doi:10.20944/preprints202007.0336.v1

Genomic Prediction of Tropical Maize Resistance to Fall Armyworm and Weevils: Genomic Selection Should Focus on Effective Training Set Determination

A. Badji[1*], L. Machida[2], D. B. Kwemoi[3*], F. Kumi[4], D. Okii[1], N. Mwila[1], S. Agbahoungba[5], A. Ibanda[1], A. Bararyenya[1], S. N. Nghituwamhata[1], T. Odong[1], P. Wasswa[1], M. Otim[3], M. Ochwo-Ssemakula[1], H. Talwana[1], G. Asea[3], S. Kyamanywa[1], and P. Rubaihayo[1]

[1] Department of Agricultural Production, Makerere Univesity, P.O.Box: 7062, Kampala, Uganda; arbad2009@caes.mak.ac.ug (A. Badji); 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 International-CIAT, P.O. Box: 24384, Kampala, Uganda; b
[3] Cereals Program, National Crop Resource Research Institute, P.O. Box: 7084, Kampala, Uganda; kdbomet@gmail.com (D.B.K.); otim_michael@yahoo.com (M.O.); grasea9@gmail.com (G.A.)
[4] Department of Crop Science, University of Cape Coast, PMB, Ghana; frankkumifk@gmail.com
[5] Laboratory of Applied Ecology, University of Abomey-Calavi, 01BP: 526, Cotonou, Benin; agbahoungbasymphorien@gmail.com
* Correspondence: A. Badji: arbad2009@caes.mak.ac.ug and D.B.K: kdbomet@gmail.com

Abstract: Genomic selection (GS) can accelerate variety release by shortening variety development phase when factors that influence prediction accuracies (PA) of genomic prediction (GP) models such as training set (TS) size and relationship with the breeding set (BS) are optimized beforehand. In this study, PAs for the resistance to fall armyworm (FAW) and maize weevil (MW) in a diverse tropical maize panel composed of 341 double haploid and inbred lines were estimated. Both phenotypic best linear unbiased predictors (BLUPs) and estimators (BLUEs) were predicted using 17 parametric, semi-parametric, and nonparametric algorithms with a 10-fold and 5 repetitions cross-validation strategy. n. For both MW and FAW resistance datasets with an RBTS of 37%, PAs achieved with BLUPs were at least as twice as higher than those realized with BLUEs. The PAs achieved with BLUPs for MW resistance traits: grain weight loss (GWL), adult progeny emergence (AP), and number of affected kernels (AK) varied from 0.66 to 0.82. The PAs were also high for FAW resistance RBTS datasets, varying from 0.694 to 0.714 (for RBTS of 37%) to 0.843 to 0.844 (for RBTS of 85%). The PAs for FAW resistance with PBTS were generally high varying from 0.83 to 0.86, except for one dataset that had PAs ranging from 0.11 to 0.75. GP models showed generally similar predictive abilities for each trait while the TS designation was determinant. There was a highly positive correlation (R=0.92***) between TS size and PAs for the RBTS approach while, for the PBTS, these parameters were highly negatively correlated (R=-0.44***), indicating the importance of the degree of kinship between the TS and the BS with the smallest TS (31%) achieving the highest PAs (0.86). This study paves the way towards 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.

INTRODUCTION

Insect damage on maize plants and stored grains potentially impede food security in Africa (1–3). The fall armyworm (FAW) and stem borers in the field, and the maize weevil (MW) in storage facilities are some of the most devastating insect pests on the continent causing yield losses ranging from 10 to 90% leading to loss of grain marketability, and consumer health concerns due to the possible contamination of the grain with mycotoxin 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 results. Several Africa adapted maize lines were developed and successfully tested for resistance to MW damage on grains 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, FAW is a recent pest on the continent and was 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).

However, 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 GWAS methods, both with several shortcomings, especially for complex traits (21–23) and have seldom been conducted in African germplasm (8,24), further impeding the application of MAS in the development of insect resistance maize lines in Africa. Moreover, in a previous study, we discovered several quantitative trait nucleotides and genes putatively associated with FAW and MW resistance which confirmed the quantitative nature of these traits, and thus, the difficulty in trying to improve 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 to 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 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 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 on 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 (RR-BLUP) (39) and genomic BLUP (GBLUP) (40). GBLUP was the first GP method developed and replaced the traditional pedigree-based relationship matrix by a genomic information-based matrix to improve prediction accuracies (PA) (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 effect of some SNPs to be zero. The Bayesian least absolute shrinkage and selection operator (Bayes LASSO) which is a assumes that the effects of all markers follow a double exponential distribution (45). Also, Zhou *et al.*(46) developed the Bayesian Sparse Linear Mixed Model (BSLMM) a parametric method that combines the hypotheses of both GBLUP and Bayesian methods and achieves higher PAs than BayesCπ, and BayesLASSO. Nonparametric or semi-parametric approaches like random forest, reproducing kernel Hilbert space (RKHS) (47,48), support vector regression (SVR) (49), and neural networks (NN) (50) are better suited for accounting for non-additive genetic effects (37,38) in contrast with parametric genomic prediction models (23,38,47,51).

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 an additional increase 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 (52–54). 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 (55,56). Compared to the above-mentioned parametric methods, deep learning techniques such as SVR, Multilayer Perceptron, and Convolutional Neural Networks models performed poorly in some studies (57,58). However, there are also instances where RKHS, SVR, and NN outperformed one or several of parametric methods, for instance, GBLUP, RRBLUP, and Bayesian algorithms for several traits in several crops including maize (53,59,60). These results are most likely because nonparametric GS models capture more adequately 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 (12,61,70,62–69).

Additional factors that influence PAs are the sizes of the training sets (TS) 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 extent of linkage disequilibrium (21,23,31,41,71). Since phenotyping is the current bottleneck in plant breeding and one of the disadvantages of GP is the requirement of large TS for high PAs to be achieved, determination of effective TS composition and size is critical for effective implementation of GS in crop improvement programs (21,72–75). Also, the best TS determination will depend on the genetic architecture and the extent of population structure of the trait targeted for GP (74), 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 (74). 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 (76). However, deciding the best TS selection method is not straightforward and depends on contexts (77).

Furthermore, in maize, GPs were previously conducted using either genotypic best linear unbiased estimators (BLUEs) (78–80) or best linear unbiased predictors (BLUPs) (31,41,81) as a means of phenotypic corrections (81). 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 radom and allow a shrinkage of the means towards the population mean (82). 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 (81). However, the shrinkage in the BLUP procedure accounts better for outliers and environmental variabilities (83) permitting better estimates of individual genetic effects than BLUEs (82) and, therefore, yield usually more accurate predictions of phenotypic performance (81,83,84). Furthermore, BLUPs are better in handling unbalanced data where, for example, the number of individuals is not the same in different locations or in the different replications of an experiment (51,81). On this basis, the current study was conducted to evaluate the efficacy of different parametric, semiparametric, and nonparametric, statistical, and machine learning GP models in achieving high prediction accuracies (PA) for maize resistance to FAW and MW traits in a diverse panel using both genotypic BLUEs and BLUPs.

MATERIAL AND METHODS

Genetic material and field experiments

The panel used in this study consisted of 358 maize lines from a diverse genetic and geographic background 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). It consisted of 71 inbred lines developed for various purposes at NaCRRI, 28 and five stem borer (SB)-resistant inbred lines from CIMMYT and IITA, respectively (6,13,14), 19 storage pest (SP)-resistant inbred lines (7,8) and a doubled haploid (DH) panel of 235 lines developed from insect-resistant parents at CIMMYT using six parents of which three were stem borer-resistant and 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) (Table S1).

The panel was planted and evaluated in three environments, at the 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 (85).

Locations	Geographical coordinates	Altitude above sea level	Minimum rainfall	Soil Characteristics
Kasese 2017B	0°16′10′′ N 30°6′9′′ E	1330 m asl	1000 mm	Sandy loam soils with a pH of 5.68
Namulonge 2018A	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 amounting to 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 (12 blocks in 2017B, five blocks in 2018A, and 10 blocks in 2019A) containing unreplicated lines. The experiment in 2018A was replicated twice.

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 three environments was carried out two months after planting when adequate natural FAW infestation and pressure had manifested, and scoring was based on a visual assessment on a scale of 1 (no or minor leaf damage) to 9 (all leaves highly damaged) as described by (86) and illustrated in Figure S1 (18).

Rearing and bioassay of MW were performed as described in previous experiments carried out at NaCRRI (87,88). Weevils were reared prior to the 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 creating a weevil-maize grain culture of 300 to 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 of pore size smaller than one mm was fitted on each lid to allow proper ventilation while preventing the weevils from escaping.

After harvest and shelling, 30 grams of grains from each line of the panel were weighed from a bulk of all three environments with the aim of having three replicates per genotype. However, due to lack of enough seed for some of the genotypes, only 64 lines could generate three replicates while 123 provided two and 132 only one. 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 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 separated 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. Also, the number of dead and living weevils was 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), the final number of damage-affected kernels (AK), the total number of holes on grains (NH), and the cumulative weight of the flour produced (FP).

Statical analyses of the phenotypic data

For GP analyses, 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) (89) and the *ranef* function of the R package (90) lme4 (91), 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 equation is the intercept.

Strategies for TS and BS determination

MW resistance traits

To estimate GP accuracies for MW resistance, 37% (126 out of 341) of lines from the panel which had both genotypic and phenotypic data on grain weight loss (GWL), adult progeny emergence (AP), and number of affected kernels (AK) were used as the TS. The remaining 215 lines with no phenotypic data were considered as the breeding set (BS) for which the genomic-estimated breeding values (GEBV) were predicted.

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.

Random-based TS determination

For the RBTS, first, 126 (37%) lines used for GPs of MW resistance traits were used as the TS to predict the GEBVs of the remaining lines. To build the second TS, the 215 (67%) lines used earlier as BS were considered as a TS. Then random selections of 75 and 85% of the lines in the entire panel were performed to determine the third and fourth TS to predict the GEBVs of the remaining lines.

Pedigree-based TS determination

The PBTS strategy is presented in Table 2. For the first dataset (FAW.Ped1), the 235 (68.91%) doubled haploid (DH) lines were used as a TS and the remainder 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 (86.22%) that were stem borer (SB) and storage pest (SP) resistance lines not from the International Center for Maize and Wheat Improvement (CIMMYT) whilst the CIMMYT SB and SP resistant lines constituted the BS. In the last dataset, FAW.Ped4, the 235 DH lines, and the SB and SP lines from CIMMYT and the International Institute for Tropical Agriculture (IITA) amounting to 287 (84.16%) genotypes were considered as the TS and the remaining from NaCRRI lines were considered as the BS (Table 2).

Table 2. Composition of the pedigree-based TS for fall armyworm (FAW) datasets.

FAW Datasets	FAW.Ped1	FAW.Ped2	FAW.Ped3	FAW.Ped4
TS composition	235 DH 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 to 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.

Genomic prediction algorithms

The GP analyses were performed using the BLUEs and BLUPs of the phenotypes and the 28919 single nucleotide polymorphism (SNP) markers that remained after duplicate removal (DRSNPs) and distributed across all the 10 chromosomes of the entire maize genome. Generation of the SNPs from this maize panel and duplicate removal were described in our previous study (25). The analyses were based on 17 prediction methods (a set of statistical and machine learning models, parametric, semi-parametric, and non-parametric methods) available in 12 GP algorithms implemented on 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. The 12 algorithms were drawn from five R packages designed for GP analyses:

The Bayesian Generalized Linear Regression (BGLR) package fits various types of parametric and semi-parametric Bayesian regressions (92). Methods used from this package are BL (Bayesian least absolute shrinkage and selection operator [LASSO]) (93), BLR (Bayesian Linear Regression) (94), BRR (Bayesian Ridge Regression) (95), RKHS (Reproducing Kernel Hilbert Space) (47,48), BayesA (26), BayesB (43) , and BayesC (44). The Sommer (solving mixed model equations in R) package (96) implements the function mmer (mixed model equations in R) for GP. It uses four algorithms: Efficient Mixed Model Association (EMMA) (97), Average Information (AI) (98), Expectation Maximization (EM) (99), and the default Newton-Raphson (NR) (100).

The BRNN (Bayesian regularization for feed-forward neural networks) package (101) fits a two-layer neural network (102,103). BRNN uses the Nguyen and Widrow algorithm (104) to assign initial weights and the Gauss-Newton algorithm to perform the optimization.

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. This package also implements the Mixed Model Solution (MMS) that calculates the maximum-likelihood (ML) or restricted-ML (REML) solutions for mixed models to perform GP (105). The R package missForest uses the function missForest that implements Breiman's random forest algorithm for classification and regression(106). On the KDCompute 1.5.2.beta platform, the user can choose either Square root or Regression options for the classification of the number of variables randomly sampled as candidates at each split (https://kdcompute.igss-africa.org/ kdcompute/plugins). 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
Sommer function mmer with Average information (AI)	mmer_AI	Parametric/Mixed model
mmer with Expectation Maximization (EM)	mmer_EM	Parametric/Mixed model
mmer with Efficient Mixed Model Association (EMMA)	mmer_EMMA	Parametric/Mixed model
mmer with default Newton-Raphson (NR)	mmer-NR	Parametric/Mixed model
rrBLUP	rrBLUP	Parametric/Mixed model
Mixed Model solution with Maximum Likelihood (ML)	mms_ML	Parametric/Mixed model
Mixed Model solution with Restricted Maximum Likelihood (REML)	mms_REML	Parametric/Mixed model
BayesB	BayesB	Parametric/Bayesian
BayesA	BayesA	Parametric/Bayesian
BayesC	BayesC	Parametric/Bayesian
Bayesian LASSO	BL	Parametric/Bayesian
Bayesian Ridge Regression	BRR	Parametric/Bayesian
Bayesian Linear Regression	BLR	Parametric/Bayesian
Bayesian regularization for feed-forward neural networks	*brnn*	*Parametric/Bayesian*
RKHS	RKHS	Semi-parametric/Bayesian
missForest with Square root (Sqt)	missForest_Sqt	Nonparametric/Machine learning
missForest with Regression (Regression)	missForest_Reg	Nonparametric/Machine learning

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).

RESULTS

Higher PAs achieved with BLUPs for FAW and MW resistance traits as compared with 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. Generally, BLUPs produced better predictions than BLUEs by at least two orders of magnitude in terms of PAs (Figure 1). The PAs realized

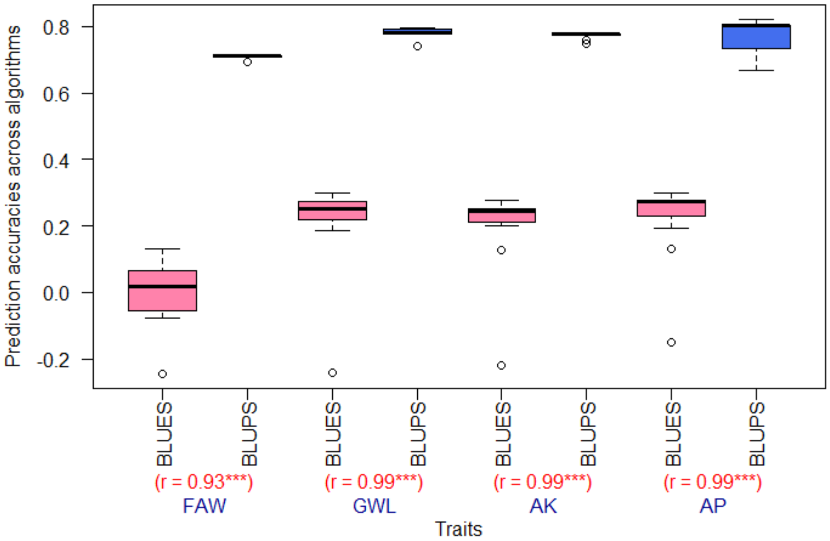

Figure 1. Boxplot of PAs (prediction accuracies) for best linear unbiased estimators (BLUEs) (in pink) and predictors (BLUPs) (in blue) of maize resistance to FAW and MW across prediction models and correletations (r) between BLUEs and BLUPs of each trait. FAW (fall armyworm),MW (Maize weevil)resistance traits; GWL (grain weight loss), AP (adult progeny emergence (AP), number of affected kernels (AK), flour production (FP), and number of holes (NH).

with BLUEs (Figure S2) varied from -0.246 for FAW (mms_ML) to 0.299 for AP (BayesB), while PAs for BLUEs 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 MW resistance traits such as AP, AK, and GWL (Figure 1), therefore, only results for BLUPs of resistance traits will be presented hereafter.

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

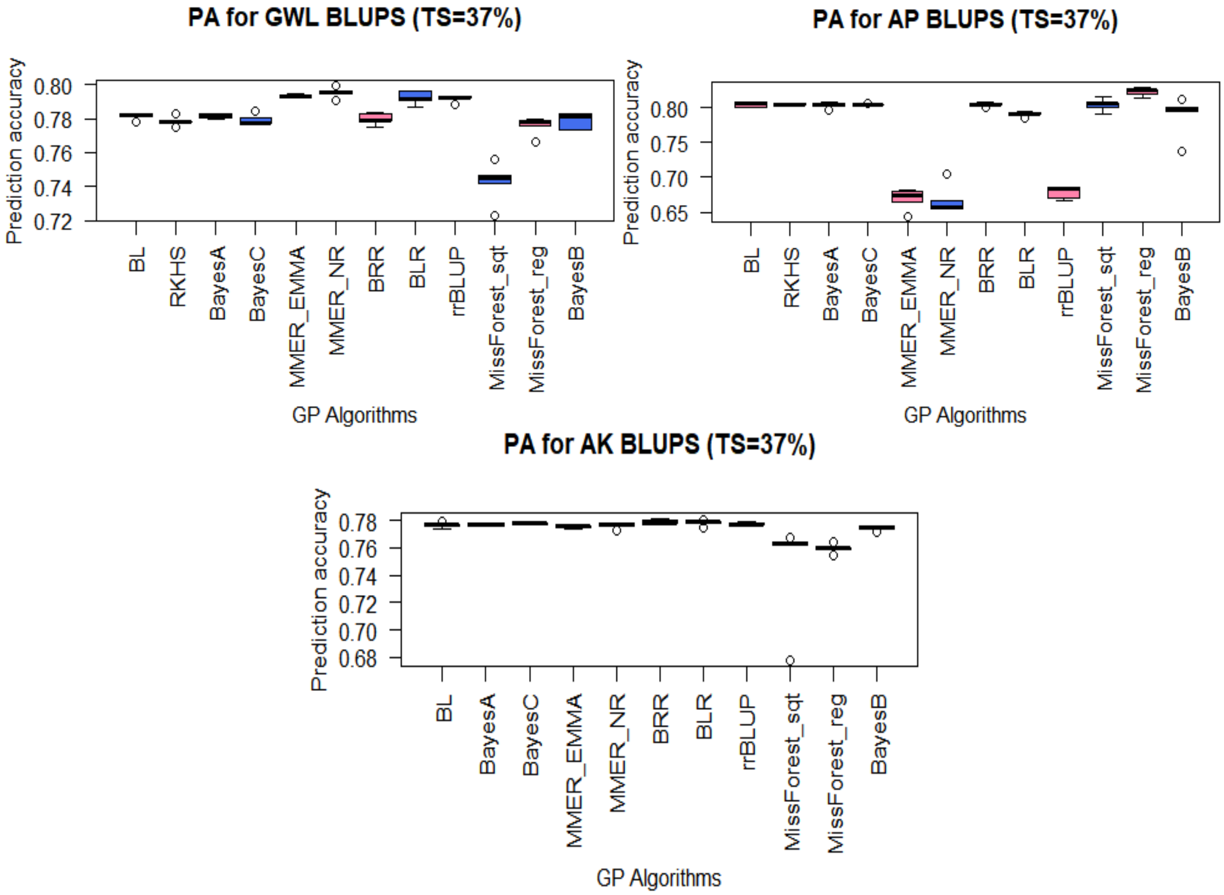

Figure 2. Boxplots of the genomic prediction accuracies (PA) of BLUPs for MW resistance traits: GWL (grain weight loss), AP (adult progeny emergence (AP), and number of affected kernels (AK) (See Table 3 for GP algorithms).

to work on the AK trait. The highest PAs were achieved for AP with missForest_reg (0.823), followed by BRR (0.805) and RKHS (0.804) with the lowest prediction accuracy of 0.667 obtained through the mmer_NR algorithm (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). Generally, Bayesian models predicted better than both mixed model and machine learning methods, although the differences were small (Figure S3).

PA for FAW resistance using BLUPs

The different maize resistance to FAW datasets showed high predictive abilities achieved through 10 of the 17 GP algorithms used in the study. Regarding the RBTS approach, the PAs were the lowest with the dataset that had a TS composed of 37% (lowest size) of the panel and the 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 composition (37% of the panel), higher PAs were achieved for MW resistance

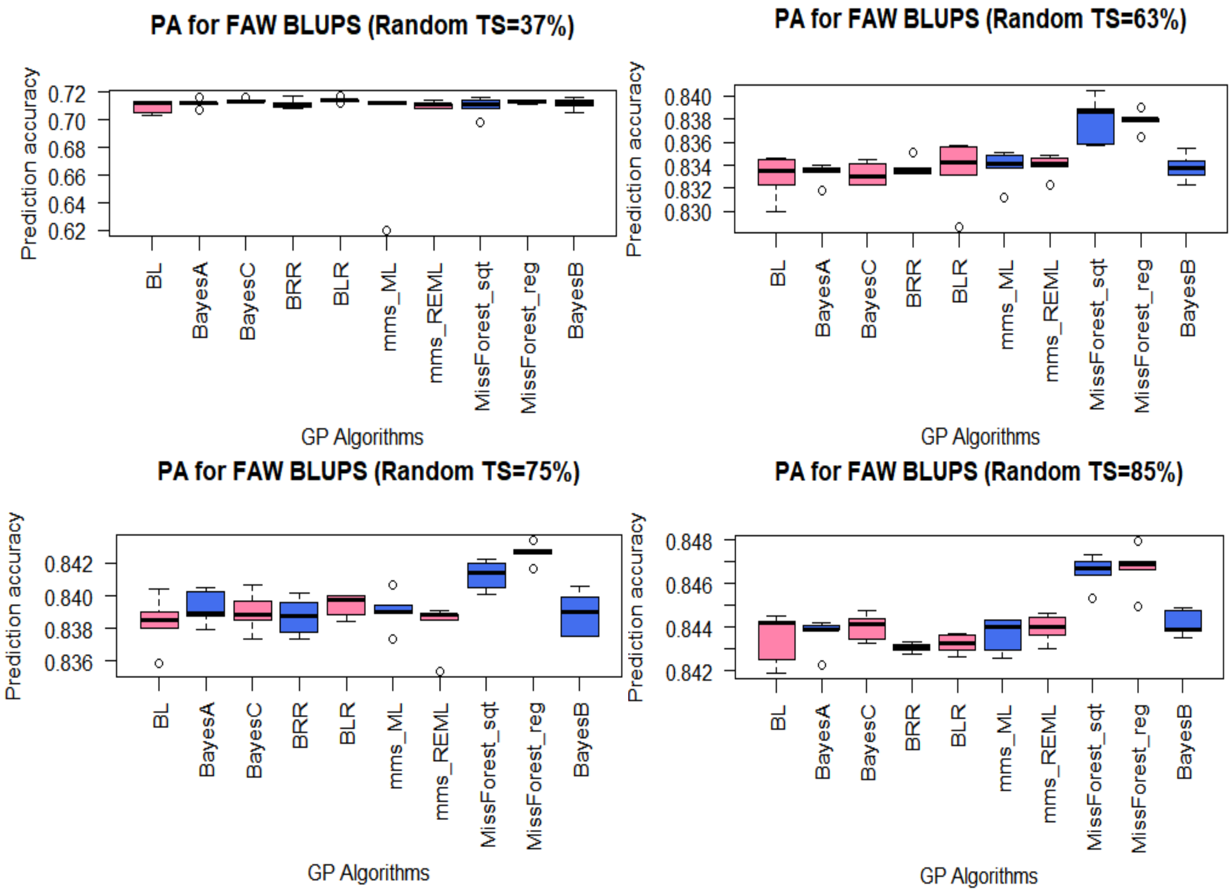

Figure 2. Boxplot of PA for maize resistance to the fall armyworm (FAW) datasets with the RBTS approach with random selection 37, 63, 75, and 87% of the entire panel (see Table 3 for GP algorithms).

traits such as GWL, AP, and AK than for FAW resistance (Figure S3). The PA for the TS of 63% varied from 0,833 for BL method to 0.838 for the missForest_Sqt, thus, there was a small variation among 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 TS 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 ($P<22e^{-16}$) positive correlation of 0.92 (Figure 4) between the PAs and TS sizes for FAW datasets for the RBTS

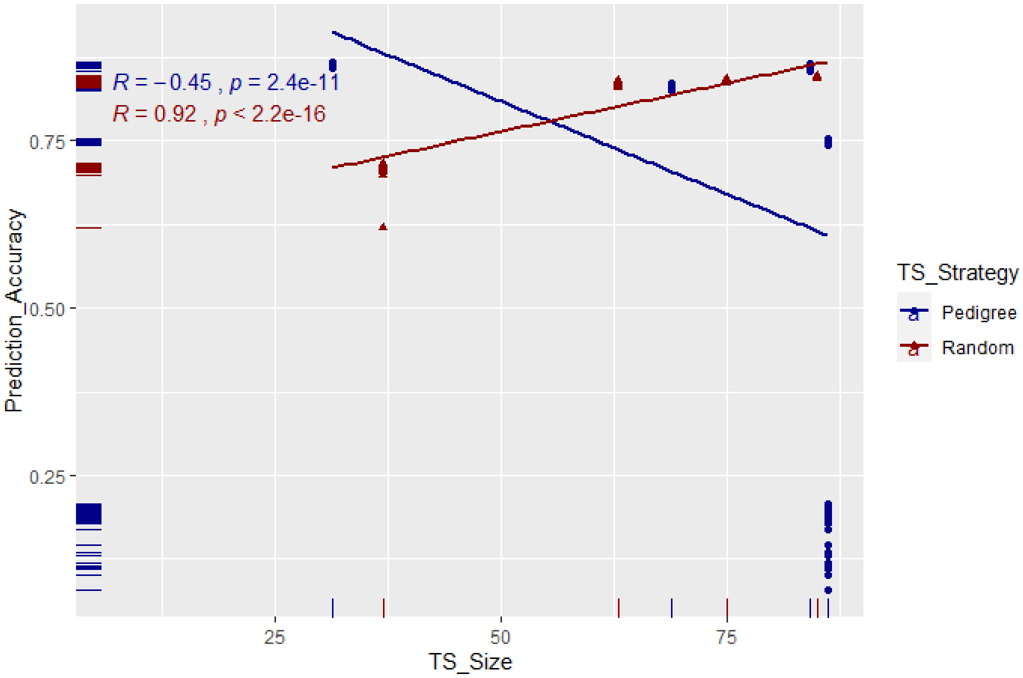

Figure 4. Pearson correlation between training set (TS) sizes and prediction accuracies across the 10 genomic prediction models conducted on RBTS and PBTS datasets for fall armyworm resistance (FAW).

denoting a steady improvement of the PAs as the TS sized increased. However, the PAs for FAW resistance reached a plateau at TS size above 63% of the panel (Figure 5).

Although the PAs do 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 almost all the traits, except for the TS of 37% where Bayesian methods such as BLR and BayesC showed a slight advantage over the former (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. Regarding FAW.Ped2 (TS=31.09%), the PAs ranged from 0.862 for BayesC to 0.864 for mms_REML.

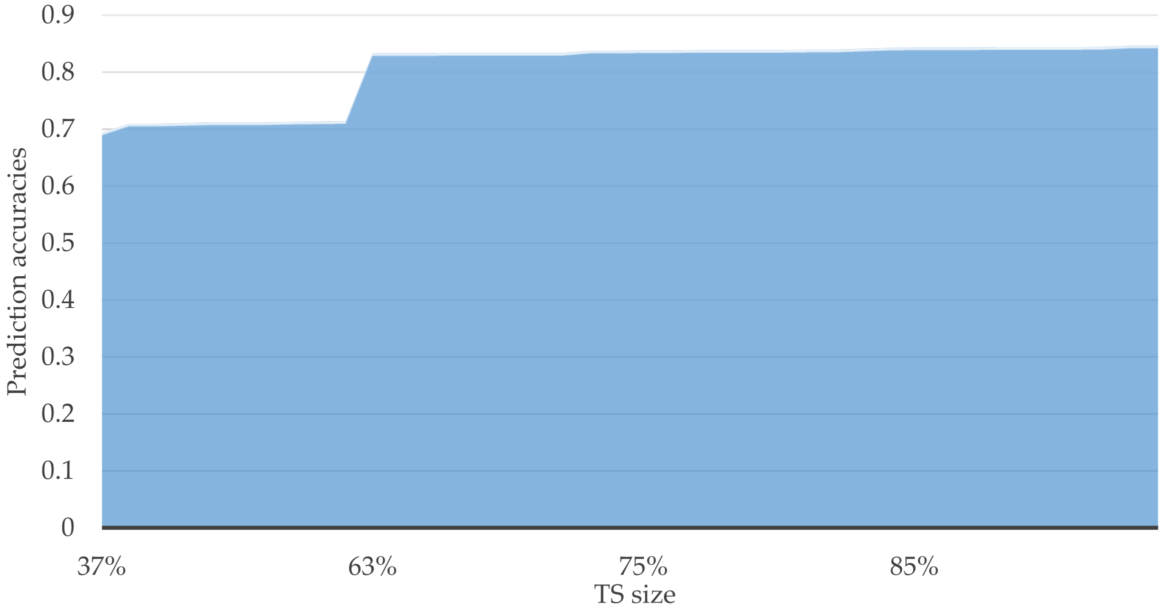

Figure 5. Prediction accuracies for FAW with RBTS accros algorithms and training set (TS) sizes in percent of the total panel.

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. Consequently, the Pearson correlation between the sizes of the PBTS datasets and the predictions accuracies for the 10 GP algorithms revealed a significant (P>2.4e-11) negative relationship of r=-0.33 (Figure 4).

In these datasets, the PAs were more influenced by the composition of the TS and its pedigree 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 higher PAs from all the 10 algorithms than when considering the CIMMYT SB and SP resistant lines as BS and the remainder as a TS (FAW.Ped3), where only Machine learning algorithms missForest_reg and missForest_Sqt achieved high PAs. Furthermore, the composition of the TS and its relationship with the BS determined which GP methods achieved the highest PAs with Machine learning algorithms working best on FAW.Ped1 and FAW.Ped3, linear mixed model approaches outperforming Bayesian and Machine Learning algorithms on FAW.Ped2 and FAW.Ped4, and Bayesian methods ranking 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).

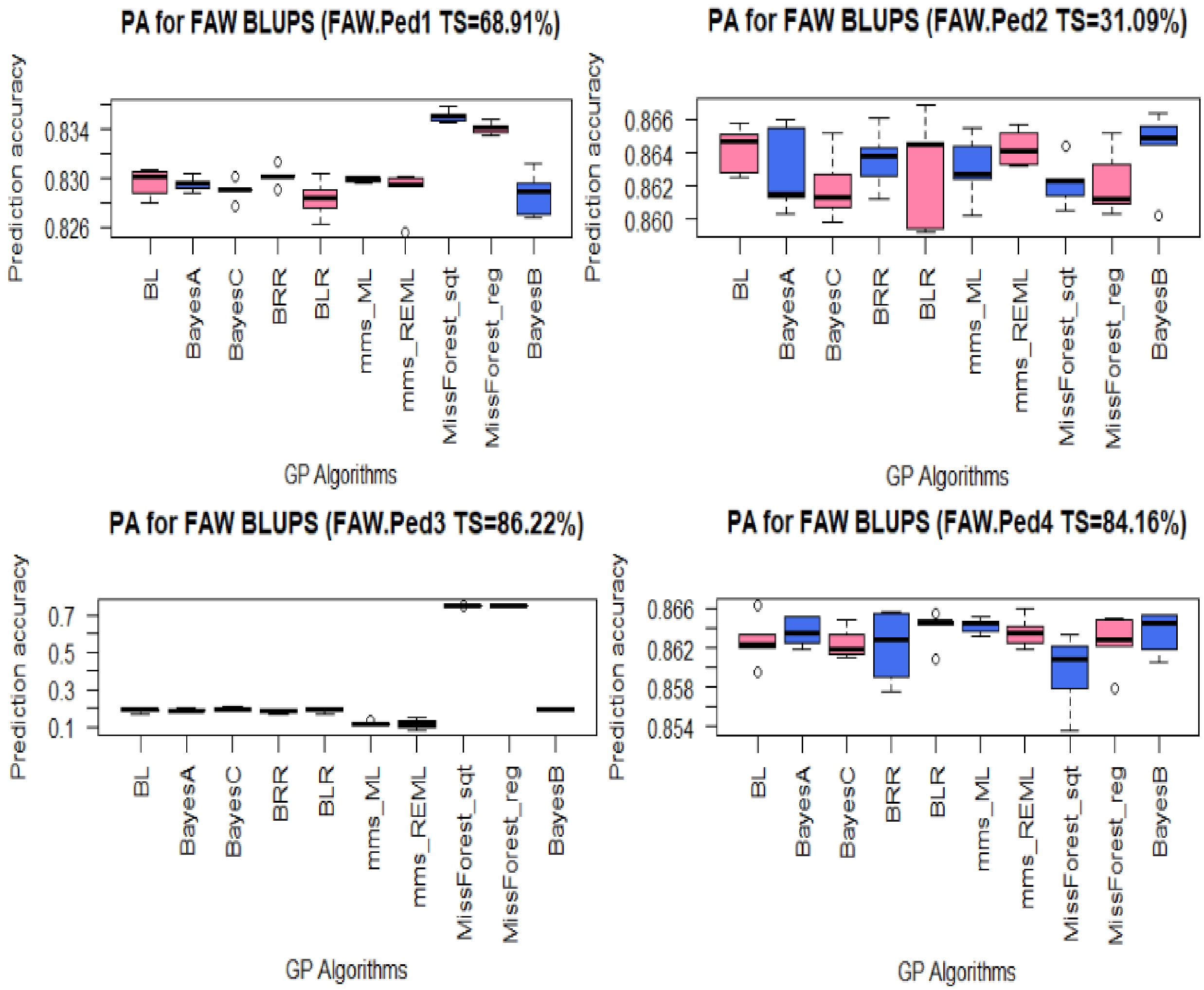

Figure 6. Boxplots of PAs for maize resistance to the fall armyworm (FAW) datasets using the PBTS approach (see Table 2 for the PBTS strategy and Table 3 for GP algorithms).

DISCUSSION

Tropical maize germplasm is characterized by rapid linkage disequilibrium (LD) decay with high diversity (107). These genetic characteristics make genomic selection (GS) a promising approach to integrate into African breeding programs (108). Given the diversity of GP models and their differential performances depending on crops and trait architectures and other other parameters such as the size of the training set (TS) and it its relationship with the breeding set (BS) (31,37), 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 genomic prediction (GP) algorithms for both traits using both BLUES and BLUPs of these phenotypes, and random and pedigree-based TS determination strategies.

Higher PAs for BLUPs compared to BLUEs for both FAW and MW resistance traits

With an RBTS of 37% of the panel, 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. Generally, 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 (81–84). 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 (51,81) as it 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 on in this study and will be further discussed.

High PAs 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), and the number of affected kernels (AK), and for FAW resistance were achieved in agreement with the moderate to high heritability values for these traits as earlier reported (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 (73,75).

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 on 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 (31,76,109,110). Very few reports of GP are available for maize resistance to biotic stresses. Similar and higher PAs were achieved for maize resistance to chlorotic mottle virus (up to 0.95) and maize lethal necrosis (reaching 0.87) in tropical germplasm (78). 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 (111) in temperate germplasm. Similarly, Gowda *et al.*(80) reported moderate PAs (close to 0.60) for maize resistance to the biotic stress, the maize lethal necrosis in tropical maize populations.

GP algorithms performed differently on FAW and MW maize resistance traits

In this study, several GP models including statistical and Machine Learning algorithms, 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, similar to earlier model benchmarking reports (112,113). 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) and, to a lesser extent, 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 resistance traits evaluated in this study could be due to differences in the genetic structures (extent of additive vs. nonadditive gene action) of the respective traits (23,38,47,51). Maize resistance to FAW, which was moderately heritable across environments (25), would be expected to be controlled by both additive and nonadditive genetic factors (62,63,70), 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 (64,68) in the current panel.

Influence of the size and the composition of TS and BS on PAs

Two factors, the relative sizes of the TS and BS (RBTS approach) and their kinship (PBTS approach), affected the PAs across FAW resistance datasets which corroborated earlier reports

(31,74,76,109,110,114,115). 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 (116). This increase was followed by a slight gain in predictability at 75% (0.837 to 0.843) and 85% (0.843 to 0.847). Similar to results in the current study, PAs were reported to reach a plateau and not increase significantly when TS reached a certain size (21,75,116–118). Consequently, 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>2.2e-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.*(116).

The composition of the TS and its relationship with the BS are determinant factors for the genomic predictability of complex traits (74,119–121). In the current study, using the PBTS approach, these 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% than in 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<13e-08) negative correlation (R=-0.33) between the sizes of the PBTS and the achieved PAs.

However, it is not very clear why the prediction between the BS FAW.Ped3 (47 CIMMYT SB and SP resistant lines) and the TS (DH, IITA SB, and NaCRRI lines) led to lower PAs in FAW.Ped3. A possible explanation is 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.*(118) 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 have sampled 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 (118). 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 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 would be of interest in future breeding activities targeted at improving insect resistance in maize (23,122–124) and potentially useful for GS of complex traits with low to moderate heritability (125).

Conclusion

This study assessed the influence of the size and composition of the training set on the genomic predictability of maize resistance to fall armyworm (FAW) and maize weevil (MW) in a diverse Africa adapted maize population using several 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 correlatedwith the size of the training set, the composition and the relationship of the training set with the breeding set were more influencial in predicting line performance. Also, TS determination-related parameters were more important than the type of genomic prediction models in predicting FAW and MW resistance traits. Therefore, future genomic selection programs for maize resistance to insect pests such as the Fall armyworm and the maize weevil in Africa should put more effort in designing effective training set for improved genetic gains coupled with short breeding cycles and accelerated variety release. Such programs would 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 agroecologies.

SUPPLEMENTARY: TABLES AND FIGURES

Table S1. Descriptions of parents and crosses that constituted the doubled-haploid population

		Parents of the doubled-haploid population			
	Descriptions	**CKSBL10007**	**CKSBL10011**	**CKSBL10203**	**CKSPL10066**
CKSBL10007	CIMMYT Stem borer-resistant line				
CKSBL10011	CIMMYT Stem borer-resistant line				
CKSBL10203	CIMMYT Stem borer-resistant line				
CKSPL10066	CIMMYT Storage pest- resistant line		X		
CML312	CIMMYT elite line	X		X	
CML485	CIMMYT elite line				X

X=Crosses between lines

Figure S1. Rating of maize plants based on foliar damage by FAW (18)**.**

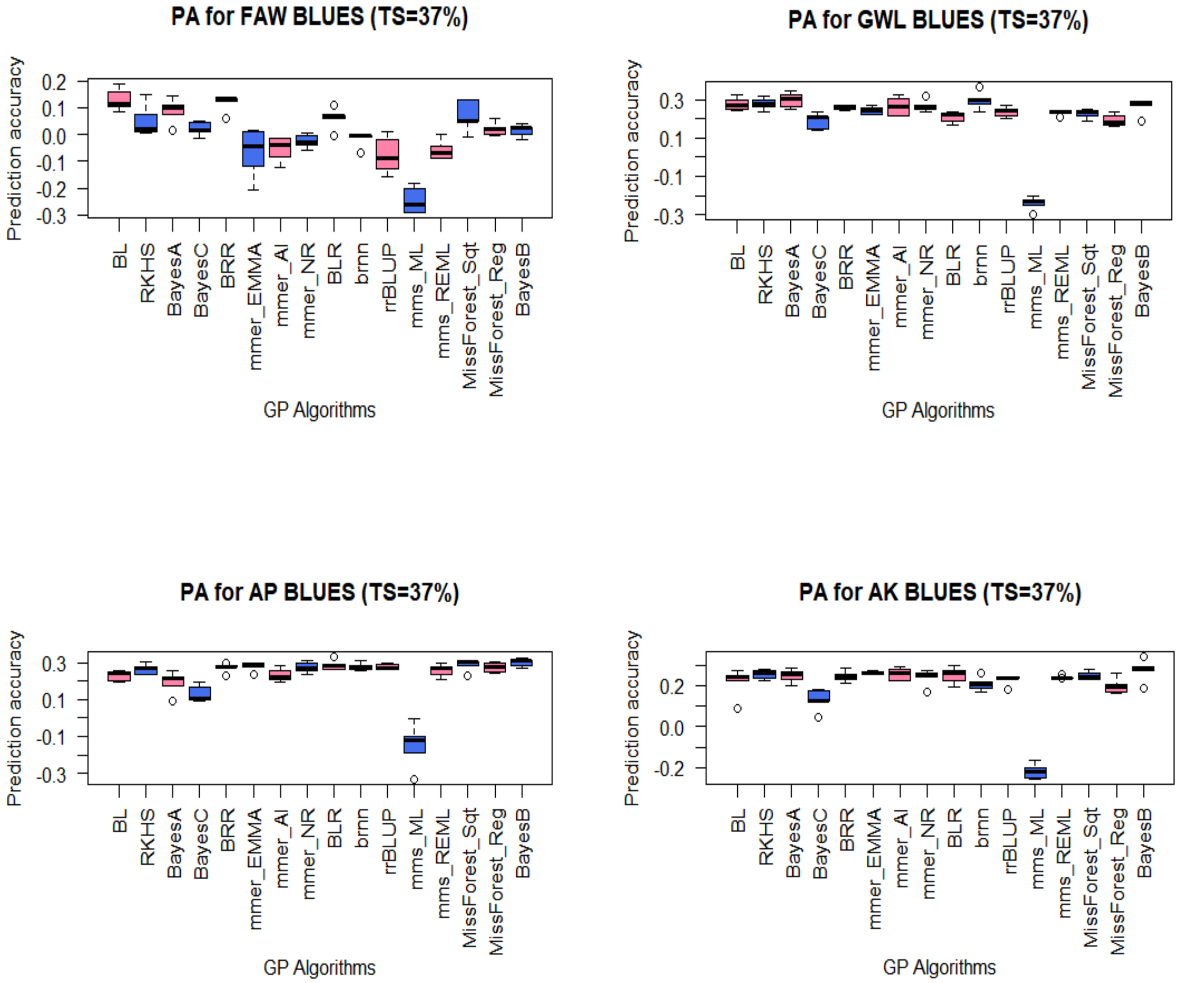

Figure S2. Boxplot of PA for best linear unbiased estimators (BLUEs) of maize resistance to the fall armyworm (FAW) and maize weevil (MW) with identical training set size (37%) and compositions

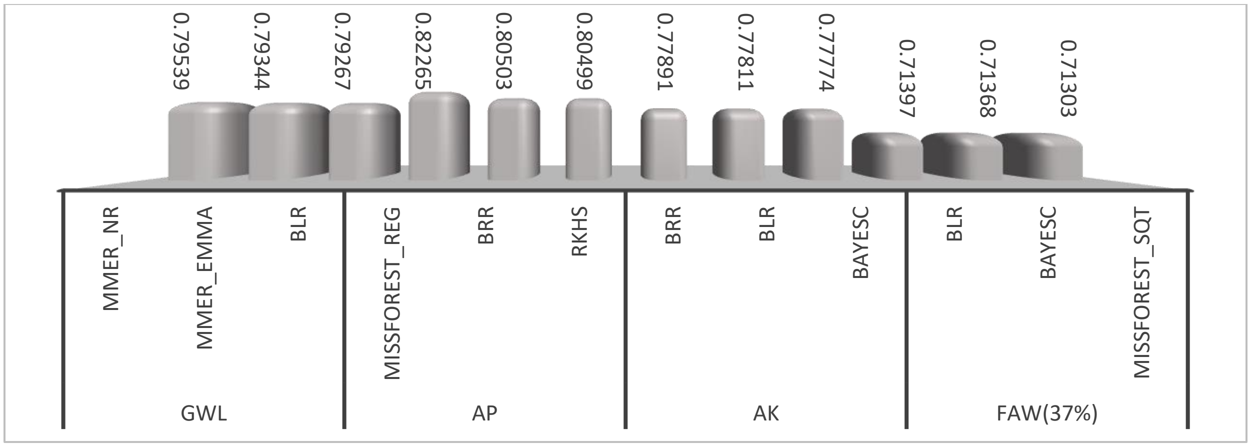

Figure S3. Comparisons of genomic prediction accuracies of the three best algorithms for best linear unbiased predictors (BLUPs) of maize weevil resistance traits: number of affected kernels (AK), adult progeny emergence (AP), and grain weight loss (GWL) vs., fall armyworm resistance dataset with identical TS.

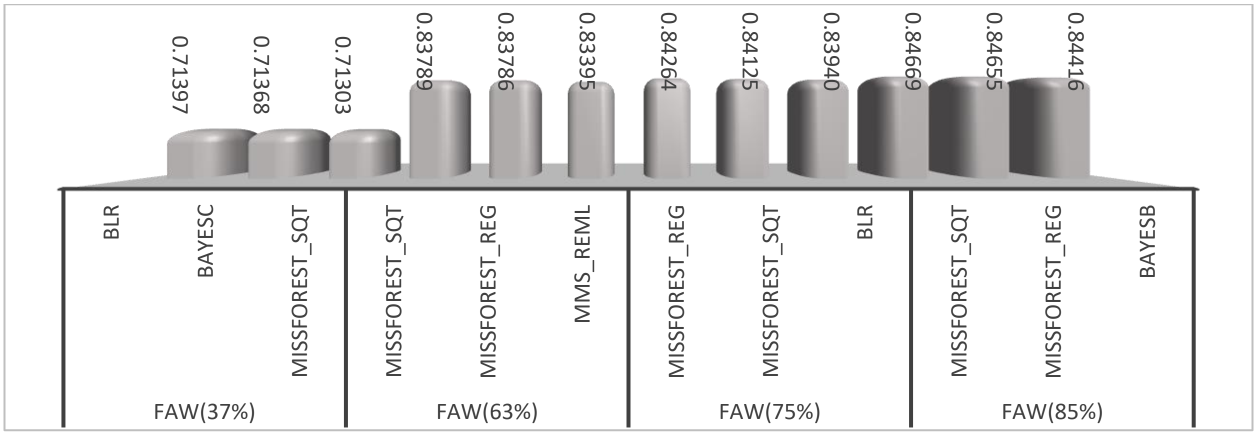

Figure S4. Genomic prediction accuracies of the three best algorithms for each fall armyworm resistance BLUPs datasets with RBTS of 37, 62, 75, and 85% of the entire dataset.

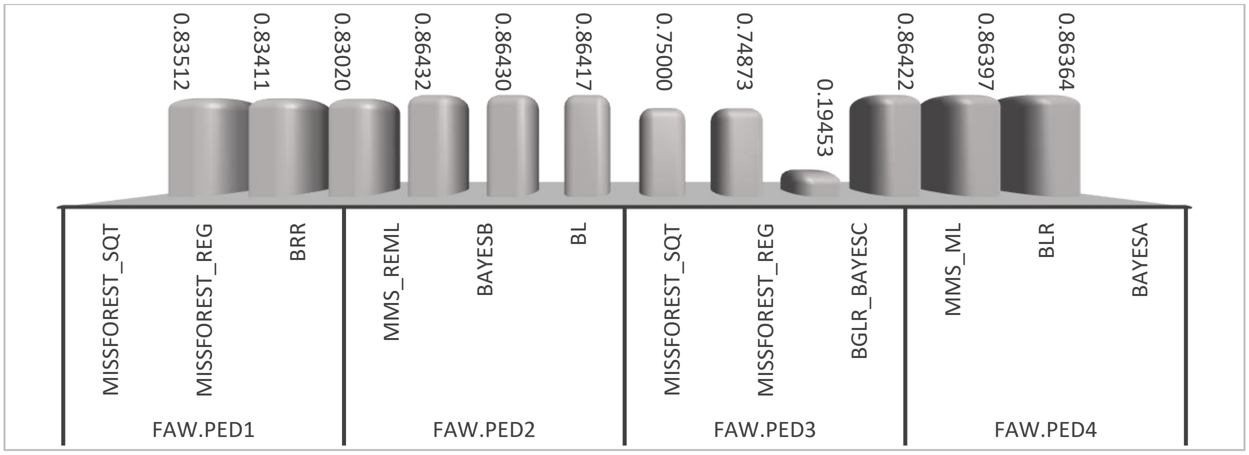

Figure S5. Genomic prediction accuracies of the three best algorithms for each fall armyworm resistance BLUPs datasets with PBTS.

AUTHOR CONTRIBUTION

Conceptualization, A.Badji, P.R., S.K., M.O., D.B.K. and L.M.; Methodology, A.Badji., D.B.K. and L.M.; Investigation, A.Badji and D.B.K.; Formal Analysis, A.Badji and LM; Resources, A.Badji, G.A., M.O., D.B.K. and L.M.; Visualization, A.Badji; Supervision, P.R., S.K., M.O. and L.M.; Project Administration, P.R. and M.O.; Funding Acquisition, A.Badji, M.O., G.A., D.B.K. and P.R. ; Writing—Original Draft Preparation, A.Badji; Writing—Review and Editing, All authors read and critically revised the manuscript.

FUNDING

This research was funded by the capacity building competitive grant training the next generation of scientists provided by Carnegie Cooperation of New York through the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM: RU/2016/Intra-ACP/RG/001). A. Badji received a Ph.D. scholarship from the Intra- ACP Academic mobility for Crop Scientists for Africa Agriculture (CSAA) project. Genotyping of the lines was carried out through a project of D.B.K. thanks to the Integrated Genotyping Service and Support (IGSS) coordinated by the International Livestock Research Institute (ILRI) and Bioscience east and central Africa (BecA) grant number: IGSS-DL0274. The National Crop Resources Research Institute (NaCRRI) of Namulonge (UGANDA) through a grant of the USAID Feed-the-Future Uganda, Agriculture Research Activity/Maize payed the article processing charges. Further, NaCRRI financially and logistically supported field and laboratory activies of this research.

ACKNOWLEDGEMENT

The authors thank all the technicians for experiment setup and data collection in fields and laboratories of NaCRRI at Namulonge and Kasese, UGANDA. The authors acknowlege NaCRRI, the International Maize and Wheat Improvement Center (CIMMYT) of Nairobi, KENYA, and the International Institute of Tropical Agriculture (IITA) of Ibandan, Nigeria for providing the original germplasm used for this research. The authors thank Prof. Clay SNELLER of the Ohio State University and all the personel of ILRI and BecA who provided the genotyping support at BecA/ILRI.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

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. African J Biotechnol.

2008;7(20):3777–82.
2. Shiferaw B, Prasanna BM, 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(3):307–27.
3. Awata LAO, Tongoona P, Danquah E, Ifie BE, Suresh LM, Jumbo MB, et al. Understanding tropical maize (Zea mays L .): The major monocot in modernization and sustainability of agriculture in sub-Saharan Africa. Int J Adv Agric R|esearch. 2019;7:32–77.
4. Nyukuri RW, Wanjala FM, Kirui SC, Cheramgoi E, Chirchir E, Mwale R. Amage of stem borer species to zea mays l., sorghum bicolor l.and three refugia graminae. Adv Agric Biol [Internet]. 2014;1(2):37–45. Available from: http://pscipub.com/Journals/Data/JList/Advance in Agriculture and Biology/2014/Volume 1/Issue 2/4.pdf
5. Tefera T, Goftishu M, Ba M, Rangaswamy M. A Guide to Biological Control of Fall Armyworm in Africa Using Egg Parasitoids. 1st ed. Nairobi, Kenya; 2019.
6. Munyiri SW, Mugo SN, Otim M, Mwololo JK, Okori P. Mechanisms and Sources of Resistance in Tropical Maize Inbred Lines to Chilo partellus Stem Borers. J Agric Sci [Internet]. 2013;5(7):51–60. Available from: http://www.ccsenet.org/journal/index.php/jas/article/view/26063
7. Mwololo JK, Mugo S, Okori P, Tefera T, Otim M, Munyiri SW. Sources of Resistance to the Maize Weevil Sitophilus Zeamais in Tropical Maize. J Agric Sci. 2012;4(11):206–15.
8. Mwololo JK. Resistance in Tropical Maize To the Maize Weevil and Larger Grain Borer. Makerere Univ. 2013;PhD thesis.
9. Kasozi LC, 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(1):83–97.
10. Tende R, Derera J, Mugo S, Oikeh S. Original Paper Open Access Estimation of genetic diversity of germplasm used to develop insect-pest resistant maize. 2016;
11. Khakata S, Mbute FN, Chemining’wa GN, Mwimali M, Karanja J, Harvey J, et al. Post-harvest evaluation of selected inbred lines to maize weevil Sitophilus zeamais resistance. J Plant Breed Crop Sci [Internet]. 2018;10(5):105–14. Available from: http://academicjournals.org/journal/JPBCS/article-abstract/4A7E8F556752
12. Sodedji FKA, Kwemoi DB, Kasozi CL, Asea G, Kyamanywa S. Genetic analysis for resistance to Sitophilus zeamais (Motschulsky) among provitamin-A maize germplasm. Maydica [Internet]. 2018 Sep 25 [cited 2019 Nov 13];63(2):8. Available from: https://journals-crea.4science.it/index.php/maydica/article/view/1698
13. Munyiri WS, Mugo NS, Otim M, Tefera T, Beyene Y, Mwololo KJ, et al. Responses of tropical maize landraces to damage by Chilo partellus stem borer. African J Biotechnol. 2013;12(11):1229–35.
14. Munyiri SW, Mugo SN, Mwololo JK. 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(3):81–90.
15. Mwololo JK, Munyiri SW, 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(15):1–22.
16. Goergen G, Kumar PL, Sankung SB, Togola A, Tamò M. First report of outbreaks of the fall armyworm spodoptera frugiperda (J E Smith) (Lepidoptera, Noctuidae), a new alien invasive pest in West and Central Africa. PLoS One. 2016;11(10):1–9.
17. Padhee AK, Prassanna BM. The emerging threat of Fall Armyworm in India. Indian Farming. 2019;69(1):51–4.
18. Prasanna BM, Huesing JE, Eddy R, Peschke VM, Regina E, Virginia, M P. Fall Armyworm in Africa: a guide for integrated pest management. First Edit. Prasanna BM, Regina E, Virginia, M P, editors. West Africa Regional Training of Trainers and Awareness Generation Workshop on Fall Armyworm Management, IITA, Cotonou, Bénin. CIMMYT; 2018.
19. Gedil M, Menkir A. An Integrated Molecular and Conventional Breeding Scheme for Enhancing Genetic Gain in Maize in Africa. 2019;10(November):1–17.
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(M4).
21. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci [Internet]. 2017;xx(11):1–15. Available from: https://doi.org/10.1016/j.tplants.2017.08.011
22. Roorkiwal M, Jarquin D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, 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):1–11.
23. Robertsen CD, Hjortshøj RL, Janss LL. Genomic Selection in Cereal Breeding. 2019;1–16.
24. Munyiri SW, Mugo SN. 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. African J Biotechnol [Internet]. 2017 Jul 12 [cited 2017 Aug 19];16(28):1579–89. Available from: http://academicjournals.org/journal/AJB/article-abstract/2B7761765096
25. Badji A, Kwemoi DB, Machida L, Okii D, Mwila N, Agbahoungba S, et al. Genetic Basis of Maize Resistance to Multiple-Insect Pests : Integrated Genome-Wide Comparative Mapping and Candidate. Genes (Basel). 2020;11(689):1–27.
26. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. 2001;
27. Jannink J, Lorenz AJ, Iwata H. Genomic selection in plant breeding : from theory to practice. Brief Funct Genomics. 2010;9(2):166–77.
28. Massman JM, Jung HJG, 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(1):58–66.
29. Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, et al. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 2015;55(1):154–63.
30. Vivek BS, Krishna GK, Vengadessan V, Babu R, Zaidi PH, Kha LQ, et al. Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. Plant Genome. 2017;10(1):1–8.
31. Liu X, Wang H, Wang H, Guo Z, Xu X, Liu J, et al. Factors affecting genomic selection revealed by empirical evidence in maize. Crop J [Internet]. 2018;6(4):341–52. Available from: https://doi.org/10.1016/j.cj.2018.03.005
32. Yuan Y, Scheben A, Batley J, Edwards D. Using Genomics to Adapt Crops to Climate Change. In: Sustainable Solutions for Food Security [Internet]. Cham: Springer International Publishing; 2019. p. 91–109. Available from: http://link.springer.com/10.1007/978-3-319-77878-5_5
33. Zhang X, Pérez-Rodríguez P, Burgueño J, Olsen M, Buckler E, Atlin G, et al. Rapid cycling genomic selection in a multiparental tropical maize population. G3 Genes, Genomes, Genet. 2017;7(7):2315–26.
34. Hassen M Ben, Bartholomé J, Valè G, Cao TV, 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(7):2319–32.
35. Muleta KT, Pressoir G, Morris GP. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti : A Simulation Study. 2019;9(February):391–401.
36. Suontama M, Klápště J, Telfer E, Graham N, Stovold T, Low C, et al. Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. Heredity (Edinb). 2019;122(3):370–9.
37. Voss-Fels KP, Cooper M, Hayes BJ. Accelerating crop genetic gains with genomic selection. Theor Appl Genet [Internet]. 2019 Mar 19;132(3):669–86. Available from: https://doi.org/10.1007/s00122-018-3270-8
38. Kadam DC, Lorenz AJ. Evaluation of nonparametric models for genomic prediction of early-stage single crosses in maize. Crop Sci. 2019;59(4):1411–23.
39. Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. Genet Res. 2000;75(2):249–52.
40. Clark SA, van der Werf J. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. In 2013. p. 321–30. Available from: http://link.springer.com/10.1007/978-1-62703-447-0_13
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 [Internet]. 2019 Mar 14;10(March):1–10. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.00189/full
42. Hayashi T, Iwata H. EM algorithm for Bayesian estimation of genomic breeding values. BMC Genet. 2010;11.
43. Meuwissen TH, Solberg TR, Shepherd R, Woolliams JA. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet Sel Evol. 2009;41(2):1–10.
44. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12.
45. De Los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics. 2009;182(1):375–85.

46. Zhou X, Carbonetto P, Stephens M. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. 2013;9(2).
47. Gianola D, Fernando RL, Stella A. Genomic-Assisted Prediction of Genetic Value with Semiparametric Procedures. Genetics. 2006;173(3):1761–76.
48. Gianola D, de los Campos G, González-Recio O, Long N, Okut H, Rosa GJM, et al. Statistical learning methods for genome-based analysis of quantitative traits. Proc 9th World Congr Genet Appl to Livest Prod. 2010;14:1–6.
49. Long N, Gianola D, Rosa GJM, Weigel KA. Application of support vector regression to genome-assisted prediction of quantitative traits. Theor Appl Genet. 2011;123(7):1065–74.
50. Gianola D, Okut H, Weigel KA, Rosa GJM. Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. BMC Genet. 2011;12:4–7.
51. Howard R, Carriquiry AL, Beavis WD. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3 Genes, Genomes, Genet. 2014;4(6):1027–46.
52. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 2013;193(2):327–45.
53. Heslot N, Yang HP, Sorrells ME, Jannink JL. Genomic selection in plant breeding: A comparison of models. Crop Sci. 2012;52(1):146–60.
54. Maltecca C, Parker KL, Cassady JP, Maltecca C, Parker KL, Cassady JP. Application of multiple shrinkage metho ds to genomic predictions. 2014;(2010):1777–87.
55. Bandeira e Sousa M, Cuevas J, de Oliveira Couto EG, Pérez-Rodríguez P, Jarquín D, Fritsche-Neto R, et al. Genomic-Enabled Prediction in Maize Using Kernel Models with Genotype × Environment Interaction. G3&#58; Genes|Genomes|Genetics [Internet]. 2017 Jun;7(6):1995–2014. Available from: http://g3journal.org/lookup/doi/10.1534/g3.117.042341
56. Cuevas J, Granato I, Fritsche-Neto R, Montesinos-Lopez OA, Burgueño J, Sousa MB, et al. Genomic-enabled prediction Kernel models with random intercepts for multi-environment trials. G3 Genes, Genomes, Genet. 2018;8(4):1347–65.
57. Bellot P, de los Campos G, Pérez-Enciso M. Can Deep Learning Improve Genomic Prediction of Complex Human Traits? Genetics [Internet]. 2018 Nov;210(3):809–19. Available from: http://www.genetics.org/lookup/doi/10.1534/genetics.118.301298
58. Montesinos-López OA, Martín-Vallejo J, Crossa J, Gianola D, Hernández-Suárez CM, Montesinos-López A, et al. 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(2):601–18.
59. Crossa J, Beyene Y, Semagn K, Pérez P, Hickey JM, Chen C, et al. Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes, Genomes, Genet. 2013;3(11):1903–26.
60. Pérez-Rodríguez P, Gianola D, González-Camacho JM, 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(12):1595–605.
61. Widstrom NW, Wiseman BR, McMillian WW. Resistance Among Some Maize Inbreds and Single Crosses to Fall Armyworm Injury1. Crop Sci. 1972;12(3):290.
62. Viana PA, Guimarães PEO. Maize resistance to the lesser cornstalk borer and fall armyworm in Brazil. In: Embrapa Milho e Sorgo-Artigo em anais de congresso (ALICE). In: INTERNATIONAL SYMPOSIUM ON INSECT RESISTANT MAIZE: recent advances and …; 1997.
63. Alvarez P, Branco J, Filho DM. Diallel crossing among miaze populations for resistance to fall armyworm. Sci Agric. 2002;59(4):731–41.
64. 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(2):662–7.
65. Kasozi LC. Genetic analysis and selection for weevil resistance in maize. 2013;(February):208.
66. Derera J, Pixley K V., Giga DP, Makanda I. Resistance of maize to the maize weevil: III. Grain weight loss assessment and implications for breeding. J Stored Prod Res [Internet]. 2014;59:24–35. Available from: http://www.sciencedirect.com/science/article/pii/S0022474X14000472
67. Zunjare R, Hossain F, Muthusamy V, Jha SK, Kumar P, Javaji C. Genetics of resistance to stored grain weevil

(Sitophilus oryzae L .) in maize. Cogent Food Agric [Internet]. 2015;8(1):1–9. Available from: http://dx.doi.org/10.1080/23311932.2015.1075934
68. Musundire L, Dari S, Derera J, Co-zimbabwe S, Wgt POB. Original Paper Open Access Genetic analysis of grain yield performance and weevil [Sitophilus zeamais (Motschulsky)] resistance in southern African maize hybrids. Maydica. 2015;60.
69. Butron A, Samayoa LF, Santiago R, Ordás B, Malvar RA. Genomics of Insect Resistance. In 2018. p. 163–83. Available from: http://link.springer.com/10.1007/978-3-319-97427-9_11
70. Drouaillet BE, Mendez CAR. Combinatorial aptitude and resistance to leaf damage of Spodoptera frugiperda (J . E . Smith) in maize germplasm native to Tamaulipas. 2018;9(1):81–93.
71. Hickey JM, Chiurugwi T, Mackay I, Powell W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat Genet. 2017;49(9):1297–303.
72. Akdemir D, Isidro-sánchez J. Design of training populations for selective phenotyping in genomic prediction. 2019;1–15.
73. Gosal SS, Wani SH, Editors. Accelerated Plant Breeding, Volume 1: Cereal Crops [Internet]. Springer N. Gosal SS, Wani SH, editors. Vol. 1. Cham: Springer International Publishing; 2020. Available from: http://link.springer.com/10.1007/978-3-030-41866-3
74. Sarinelli JM, Murphy JP, Tyagi P, Holland JB, Johnson JW, Mergoum M, et al. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theor Appl Genet [Internet]. 2019;132(0123456789):1247–61. Available from: https://doi.org/10.1007/s00122-019-03276-6
75. Cericola F, Jahoor A, Orabi J, Andersen JR, Janss LL, 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(1):1–20.
76. Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME. Training set optimization under population structure in genomic selection. Theor Appl Genet. 2015;128(1):145–58.
77. Andres RJ, Dunne JC, Samayoa LF, Holland JB. Enhancing Crop Breeding Using Population Genomics Approaches. 2020;
78. Sitonik C, Suresh LM, Beyene Y, Olsen MS, Makumbi D, Oliver K, 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 [Internet]. 2019;132(8):2381–99. Available from: https://doi.org/10.1007/s00122-019-03360-x
79. Nyaga C, Gowda M, Beyene Y, Muriithi WT, Makumbi D, Olsen MS, et al. Genome-wide analyses and prediction of resistance to mln in large tropical maize germplasm. Genes (Basel). 2020;11(1):1–16.
80. Gowda M, Das B, Makumbi D, Babu R, Semagn K, Mahuku G, et al. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor Appl Genet. 2015;128(10):1957–68.
81. Galli G, Lyra DH, Alves FC, Granato ÍSC, e Sousa MB, Fritsche-Neto R. Impact of phenotypic correction method and missing phenotypic data on genomic prediction of maize hybrids. Crop Sci. 2018;58(4):1481–91.
82. Molenaar H, Boehm R, Piepho HP. Phenotypic selection in ornamental breeding: It's better to have the BLUPs than to have the BLUEs. Front Plant Sci. 2018;871(November):1–14.
83. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica. 2008;161(1–2):209–28.
84. Piepho HP, Möhring J. Selection in cultivar trials—Is it ignorable? Crop Sci. 2006;46(1):192–201.
85. Dramadri IO, Nkalubo ST, Kelly JD. Identification of QTL Associated with Drought Tolerance in Andean Common Bean. Crop Sci [Internet]. 2019 May;59(3):1007–20. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/0/0/cropsci2018.10.0604
86. Williams WP, Buckley PM, Davis FM. Combining ability for resistance in corn to fall armyworm and southwestern corn borer. Crop Sci. 1989;29(4):913–5.
87. Sodedji FAK, Kwemoi DB, Asea G, Kyamanywa S. Response of provitamin-A maize germplasm to storage weevil Sitophilus zeamais (Motschulsky). Int J Agron Agri R Int J Agron Agric Res. 2016;9(5):1–13.
88. Kasozi LC, Derera J, Tongoona P, Zziwa S, Foundation MG, Box PO. Comparing the Effectiveness of the " weevil warehouse " and " laboratory bioassay " as Techniques for Screening Maize Genotypes for Weevil Resistance. 2018;6(4):170–7.
89. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association

mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.
90. R Development Core Team R. R: A Language and Environment for Statistical Computing [Internet]. Team RDC, editor. R Foundation for Statistical Computing. R Foundation for Statistical Computing; 2011. p. 409. (R Foundation for Statistical Computing; vol. 1). Available from: http://www.r-project.org
91. Bates DM, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw [Internet]. 2015;67:1–48. Available from: http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf
92. de los Campos G, Pérez-Rodríguez P. BGLR: Bayesian generalized linear regression. R Packag version. 2016;1(5).
93. Park T, Casella G. The Bayesian Lasso. J Am Stat Assoc. 2008;103(482):681–6.
94. de los Campos G, Pérez P, Vazquez AI, Crossa J. Genome-Enabled Prediction Using the BLR (Bayesian Linear Regression) R-Package. In: Methods in Molecular Biology [Internet]. 2013. p. 299–320. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23756904%5Cnhttp://link.springer.com/10.1007/978-1-62703-447-0
95. 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(2):106.
96. Covarrubias-pazaran G. Genetic analysis using the sommer package. 2016;1–16.
97. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008;178(3):1709–23.
98. Gilmour AR, Thompson R, Cullis BR. Average Information REML : An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models Author (s): Arthur R . Gilmour , Robin Thompson and Brian R . Cullis Published by : International Biometric Society Stable. 1995;51(4):1440–50.
99. Searle SR. Applying the EM algorithm to calculating ML and REML estimates of variance components. 1993;1–9.
100. Tunnicliffe GW. On the Use of Marginal Likelihood in Time Series Model Estimation. JRSS. 1989;51(1):15–27.
101. Rodriguez PP, Gianola D. brnn: Bayesian regularization for feed-forward neural networks. R Packag version 06. 2016;
102. MacKay D. Bayesian model comparison and backprop nets. Proc Int Conf Neural Networks [Internet]. 1992;3(IEEE):839–46. Available from: http://www.cs.toronto.edu/~mackay/nips91.pdf
103. Dan Foresee F, Hagan MT. Gauss-Newton approximation to bayesian learning. IEEE Int Conf Neural Networks - Conf Proc. 1997;3:1930–5.
104. Nguyen D, Widrow B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. IJCNN Int Jt Conf Neural Networks. 1990;21–6.
105. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome J. 2011;4(3):250.
106. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112–8.
107. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 2013;14(6):R55.
108. Peiffer JA, Flint-Garcia SA, De Leon N, McMullen MD, Kaeppler SM, Buckler ES. The genetic architecture of maize stalk strength. PLoS One. 2013;8(6):e67066.
109. Lorenz AJ, Smith KP, Jannink JL. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 2012;52(4):1609–21.
110. Arruda MP, Brown PJ, Lipka AE, Krill AM, Thurber C, Kolb FL. Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. Plant Genome. 2015;8(3):1–12.
111. Foiada F, Westermeier P, Kessel B, Ouzunova M, Wimmer V, Mayerhofer W, 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(5):875–91.
112. Riedelsheimer C, Technow F, Melchinger AE. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics. 2012;13(1).
113. Azodi CB, McCarren A, Roantree M, Campos G de los, Shiu S-H. Benchmarking algorithms for genomic prediction of complex traits. bioRxiv [Internet]. 2019;614479. Available from: https://www.biorxiv.org/content/10.1101/614479v1.full
114. Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. Plant Genome J. 2011;4(2):132.

115. Lenz PRN, Beaulieu J, Mansfield SD, 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 Genomics. 2017;18(1):1–17.
116. Edwards SMK, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, et al. The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet [Internet]. 2019;132(7):1943–52. Available from: https://doi.org/10.1007/s00122-019-03327-y
117. Wang W, Cao XH, Miclău M, Xu J, Xiong W. The Promise of Agriculture Genomics. Int J Genomics. 2017;2017.
118. Spindel J, Iwata H. Genomic selection in rice breeding. Rice Genomics, Genet Breed. 2018;473–96.
119. Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb). 2014;112(April 2013):48–60.
120. Ou JH, Liao CT. Training set determination for genomic selection. Theor Appl Genet [Internet]. 2019;(Lmm). Available from: https://doi.org/10.1007/s00122-019-03387-0
121. Mangin B, Rincent R, Rabier CE, Moreau L, Goudemand-Dugue E. Training set optimization of genomic prediction by means of EthAcc. PLoS One. 2019;14(2):1–21.
122. Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, et al. Genome-based prediction of testcross values in maize. Theor Appl Genet. 2011;123(2):339–50.
123. Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 2014;54(4):1476–88.
124. 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(5):2091–9.
125. Mayor PJ, Bernardo R. Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F2 populations. Crop Sci. 2009;49(5):1719–25.