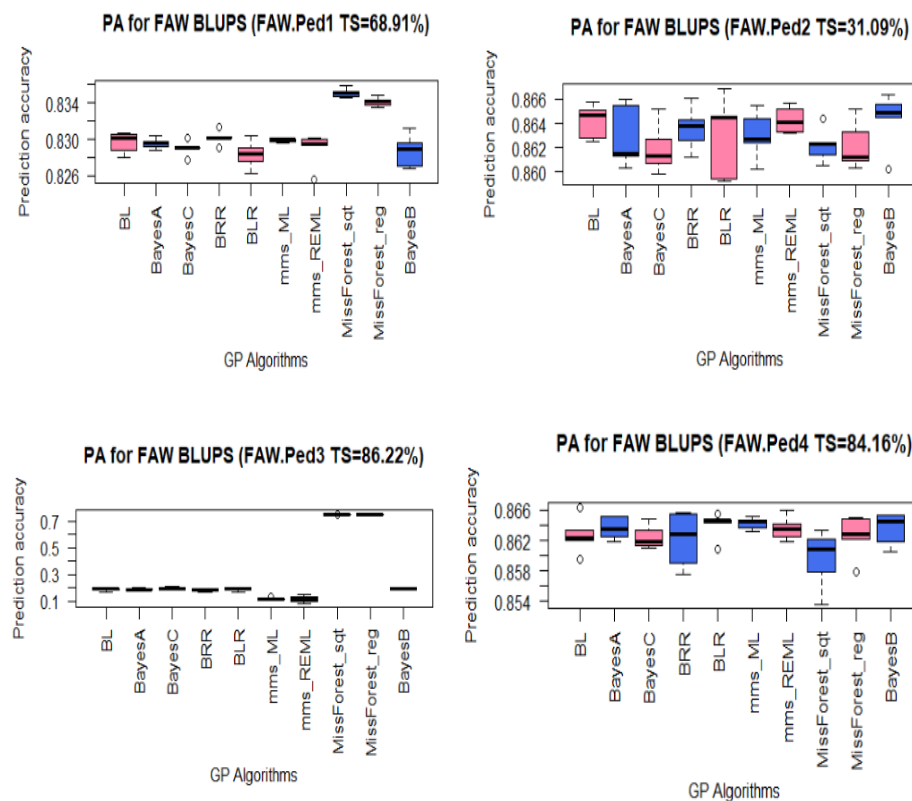
Figure 6. Boxplots of PAs for maize resistance to the fall armyworm (FAW) datasets using the PBTS approach (see Table 2 for the PBTS strategy and Table 3 for GP algorithms).

For FAW.Ped4, with a TS of 84.16%, PAs varied between 0.860 to 0.864 for missForest_Sqt and mms_ML, respectively. However, for FAW.Ped3 with the largest TS (86.22%), eight of the 10 algorithms achieved low PAs (below 0.20) and only missForest_Reg and missForest_Sqt attained PAs of 0.749 and 0.750, respectively. Thus, the Pearson correlation between the sizes of the PBTS datasets and the predictions accuracies for the 10 GP algorithms revealed a significant ($p > 0.0036$) negative relationship of r = -0.45 (Figure 4).

In the FAW datasets, the PAs were more influenced by the composition of the TS and its genetic relationship with the BS (see Table 2). Using the doubled haploid (DH) lines as TS (FAW.Ped1) and vice-versa (FAW.Ped2) or DH and stem borer (SB) and storage pest (SP)-resistant lines as TS (FAW.Ped4) permitted achieving relatively high PAs from all the 10 algorithms, which when considering the CIMMYT SB and SP-resistant lines as BS and the remainder as a TS (FAW.Ped3), only resulted in machine learning algorithms missForest_reg and missForest_Sqt achieving relatively high PAs. Furthermore, the composition of the TS and its relationship with the BS determined which GP methods achieved the highest Pas; machine learning algorithms worked best on FAW.Ped1 and FAW.Ped3, linear mixed model approaches outperformed Bayesian and machine learning algorithms on FAW.Ped2 and FAW.Ped4, and Bayesian methods ranked either second or third on all datasets (Figure S5). It should be noted that the PBTS strategy generally achieved better PAs than the RBTS irrespective of the size of the TS, except for the FAW.Ped3 dataset (Figures 3 and 6).

4. Discussion

Tropical maize germplasm is characterized by rapid linkage disequilibrium (LD) decay with high diversity [95]. These germplasm genetic characteristics make genomic selection (GS) a promising approach to integrate into African breeding programs [96]. However, genomic prediction (GP) models are very diverse and their differential performances depend on crops and trait architectures, besides other parameters such as the size of the training set (TS) and its genetic relationship with the breeding set (BS) [31,37]. Therefore, this study aimed at assessing the feasibility of genomic selection for maize resistance to FAW and MW through estimation of the genomic prediction accuracies achieved by parametric, semiparametric, and nonparametric (machine learning) genomic prediction (GP) algorithms using phenotypic BLUEs and BLUPs, and random and pedigree-based TS determination strategies.

4.1. Higher Pas Were Achieved for BLUPs Compared to BLUEs for Both FAW and MW-resistance Traits

With a RBTS of 37% of the panel, which was the smallest and expected to give the worst PAs, PAs were higher (at least two-fold) across both FAW and MW-resistance traits and for all GP models when trait BLUPs were used as phenotypes compared to BLUEs, although there were high Pearson correlations between these two categories of phenotypic data for each trait. In general, BLUPs were reported to have higher predictability than BLUEs owing to better accounting for outliers and environmental variabilities permitted by the shrinkage procedure in BLUPs, which results in more accurate estimates of individual genetic effects [70–73]. Furthermore, most of the predictive differences between BLUPs and BLUEs might have stemmed from BLUPs being more suitable than BLUEs in fitting data recorded from unbalanced experiments [49,70] as was the case for both FAW damage scores across environments and MW bioassay in this study. Therefore, for all subsequent analyses with higher RBTS sizes and the PBTS strategy for FAW, only BLUPs were focused at in this study and will be further discussed.

4.2. High PAs Were Achieved for FAW and MW-Resistance Traits Using Moderately Sized Training Sets

The obtained PAs were high for both MW and FAW-resistance traits even with TS of moderate sizes confirming the potential of genomic selection (GS) in Africa-adapted germplasms [28–30,33]. With a TS of 37% of the entire panel, high PAs (above 0.70) for MW-resistance traits, grain weight loss (GWL), adult progeny emergence (AP), the number of

affected kernels (AK), and FAW resistance were achieved in agreement with the moderate to high heritability values for these traits as, reported earlier [21,31,41]. These results are significantly important considering that one of the disadvantages of GS is the requirement of large TS which negatively impacts the reduction of phenotyping cost [62,64].

The PAs increased up to above 0.85 in proportion to the increase in TS (RBTS approach) size for FAW resistance which was the only trait phenotyped for all the lines of the panel. It would be interesting to phenotype other lines of the panel that were not evaluated for MW-resistance traits to establish larger TS which may improve the PAs [31,65,97,98]. Very few reports of GP are available for maize resistance to biotic stresses. High PAs were achieved for maize resistance to chlorotic mottle virus (up to 0.95) and maize lethal necrosis (reaching 0.87) in tropical germplasm [67]. However, lower PAs of up to 0.59 were obtained in a study that assessed the predictability of maize resistance to the European corn borer [99] in temperate germplasm. Additionally, Gowda et al. (69) reported moderate PAs (close to 0.60) for maize resistance to a biotic stress, maize lethal necrosis in tropical maize populations.

4.3. GP Algorithms Performed Differently on FAW and MW Maize Resistance Traits

In this study, several GP models that included statistical and machine learning algorithms from parametric, semi-parametric, and nonparametric approaches were used to predict FAW and MW-resistance traits. These GP algorithms, as expected, performed differently on the different traits although the predictive variations were generally minimal, especially when large TS were involved, similarly to earlier model benchmarking reports [100,101]. Bayesian models (parametric: BLR and BRR, and semi-parametric: RKHS) performed better on MW traits, GWL, AP, and AK, while nonparametric machine learning algorithms (missForest, here), and to a lesser extent, the linear mixed model (especially in the PBTS approach), achieved the highest PAs on FAW datasets. The differential performances of the different GP algorithms on the insect resistance traits evaluated in this study could be due to differences in the genetic structures (extent of additive vs. non-additive gene action) of the respective traits [23,38,47,49]. Maize resistance to FAW, which was moderately heritable across environments [25], would be expected to be controlled by both additive and non-additive genetic factors, including epistasis [102–104], whereas, MW-resistance traits such as GWL, AP, and AK with heritability values above 90% [25] were most likely characterized by a prevalence of additive gene action [105,106] in the current panel.

This supposed genetic architecture difference between FAW and MW-resistance trait could be the reason for non-linear methods such as random forest performing better at predicting FAW resistance, since these are more capable of integrating epistasis in the statistical modelling [27,51]. However, the RKHS algorithm, also a non-linear GP approach known to efficiently handle epistatic genetic relation [51,59], did not successfully run on FAW dataset, although it was among the best models for predicting MW-resistance traits, except BLUPs for the number of affected kernels (AK), for which the RKHS algorithm did not run successfully. In this study, the reasons for some GP algorithms failing to run either on MW or FAW-resistance datasets are unclear, but this could be related to the BLUPS structure of the datasets that failed to run. It should be noted that all the algorithms ran successfully on phenotypic BLUEs datasets with the smallest TS (37% of the panel) being used to compare PAs between BLUPs and BLUEs in this study. However, the two to three-fold predictive ability gain with BLUPs compared to BLUEs would be an incentive to consider BLUPs in future GS activities for maize resistance to MW and FAW. Overall, future GS efforts for maize resistance to MW and FAW are recommended to focus more on Bayesian and machine learning algorithms such as random forest, BayesA, BayesB, BayesC, BRR, and BLR which outperformed mixed linear models for most datasets considered in the current study.

4.4. Influences of the Sizes and the Compositions of TS and BS on PAs

Two factors, the relative sizes of the TS and BS (RBTS approach) and their genetic relationship (PBTS approach), influenced the levels of PAs across FAW-resistance datasets, corroborating earlier reports [31,63,65,97,98,107,108]. A net increase in PAs for maize resistance to FAW was realized when the size of the TS was increased from 37% (0.694 to 0.714) to 63% (0.833 to 0.838), similar to earlier reports on wheat yield [109]. This increase was followed by a slight gain in predictability at 75% (0.837 to 0.843) and 85% (0.843 to 0.847), and thus, the PAs plateaued when TS sizes above 63% were considered in this study as reported earlier in other studies [21,64,109–111]. Thus, future GS programs for maize resistance to FAW could be designed around TS composed of a minimum of 60% of the entire breeding germplasm to achieve high genetic gains. These results were further supported by the highly significant ($p > 2.2.10^{-16}$) positive correlation (R = 0.92) between TS size and PAs. Similarly, positive correlations between the number of lines in the TS and the PAs, and plateau for the PAs were also reported by Edwards et al. [109].

The composition of the TS and its relationship with the BS are determinant factors for the genomic predictability of complex traits [63,112–114]. In the current study, using the PBTS approach, these two parameters were more important than the size of the TS since higher PAs were achieved in FAW.Ped2 (0.862 to 0.864) with a TS of 31.09% compared to all other FAW PBTS datasets, including FAW.Ped3 (0.114 to 0.750), with the largest TS of 66.22%. In fact, FAW.Ped3 achieved the lowest PAs among all the PBTS FAW datasets. These results were further illustrated by the significantly ($p < 0.0036$) negative correlation (R = −0.45) between the sizes of the PBTS and the achieved PAs.

However, it is not very clear why the predictions for the BS FAW.Ped3 (47 CIMMYT SB and SP-resistant lines) and the TS (DH, IITA SB, and NaCRRI lines) led to lower PAs for FAW.Ped3. A possible explanation could be that these two sets were distantly related since only two and one CIMMYT SB and SP-resistant lines, respectively, were used as parents to develop the DH lines. Spindel et al. [111] argued that high PAs can be achieved with small-sized TS when lines in the TS and the BS are closely related, since such TS would sample the full genetic diversity of the population. However, the more distantly related the TS and the BS are, the larger the required TS size to reach high PAs [111]. Using the CIMMYT SB and SP-resistant lines as a TS would most likely lead to lower PAs since such a TS would be additionally disadvantaged by its small size (47 lines). The DH lines in the current study are involved as a TS in most of the best performing GP datasets evaluated in the current study (both in the RBTS and PBTS approaches) and as unique lines in the BS of the best performing pedigree-based BS (FAW.Ped2). This DH population could be of interest in future breeding activities targeted at improving insect resistance in maize [23,115–117] and potentially useful for GS of complex traits with low to moderate heritability [118].

5. Conclusions

This study assessed prediction accuracies of genomic-estimated breeding values for fall armyworm (FAW) and maize weevil (MW)-resistance traits in a diverse Africa-adapted maize panel using several parametric, semi-parametric, and non-parametric genomic prediction models. Prediction accuracies for maize resistance to FAW and MW traits were relatively high, even with a moderate training set size. For FAW resistance, although the prediction accuracies were positively correlated with the size of the training set, the composition and the relationship of the training set with the breeding set were more influential in predicting line performance. Additionally, TS determination-related parameters were more important than the type of genomic prediction models in predicting FAW and MW-resistance traits. However, Bayesian models on MW-resistance traits and machine learning models on FAW damage resistance outperformed mixed linear models in almost all the datasets used in this study. Therefore, future genomic selection programs for maize resistance to insect pests such as FAW and MW in Africa should put more effort into designing effective training sets and use selected Bayesian and machine learning GP algorithms to improve genetic gains, shorten breeding cycles, and accelerate variety release.

Such programs could greatly benefit from using the genetically diverse maize panel used in this study as a base population, since it consists of lines adapted to several African agro-ecologies.

References

1. Demissie, G.; Tefera, T.; Tadesse, A. Importance of husk covering on field infestation of maize by *Sitophilus zeamais* Motsch (Coleoptera: Curculionidea) at Bako, Western Ethiopia. *Afr. J. Biotechnol.* **2008**, *7*, 3777–3782.
2. Shiferaw, B.; Prasanna, B.M.; Hellin, J.; Bänziger, M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* **2011**, *3*, 307–327. [CrossRef]
3. Awata, L.A.O.; Tongoona, P.; Danquah, E.; Ifie, B.E.; Suresh, L.M.; Jumbo, M.B.; Marchelo-D'ragga, P.W.; Sitonik, C. Understanding tropical maize (*Zea mays* L.): The major monocot in modernization and sustainability of agriculture in sub-Saharan Africa. *Int. J. Adv. Agric. Res.* **2019**, *7*, 32–77.
4. Nyukuri, R.W.; Wanjala, F.M.; Kirui, S.C.; Cheramgoi, E.; Chirchir, E.; Mwale, R. Damage of stem borer species to Zea mays L.,Sorghum bicolor L. and three refugia graminae. *Adv. Agric. Biol.* **2014**, *1*, 37–45.
5. Tefera, T.; Goftishu, M.; Ba, M.; Rangaswamy, M. *A Guide to Biological Control of Fall Armyworm in Africa Using Egg Parasitoids*, 1st ed.; ICIPE: Nairobi, Kenya, 2019.

6. Munyiri, S.W.; Mugo, S.N.; Otim, M.; Mwololo, J.K.; Okori, P. Mechanisms and sources of resistance in tropical maize inbred lines to *Chilo partellus* stem borers. *J. Agric. Sci.* **2013**, *5*, 51–60. [CrossRef]

7. Mwololo, J.K.; Mugo, S.; Okori, P.; Tefera, T.; Otim, M.; Munyiri, S.W. Sources of resistance to the maize weevil *Sitophilus zeamais* in tropical maize. *J. Agric. Sci.* **2012**, *4*, 206–215.

8. Mwololo, J.K. Resistance in Tropical Maize to the Maize Weevil and Larger Grain Borer. Ph.D. Thesis, Makerere University, Kampala, Uganda, 2013.

9. Kasozi, L.C.; Derera, J.; Tongoona, P.; Tukamuhabwa, P.; Muwonge, A.; Asea, G. Genotypic variation for maize weevil resistance in eastern and southern Africa maize inbred lines. *Uganda J. Agric. Sci.* **2016**, *17*, 83–97. [CrossRef]

10. Tende, R.; Derera, J.; Mugo, S.; Oikeh, S. Estimation of genetic diversity of germplasm used to develop insect-pest resistant maize. *Maydica* **2016**, *61*, 1–8.

11. Khakata, S.; Mbute, F.N.; Chemining'wa, G.N.; Mwimali, M.; Karanja, J.; Harvey, J.; Mwololo, J.K. Post-harvest evaluation of selected inbred lines to maize weevil *Sitophilus zeamais* resistance. *J. Plant Breed. Crop Sci.* **2018**, *10*, 105–114. [CrossRef]

12. Sodedji, F.K.A.; Kwemoi, D.B.; Kasozi, C.L.; Asea, G.; Kyamanywa, S. Genetic analysis for resistance to *Sitophilus zeamais* (Motschulsky) among provitamin-A maize germplasm. *Maydica* **2018**, *63*, 8. Available online: https://journals-crea.4science.it/index.php/maydica/article/view/1698 (accessed on 13 November 2019).

13. Munyiri, W.S.; Mugo, N.S.; Otim, M.; Tefera, T.; Beyene, Y.; Mwololo, K.J.; Okori, P. Responses of tropical maize landraces to damage by *Chilo partellus* stem borer. *Afr. J. Biotechnol.* **2013**, *12*, 1229–1235.

14. Munyiri, S.W.; Mugo, S.N.; Mwololo, J.K. Mechanisms and levels of resistance in hybrids, open pollinated varieties and landraces to *Chilo partellus* maize stem borers. *Int. Res. J. Agric. Sci. Soil Sci.* **2015**, *5*, 81–90.

15. Mwololo, J.K.; Munyiri, S.W.; Semagn, K.; Mugo, S. Genetic diversity analysis in tropical maize germplasm for stem borer and storage pest resistance using molecular markers and phenotypic traits. *Mol. Plant Breed.* **2015**, *6*, 1–22.

16. Goergen, G.; Kumar, P.L.; Sankung, S.B.; Togola, A.; Tamò, M. First report of outbreaks of the fall armyworm *Spodoptera frugiperda* (JE Smith) (Lepidoptera, Noctuidae), a new alien invasive pest in West and Central Africa. *PLoS ONE* **2016**, *11*, e0165632. [CrossRef]

17. Padhee, A.K.; Prassanna, B.M. The emerging threat of Fall Armyworm in India. *Indian Farming* **2019**, *69*, 51–54.

18. Prasanna, B.M.; Huesing, J.E.; Eddy, R.; Peschke, V.M.; Regina, E.; Virginia, M.P. *Fall Armyworm in Africa: A Guide for Integrated Pest Management*, 1st ed.; West Africa Regional Training of Trainers and Awareness Generation Workshop on Fall Armyworm Management, IITA, Cotonou, Bénin; Prasanna, B.M., Regina, E., Virginia, M.P., Eds.; CIMMYT: Queretaro, Mexico, 2018.

19. Gedil, M.; Menkir, A. An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Front. Plant Sci.* **2019**, *10*, 1–17. [CrossRef] [PubMed]

20. Murenga, M.; Derera, J.; Mugo, S.; Tongoona, P. A review of genetic analysis and response to selection for resistance to *Busseola fusca* and *Chilo partellus*, stem borers in tropical maize germplasm: A Kenyan perspective. *Maydica* **2016**, *61*, 1–11.

21. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueno, J.; Gonzales-Camacho, J.M.; Perez-Elizade, S.; Beyene, Y.; et al. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 1–15. [CrossRef]

22. Roorkiwal, M.; Jarquin, D.; Singh, M.K.; Gaur, P.M.; Bharadwaj, C.; Rathore, A.; Howard, R.; Srinivasan, S.; Jain, A.; Garg, V.; et al. Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. *Sci. Rep.* **2018**, *8*, 1–11. [CrossRef]

23. Robertsen, C.D.; Hjortshøj, R.L.; Janss, L.L. Genomic selection in cereal breeding. *Agronomy* **2019**, *9*, 95. [CrossRef]

24. Munyiri, S.W.; Mugo, S.N. Quantitative trait loci for resistance to spotted and African maize stem borers (*Chilo partellus* and *Busseola fusca*) in a tropical maize (*Zea mays* L.) population. *Afr. J. Biotechnol.* **2017**, *16*, 1579–1589. [CrossRef]

25. Badji, A.; Kwemoi, D.B.; Machida, L.; Okii, D.; Mwila, N.; Agbahoungba, S.; Kumi, F.; Ibanda, A.; Bararyenya, A.; Solemangey, M.; et al. Genetic basis of maize resistance to multiple-insect pests: Integrated genome-wide comparative mapping and candidate. *Genes* **2020**, *11*, 689. [CrossRef] [PubMed]

26. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [PubMed]

27. Jannink, J.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom.* **2010**, *9*, 166–177. [CrossRef]

28. Massman, J.M.; Jung, H.J.G.; Bernardo, R. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Sci.* **2013**, *53*, 58–66. [CrossRef]

29. Beyene, Y.; Semagn, K.; Mugo, S.; Tarekegne, A.; Babu, R.; Meisel, B.; Sehabiague, P.; Makumbi, D.; Magorokosho, C.; Oukeh, S.; et al. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Sci.* **2015**, *55*, 154–163. [CrossRef]

30. Vivek, B.S.; Krishna, G.K.; Vengadessan, V.; Babu, R.; Zaidi, P.H.; Kha, L.Q.; Mandal, S.S.; Grudloyma, P.; Takalkar, S.; Krothapalli, K.; et al. Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. *Plant Genome* **2017**, *10*, 1–8. [CrossRef]

31. Liu, X.; Wang, H.; Wang, H.; Guo, Z.; Xu, X.; Liu, J.; Wang, S.; Li, W.-X.; Zou, C.; Prasanna, B.M.; et al. Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J.* **2018**, *6*, 341–352. [CrossRef]

32. Yuan, Y.; Scheben, A.; Batley, J.; Edwards, D. Using genomics to adapt crops to climate change. In *Sustainable Solutions for Food Security*; Springer International Publishing: Cham, Switzerland, 2019; pp. 91–109.

33. Zhang, X.; Pérez-Rodríguez, P.; Burgueño, J.; Olsen, M.; Buckler, E.; Atlin, G.; Prasanna, B.M.; Vargas, M.; San Vicente, F.; Crossa, J. Rapid cycling genomic selection in a multiparental tropical maize population. *G3 Genes Genomes Genet.* **2017**, *7*, 2315–2326. [CrossRef]

34. Hassen, M.B.; Bartholomé, J.; Valè, G.; Cao, T.V.; Ahmadi, N. Genomic prediction accounting for genotype by environment interaction offers an effective framework for breeding simultaneously for adaptation to an abiotic stress and performance under normal cropping conditions in rice. *G3 Genes Genomes Genet.* **2018**, *8*, 2319–2332. [CrossRef]

35. Muleta, K.T.; Pressoir, G.; Morris, G.P. Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. *G3 Genes Genomes Genet.* **2019**, *9*, 391–401. [CrossRef]

36. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2019**, *122*, 370–379. [CrossRef] [PubMed]

37. Voss-Fels, K.P.; Cooper, M.; Hayes, B.J. Accelerating crop genetic gains with genomic selection. *Appl. Genet.* **2019**, *132*, 669–686. [CrossRef] [PubMed]

38. Kadam, D.C.; Lorenz, A.J. Evaluation of nonparametric models for genomic prediction of early-stage single crosses in maize. *Crop Sci.* **2019**, *59*, 1411–1423. [CrossRef]

39. Whittaker, J.C.; Thompson, R.; Denham, M.C. Marker-assisted selection using ridge regression. *Genet. Res.* **2000**, *75*, 249–252. [CrossRef] [PubMed]

40. Clark, S.A.; van der Werf, J. *Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 321–330.

41. Zhang, H.; Yin, L.; Wang, M.; Yuan, X.; Liu, X. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Front. Genet.* **2019**, *10*, 1–10. [CrossRef] [PubMed]

42. Hayashi, T.; Iwata, H. EM algorithm for Bayesian estimation of genomic breeding values. *BMC Genet.* **2010**, *11*, 3. [CrossRef]

43. Meuwissen, T.H.; Solberg, T.R.; Shepherd, R.; Woolliams, J.A. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genet. Sel. Evol.* **2009**, *41*, 1–10. [CrossRef]

44. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef]

45. De Los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A.; Manfredi, E.; Weigel, K.; Cotes, J.M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **2009**, *182*, 375–385. [CrossRef]

46. Zhou, X.; Carbonetto, P.; Stephens, M. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genet.* **2013**, *9*, e1003264. [CrossRef] [PubMed]

47. Gianola, D.; Fernando, R.L.; Stella, A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **2006**, *173*, 1761–1776. [CrossRef] [PubMed]

48. Gianola, D.; de los Campos, G.; González-Recio, O.; Long, N.; Okut, H.; Rosa, G.J.M.; Weigel, K.A.; Wu, X.-L. Statistical learning methods for genome-based analysis of quantitative traits. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, 1–6 August 2010; Gesellschaft für Tierzuchtwissenschaften eV: Neustadt am Rübenberge, Germany, 2010; Volume 14, pp. 1–6.

49. Howard, R.; Carriquiry, A.L.; Beavis, W.D. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3 Genes Genomes Genet.* **2014**, *4*, 1027–1046. [CrossRef] [PubMed]

50. De los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013**, *193*, 327–345. [CrossRef] [PubMed]

51. Heslot, N.; Yang, H.P.; Sorrells, M.E.; Jannink, J.L. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* **2012**, *52*, 146–160. [CrossRef]

52. Maltecca, C.; Parker, K.L.; Cassady, J.P.; Maltecca, C.; Parker, K.L.; Cassady, J.P. Application of multiple shrinkage methods to genomic predictions. *J. Anim. Sci.* **2014**, *90*, 1777–1787. [CrossRef]

53. E Sousa, M.B.; Cuevas, J.; de Oliveira Couto, E.G.; Pérez-Rodríguez, P.; Jarquín, D.; Fritsche-Neto, R.; Burgueno, J.; Crossa, J. Genomic-enabled prediction in maize using Kernel models with genotype × environment interaction. *G3 Genes Genomes Genet.* **2017**, *7*, 1995–2014.

54. Cuevas, J.; Granato, I.; Fritsche-Neto, R.; Montesinos-Lopez, O.A.; Burgueño, J.; e Sousa, M.B.; Crossa, J. Genomic-enabled prediction Kernel models with random intercepts for multi-environment trials. *G3 Genes Genomes Genet.* **2018**, *8*, 1347–1365. [CrossRef]

55. Bellot, P.; de los Campos, G.; Pérez-Enciso, M. Can deep learning improve genomic prediction of complex human traits? *Genetics* **2018**, *210*, 809–819. [CrossRef]

56. Montesinos-López, O.A.; Martín-Vallejo, J.; Crossa, J.; Gianola, D.; Hernández-Suárez, C.M.; Montesinos-López, A.; Juliana, P.; Singh, R. A benchmarking between deep learning, support vector machine and Bayesian threshold best linear unbiased prediction for predicting ordinal traits in plant breeding. *G3 Genes Genomes Genet.* **2019**, *9*, 601–618. [CrossRef]

57. Crossa, J.; Beyene, Y.; Semagn, K.; Pérez, P.; Hickey, J.M.; Chen, C.; de los Campos, G.; Burgueno, J.; Windhausen, V.S.; Buckler, E.; et al. Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 Genes Genomes Genet.* **2013**, *3*, 1903–1926. [CrossRef] [PubMed]

58. Pérez-Rodríguez, P.; Gianola, D.; González-Camacho, J.M.; Crossa, J.; Manès, Y.; Dreisigacker, S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes Genomes Genet.* **2012**, *2*, 1595–1605. [CrossRef] [PubMed]

59. De Andrade, L.R.B.; Bandeira e Sousa, M.; Oliveira, E.J.; de Resende, M.D.V.; Azevedo, C.F. Cassava yield traits predicted by genomic selection methods. *PLoS ONE* **2019**, *14*, e0224920. [CrossRef] [PubMed]

60. Hickey, J.M.; Chiurugwi, T.; Mackay, I.; Powell, W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **2017**, *49*, 1297–1303. [CrossRef] [PubMed]

61. Akdemir, D.; Isidro-sánchez, J. Design of training populations for selective phenotyping in genomic prediction. *Sci. Rep.* **2019**, *9*, 1–15. [CrossRef]

62. Gosal, S.S.; Wani, S.H. (Eds.) *Accelerated Plant Breeding, Volume 1: Cereal Crops*; Springer Nature; Springer International Publishing: Cham, Switzerland, 2020; Volume 1.

63. Sarinelli, J.M.; Murphy, J.P.; Tyagi, P.; Holland, J.B.; Johnson, J.W.; Mergoum, M.; Mason, R.E.; Babar, A.; Harrison, S.; Sutton, R.; et al. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor. Appl. Genet.* **2019**, *132*, 1247–1261. [CrossRef]

64. Cericola, F.; Jahoor, A.; Orabi, J.; Andersen, J.R.; Janss, L.L.; Jensen, J. Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. *PLoS ONE* **2017**, *12*, e0169606. [CrossRef]

65. Isidro, J.; Jannink, J.L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [CrossRef]

66. Andres, R.J.; Dunne, J.C.; Samayoa, L.F.; Holland, J.B. *Enhancing Crop Breeding Using Population Genomics Approaches*; Springer: Berlin/Heidelberg, Germany, 2020.

67. Sitonik, C.; Suresh, L.M.; Beyene, Y.; Olsen, M.S.; Makumbi, D.; Oliver, K.; Das, B.; Bright, J.M.; Mugo, S.; Crossa, J.; et al. Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theor. Appl Genet.* **2019**, *132*, 2381–2399. [CrossRef]

68. Nyaga, C.; Gowda, M.; Beyene, Y.; Muriithi, W.T.; Makumbi, D.; Olsen, M.S.; Suresh, L.M.; Bright, J.M.; Das, B.; Prasanna, B.M. Genome-wide analyses and prediction of resistance to mln in large tropical maize germplasm. *Genes* **2020**, *11*, 16. [CrossRef]

69. Gowda, M.; Das, B.; Makumbi, D.; Babu, R.; Semagn, K.; Mahuku, G.; Olsen, M.S.; Bright, J.M.; Beyene, Y.; Prasanna, B.M. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theor. Appl Genet.* **2015**, *128*, 1957–1968. [CrossRef] [PubMed]

70. Galli, G.; Lyra, D.H.; Alves, F.C.; Granato, Í.S.C.; e Sousa, M.B.; Fritsche-Neto, R. Impact of phenotypic correction method and missing phenotypic data on genomic prediction of maize hybrids. *Crop Sci.* **2018**, *58*, 1481–1491. [CrossRef]

71. Molenaar, H.; Boehm, R.; Piepho, H.P. Phenotypic selection in ornamental breeding: It's better to have the BLUPs than to have the BLUEs. *Front. Plant Sci.* **2018**, *871*, 1–14. [CrossRef] [PubMed]

72. Piepho, H.P.; Möhring, J.; Melchinger, A.E.; Büchse, A. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **2008**, *161*, 209–228. [CrossRef]

73. Piepho, H.P.; Möhring, J. Selection in cultivar trials—Is it ignorable? *Crop Sci.* **2006**, *46*, 192–201. [CrossRef]

74. Dramadri, I.O.; Nkalubo, S.T.; Kelly, J.D. Identification of QTL Associated with drought tolerance in Andean common bean. *Crop Sci.* **2019**, *59*, 1007–1020. [CrossRef]

75. Williams, W.P.; Buckley, P.M.; Davis, F.M. Combining ability for resistance in corn to fall armyworm and southwestern corn borer. *Crop Sci.* **1989**, *29*, 913–915. [CrossRef]

76. Sodedji, F.A.K.; Kwemoi, D.B.; Asea, G.; Kyamanywa, S. Response of provitamin—A maize germplasm to storage weevil *Sitophilus zeamais* (Motschulsky). *Int. J. Agron. Agric. Res.* **2016**, *9*, 1–13.

77. Kasozi, L.C.; Derera, J.; Tongoona, P.; Zziwa, S.; Foundation, M.G.; Box, P.O. Comparing the effectiveness of the "weevil warehouse" and "laboratory bioassay" as techniques for screening maize genotypes for weevil resistance. *J. Food Secur.* **2018**, *6*, 170–177.

78. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

79. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Development Core Team, Ed.; R Foundation for Statistical Computing: Vienna, Austria, 2011; Volume 1, p. 409.

80. Bates, D.M.; Maechler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]

81. De los Campos, G.; Pérez, P.; Vazquez, A.I.; Crossa, J. Genome-enabled prediction using the BLR (Bayesian linear regression) R-package. In *Methods in Molecular Biology*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 299–320.

82. Pérez, P.; de los Campos, G. BGLR: A statistical package for whole genome regression and prediction. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]

83. Pérez, P.; de los Campos, G.; Crossa, J.; Gianola, D. Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. *Plant Genome J.* **2010**, *3*, 106. [CrossRef]

84. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]

85. Covarrubias-Pazaran, G. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* **2016**, *11*, e0156744. [CrossRef]
86. Henderson, C.R.; Henderson, A.C.R. Best linear unbiased estimation and prediction under a selection model published by: International biometric society stable. *Biometrics* **1975**, *31*, 423–447. [CrossRef]
87. Kang, H.M.; Zaitlen, N.A.; Wade, C.M.; Kirby, A.; Heckerman, D.; Daly, M.J.; Eskin, E. Efficient control of population structure in model organism association mapping. *Genetics* **2008**, *178*, 1709–1723. [CrossRef]
88. Gilmour, A.R.; Thompson, R.; Cullis, B.R. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **1995**, *51*, 1440–1450. [CrossRef]
89. Searle, S.R. Applying the EM algorithm to calculating ML and REML estimates of variance components. In Proceedings of the American Statistical Association Meetings, San Francisco, CA, USA, 8–12 August 1993; pp. 1–9.
90. Tunnicliffe, G.W. On the use of marginal likelihood in time series model estimation. *JRSS* **1989**, *51*, 15–27.
91. Vanraden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
92. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J.* **2011**, *4*, 250. [CrossRef]
93. Stekhoven, D.J.; Bühlmann, P. Missforest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112–118. [CrossRef] [PubMed]
94. Chen, X.; Ishwaran, H. Genomics random forests for genomic data analysis. *Genomics* **2012**, *99*, 323–329. [CrossRef] [PubMed]
95. Romay, M.C.; Millard, M.J.; Glaubitz, J.C.; Peiffer, J.A.; Swarts, K.L.; Casstevens, T.M.; Elshire, R.J.; Acharya, C.B.; Mitchell, S.E.; Flint-Garcia, S.A.; et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **2013**, *14*, R55. [CrossRef]
96. Peiffer, J.A.; Flint-Garcia, S.A.; De Leon, N.; McMullen, M.D.; Kaeppler, S.M.; Buckler, E.S. The genetic architecture of maize stalk strength. *PLoS ONE* **2013**, *8*, e67066. [CrossRef]
97. Lorenz, A.J.; Smith, K.P.; Jannink, J.L. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Sci.* **2012**, *52*, 1609–1621. [CrossRef]
98. Arruda, M.P.; Brown, P.J.; Lipka, A.E.; Krill, A.M.; Thurber, C.; Kolb, F.L. Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. *Plant Genome* **2015**, *8*, 1–12. [CrossRef]
99. Foiada, F.; Westermeier, P.; Kessel, B.; Ouzunova, M.; Wimmer, V.; Mayerhofer, W.; Presterl, T.; Dilger, M.; Kreps, R.; Eder, J.; et al. Improving resistance to the European corn borer: A comprehensive study in elite maize using QTL mapping and genome-wide prediction. *Theor. Appl Genet.* **2015**, *128*, 875–891. [CrossRef]
100. Riedelsheimer, C.; Technow, F.; Melchinger, A.E. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genom.* **2012**, *13*, 452. [CrossRef]
101. Azodi, C.B.; Bolger, E.; McCarren, A.; Roantree, M.; de Los Campos, G.; Shiu, S.H. Benchmarking parametric and Machine Learning models for genomic prediction of complex traits. *G3 Genes Genomes Genet.* **2019**, *9*, 3691–3702. [CrossRef]
102. Drouaillet, B.E.; Mendez, C.A.R. Combinatorial aptitude and resistance to leaf damage of *Spodoptera frugiperda* (J.E. Smith) in maize germplasm native to Tamaulipas. *Rev. Mex. Cienc. Agríc.* **2018**, *9*, 81–93.
103. Alvarez, P.; Branco, J.; Filho, D.M. Diallel crossing among miaze populations for resistance to fall armyworm. *Sci. Agric.* **2002**, *59*, 731–741. [CrossRef]
104. Viana, P.A.; Guimarães, P.E.O. Maize resistance to the lesser cornstalk borer and fall armyworm in Brazil. In *Embrapa Milho e Sorgo-Artigo em Anais de Congresso (ALICE), Proceedings of the International Symposium on Insect Resistant Maize: Recent Advances and Utilization, Mexico City, Mexico, 27 November–3 December 1994*; Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT): Mexico City, Mexico, 1997.
105. Musundire, L.; Dari, S.; Derera, J.; Co-Zimbabwe, S.; Wgt, P.O.B. Genetic analysis of grain yield performance and weevil [*Sitophilus zeamais* (Motschulsky)] resistance in southern African maize hybrids. *Maydica* **2015**, *60*, M35.
106. Dhliwayo, T.; Pixley, K.V.; Kazembe, V. Combining ability for resistance to maize weevil among 14 southern African maize inbred lines. *Crop Sci.* **2005**, *45*, 662–667. [CrossRef]
107. Asoro, F.G.; Newell, M.A.; Beavis, W.D.; Scott, M.P.; Jannink, J.-L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome J.* **2011**, *4*, 132. [CrossRef]
108. Lenz, P.R.N.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Desponts, M.; Bousquet, J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genom.* **2017**, *18*, 1–17. [CrossRef]
109. Edwards, S.M.K.; Buntjer, J.B.; Jackson, R.; Bentley, A.R.; Lage, J.; Byrne, E.; Burt, C.; Jack, P.; Berry, S.; Flatman, E.; et al. The effects of training population design on genomic prediction accuracy in wheat. *Theor. Appl Genet.* **2019**, *132*, 1943–1952. [CrossRef]
110. Wang, W.; Cao, X.H.; Miclǎu, M.; Xu, J.; Xiong, W. The promise of agriculture genomics. *Int. J. Genom.* **2017**, *2017*. [CrossRef]
111. Spindel, J.; Iwata, H. Genomic selection in rice breeding. In *Rice Genomics, Genet Breed*; Springer: Singapore, 2018; pp. 473–496.
112. Crossa, J.; Pérez, P.; Hickey, J.; Burgueño, J.; Ornella, L.; Cerón-Rojas, J.; Zhang, X.; Dreisigacker, S.; Babu, R.; Li, Y.; et al. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* **2014**, *112*, 48–60. [CrossRef]
113. Ou, J.H.; Liao, C.T. Training set determination for genomic selection. *Theor. Appl. Genet.* **2019**, *132*, 2781–2792. [CrossRef]
114. Mangin, B.; Rincent, R.; Rabier, C.E.; Moreau, L.; Goudemand-Dugue, E. Training set optimization of genomic prediction by means of EthAcc. *PLoS ONE* **2019**, *14*, e0205629. [CrossRef] [PubMed]

115. Albrecht, T.; Wimmer, V.; Auinger, H.J.; Erbe, M.; Knaak, C.; Ouzunova, M.; Simianer, H.; Schon, C.-C. Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.* **2011**, *123*, 339–350. [CrossRef] [PubMed]

116. Hickey, J.M.; Dreisigacker, S.; Crossa, J.; Hearne, S.; Babu, R.; Prasanna, B.M.; Grondona, M.; Zambelli, A.; Windhausen, V.S.; Mathews, K.; et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* **2014**, *54*, 1476–1488. [CrossRef]

117. Krchov, L.-M.; Bernardo, R. Relative efficiency of genomewide selection for testcross performance of doubled haploid lines in a maize breeding program. *Crop Sci.* **2015**, *55*, 2091–2099. [CrossRef]

118. Mayor, P.J.; Bernardo, R. Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F2 populations. *Crop Sci.* **2009**, *49*, 1719–1725. [CrossRef]