

RESEARCH ARTICLE

Open Access



# Accuracies of univariate and multivariate genomic prediction models in African cassava

Uche Godfrey Okeke<sup>1</sup>, Deniz Akdemir<sup>1</sup>, Ismail Rabbi<sup>2</sup>, Peter Kulakow<sup>2</sup> and Jean-Luc Jannink<sup>1,3\*</sup>

## Abstract

**Background:** Genomic selection (GS) promises to accelerate genetic gain in plant breeding programs especially for crop species such as cassava that have long breeding cycles. Practically, to implement GS in cassava breeding, it is necessary to evaluate different GS models and to develop suitable models for an optimized breeding pipeline. In this paper, we compared (1) prediction accuracies from a single-trait (uT) and a multi-trait (MT) mixed model for a single-environment genetic evaluation (Scenario 1), and (2) accuracies from a compound symmetric multi-environment model (uE) parameterized as a univariate multi-kernel model to a multivariate (ME) multi-environment mixed model that accounts for genotype-by-environment interaction for multi-environment genetic evaluation (Scenario 2). For these analyses, we used 16 years of public cassava breeding data for six target cassava traits and a fivefold cross-validation scheme with 10-repeat cycles to assess model prediction accuracies.

**Results:** In Scenario 1, the MT models had higher prediction accuracies than the uT models for all traits and locations analyzed, which amounted to on average a 40% improved prediction accuracy. For Scenario 2, we observed that the ME model had on average (across all locations and traits) a 12% improved prediction accuracy compared to the uE model.

**Conclusions:** We recommend the use of multivariate mixed models (MT and ME) for cassava genetic evaluation. These models may be useful for other plant species.

## Background

Cassava (*Manihot esculenta* Crantz) [1] is a staple food for over 700 million people in Africa, South America and Asia [2]. Cassava also has immense industrial potential. White cassava starch is easy to extract and contains low levels of fat (~ 1.5%), protein (~ 0.6%) and phosphorus (~ 4%), which are desirable features for the food industry [3, 4]. Given the issues of climate change and fast-growing populations in countries that rely heavily on cassava, the need for rapid genetic improvement of cassava is critical. To enable this, genetic evaluation protocols based on best linear unbiased prediction (BLUP) analysis [5, 6] and selection on a merit index [7, 8] have been recommended to maximize gain from selection [9].

Genomic selection (GS) [10] offers crop species such as cassava a tremendous opportunity for accelerated genetic gains [11] by using whole-genome single nucleotide polymorphisms (SNPs) scored with methods such as genotyping-by-sequencing (GBS) [12]. These whole-genome SNPs could be sufficiently dense to be in linkage disequilibrium with most quantitative trait loci (QTL) that affect traits of interest. Using GS, selection is imposed at these QTL without actually identifying the QTL or the functional polymorphisms [10]. In addition, these SNPs will help to better track relatedness due to Mendelian sampling [13], which leads to improved selection accuracies especially when pedigree records are not available [14].

## GS models for plant genetic evaluation

Genetic evaluation [9] starts by accurately estimating the genetic value of an individual for a wide range of traits using its own performance records, progeny performance records, records from relatives, or a combination of the three [15]. Usually this estimation is carried out by using

\*Correspondence: jeanluc.work@gmail.com

<sup>3</sup> Robert W. Holley Center for Agriculture and Health, USDA-ARS, Tower Road, Ithaca, NY 14853-2901, USA

Full list of author information is available at the end of the article

the univariate single environment one-step (uT) BLUP model [16] to obtain estimated breeding values (EBV) for one trait at a time. In plant and animal breeding, breeders usually select multiple traits at the same time that are often genetically correlated, with correlations that can range from weak to strong. The uT model for traits that are measured in a single environment assumes zero genetic and residual covariances between these traits such that information from other traits is not used when obtaining EBV of the evaluated individuals for the traits in the analysis. However, the optimal estimation procedure is to combine information from multiple trait records and obtain EBV using the multi-trait single environment one-step BLUP model (MT) [17, 18]. The MT model does not assume zero genetic and residual covariances but provides an estimate for these and also uses this information when obtaining individual EBV for the traits in the analysis. The MT model has several advantages over the uT model including:

1. Higher prediction accuracies for individual traits in the model because of more information (direct or indirect) and better connectedness of the data [19], especially when traits with varying heritabilities are analyzed jointly. This is true if the genetic correlations in the model are significant or substantial with low error correlations.
2. Simplified index selection because optimal weight factors for the total merit index are the economic weights [19].
3. Procedures for obtaining genetic and residual covariances and incorporating these in EBV estimates for across-location, -country or -region evaluations [20–22].

While MT models have clear advantages over uT models, they require the estimation of additional parameters (i.e., genetic and error covariances), which will affect accuracies of EBV. The number of additional parameters increases as the number of traits increases. For large models, many additional parameters can lead to convergence problems in the analysis. Lastly, an appreciable amount of data is required to get good estimates of these additional parameters.

In most plant breeding programs, genotypes are evaluated in multi-environment trials (MET) usually at advanced stages of breeding. The goal is to sample the influence on selection candidates of the range of environments for which varieties will be targeted. Addressing the problem of the analysis of MET brings into focus another potential use for MT models [23]. Here, phenotypes of the same trait, but measured at different locations are parameterized as different traits in the MT model [24],

producing what we call a multivariate single-trait multi-environment BLUP (ME) model. Like the MT model, the ME model estimates genetic covariances of the same trait measured in multiple environments, which may lead to more accurate estimates of individual EBV for the trait in all the environments in which data are recorded. For the ME models used for modelling MET data, residual covariances are set to zero reflecting the assumption that no mechanism generates error covariances of a trait measured in different environments [20]. In contrast, the typical univariate BLUP model for modelling MET data, termed the univariate multi-environment one-step model (uE), fits a multi-kernel mixed model with the genotypic effect as one kernel and the genotype-by-environment (GxE) effect as the second kernel and maybe environment as the third kernel [25]. This model yields a GxE variance for a MET and individuals can be ranked on their performance in different locations. Different variants of the ME model have been used for modeling environment covariance structures in plant [26–29] and in animal breeding [30, 31]. Genetic covariances from the ME model offer a convenient tool to assess the impact of GxE on a trait and relate directly to the extent of GxE at all locations in the analysis. A low genetic correlation between EBV for a trait at different locations from the ME model indicates a high GxE impact on that trait [9, 32–35].

To select the GS model for a practical cassava breeding program, it is necessary to compare models that will be efficient at various stages of cassava breeding with MET data. Finally, fitting multivariate BLUP models is not trivial. Even with software that, in principle, can fit these models, model convergence is not guaranteed and may require several attempts [36–38] and univariate models may be more practical if the benefits of the multivariate models are not substantial.

The objectives of this paper are to (1) compare multi-trait (MT) and single trait (uT) mixed models for single environment data using cross-validation, and (2) compare the multivariate multi-environment (ME) model to a single-trait multi-environment (uE) model using cross-validation and assessing the GxE impact on the traits analyzed via genetic covariances from the ME model fit.

## Methods

### Cassava phenotype data

We used historical phenotype data from different trials that were conducted for the cassava breeding program at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. The genetic gain population represents a collection of clones selected from the 1970s to 2007 within this program [39, 40]. Some of these clones are West African landraces and some are of East African origin. Clones in the genetic gain population have gone

through advanced stages of the cassava breeding process up to on-farm variety testing trials. The data used in our analysis include data that were collected on clonal evaluation trials (CET), which are augmented design trials with typically two known checks and unreplicated plots with five plants. These data were collected from three target locations in Nigeria: Ibadan (7.40°N, 3.90°E), Mokwa (9.3°N, 5.0°E), and Ubiaja (6.66°N, 6.38°E). These locations represent regions, which encompass about 35% of the cassava production in Nigeria. Datasets were collected from 2000 to 2015 and included trials with most of the 739 clones of the genetic gain population. Six target agronomic traits were used in the analysis including seedling vigor (VIGOR), number of storage roots per plot at harvest (RTNO), fresh weight of harvested roots expressed in tons per hectare (T/ha) (FYLD), percent dry matter (DM) of storage roots, which measures root dry weight as the % of the root fresh weight, plot mean cassava mosaic disease severity (MCMDS), which is rated on a scale from 1 (no symptoms) to 5 (extremely severe), and plot mean cassava green mite (MCGM) severity, which is rated on a scale from 1 (no symptoms) to 5 (extremely severe). Cassava mosaic disease is caused by a *Begomovirus* that belongs to the Geminiviridae family, and is carried and transmitted by the whitefly *Bemisia tabaci*. The cassava green mite is *Mononychellus tanajoa* [41]. These traits are target traits used in the selection index for selection decisions in the IITA cassava breeding program. Phenotype data metrics are in Table 1. All trait records were plot averages for both clonal accessions and checks. All checks were included in the analysis.

#### Cassava genotyping data

DNA from 739 clones from the 2013 genetic gain trial at IITA was extracted by using DNeasy Plant Mini kits (Qiagen) and was quantified using PicoGreen. Genotyping-by-sequencing (GBS) was used for genotyping [12] these clones. Six 95-plex and one 75-plex *ApeKI* libraries were constructed and sequenced on Illumina HiSeq, one lane per library. SNPs were called from the sequence data using the TASSEL pipeline version 4.0 [42], using an alignment to the *Manihot esculenta* version 6 reference genome [43]. Average sequencing depth for polymorphic loci was 5×. Individuals with more than 80% and SNPs with more than 60% missing calls were removed. SNP genotyping data were converted to numeric genotypes (0, 1, 2) and missing genotyping data were imputed using a LASSO regression method (Ariel Chan, personal communication, 2014) that was implemented using the R *glmnet* package [44]. The resulting dataset was rounded to obtain numerical genotypes (0, 1, 2) and consisted of 183,201 SNPs scored in 739 clones.

**Table 1 Cassava phenotype means and standard deviations (in brackets) at three locations: Ubiaja, Mokwa, and Ibadan**

	Ubiaja	Mokwa	Ibadan
Number of records	7806	5345	5579
Number of clones	739	573	691
VIGOR (1–3)	6.51 (1.12)	6.52 (0.93)	6.11 (1.23)
RTNO (No./plot)	31.71 (17.30)	37.05 (21.76)	37.87 (23.91)
FYLD (T/ha)	12.61 (7.70)	16.51 (9.54)	15.84 (10.72)
DM (%)	31.95 (6.42)	29.01 (6.38)	30.8 (6.79)
MCMDS (1–5)	1.59 (0.93)	1.21 (0.57)	2.14 (1.01)
MCGM (1–5)	3.56 (0.97)	2.99 (0.67)	3.00 (0.85)

VIGOR seedling vigor, RTNO number of storage roots per plot at harvest, FYLD fresh weight of harvested roots in tons per hectare, DM percentage dry matter of storage roots, MCMDS plot mean cassava mosaic disease severity, and MCGM plot mean cassava green mites severity

#### Statistical analysis

We structured the cassava phenotype data described above into two types of data that are commonly used in most plant breeding programs. The first set was achieved by pooling data from multiple years at specific locations (multi-year trials data). We termed this scenario the single-environment genetic evaluation (Scenario 1). The resulting prediction accuracies from this dataset were assessed for the three locations. The second scenario was achieved by using data from multiple locations and years (MET) but, in this case, location-specific information was extracted by modeling G×E interactions. We termed this scenario the multi-environment genetic evaluation (Scenario 2) and its goal was to assess the impact of G×E and determine the best way to fit it while also using information from correlated environments.

#### Pseudo-true genetic values for prediction accuracy computations

To validate the models in this study, first we defined a univariate single-trait mixed model for each trait at each location separately (to preserve the variation embedded in each location) using an identity covariance matrix among clone effects, which assumes no relationship among the clones. The univariate mixed model was as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

with  $\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{I})$  and  $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$ , where  $\mathbf{y}$  is a vector of observations,  $\mathbf{b}$  is a vector of fixed effects with the design matrix  $\mathbf{X}$  (relating observations to fixed effects, in this case including grand mean and a nested effect of trial-within-year and the ratio of plants harvested to number planted);  $\mathbf{u}$  is a vector of clonal genetic effects with the design matrix  $\mathbf{Z}$  (relating observations to

clones). This model was fit using the *lmer* function in the R lme4 package [45] and resulting BLUP values ( $\hat{u}$ ), which we refer to as estimated genotypic values (EGV), were used as pseudo-true genetic effects to compute prediction accuracy, as commonly used in the plant breeding literature [41, 46, 47].

### GS models for Scenario 1

We defined two mixed models that were fitted as follows.

#### The single-trait mixed model (uT)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (2)$$

with  $\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{K})$  and  $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$ , where  $\mathbf{y}$  is the response vector of a trait for a given location,  $\boldsymbol{\beta}$  is the vector of fixed effects with the design matrix  $\mathbf{X}$  (with components as in Model 1 above for each location and trait);  $\mathbf{u}$  is the vector of random additive genomic effects with the design matrix  $\mathbf{Z}$  (relating trait values to clones) and  $\mathbf{K}$  is the additive genomic relationship matrix generated from SNPs as in method 1 of VanRaden [48] implemented in preGSf90 [49].

#### The multitrait mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (3)$$

where  $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \mathbf{y}'_3, \dots, \mathbf{y}'_d)'$ ,  $\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_d)'$  and  $\mathbf{e} = (\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3, \dots, \mathbf{e}'_d)'$ , and  $\mathbf{y}$  is the response vector of  $d$  traits (six core traits described above),  $\mathbf{X}$  is a design matrix for fixed effects  $\boldsymbol{\beta}$  (with components as in Model 1 above for each location and trait) and  $\mathbf{Z}$  is a design matrix for random genetic effects  $\mathbf{u}$ . Following a multivariate normal distribution ( $N_m$ ), the marginal density of  $\mathbf{y}$  is:

$$(\mathbf{y}|\boldsymbol{\beta}, \mathbf{R}, \mathbf{G}) \sim N_m(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}) \quad (4)$$

and  $\mathbf{V} = \mathbf{Z}(\mathbf{G} \otimes \mathbf{K})\mathbf{Z}^T + \mathbf{R} \otimes \mathbf{I}$ .

The matrices  $\mathbf{G}$  and  $\mathbf{R}$  are  $d \times d$  symmetric unstructured genomic and error covariance matrices respectively,  $\mathbf{K}$  remains the additive genomic relationship matrix for  $n$  clones generated from SNPs as above,  $\mathbf{I}$  is an identity matrix.

Models (2) and (3) were fitted separately for each location Ubiaja, Mokwa and Ibadan, respectively, allowing the error (co)variances associated with these locations to be distinct. Note also that, in these models, genotype-by-location effects are confounded with the main genotype effects such that variance components may change between locations. The effects of years and trials were fixed because emphasis was on location effects since these locations represented different production regions and we sought to capture consistent effects of these locations. In contrast, year effects are variable and by definition not consistent. Also following practice in cassava

breeding [46, 47], multiple observations of one clone were not considered as repeated measures. Although these subjects were clones, data were collected from distinct individuals and thus they are independent. Hence, these measurements were treated as samples of clones and should lead to better precision in the prediction of breeding values.

### GS models for Scenario 2

We also defined two mixed models with the aim of modeling genotype-by-environment interaction effects as follows.

#### The compound symmetric multi-environment model (uE)

Here, we describe the uE model, first how it is fit and then we show its compound symmetry structure. The model is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}, \quad (5A)$$

with  $\mathbf{y} = (\mathbf{y}'_a, \mathbf{y}'_b, \mathbf{y}'_c)'$ ,  $\mathbf{e} = (\mathbf{e}'_a, \mathbf{e}'_b, \mathbf{e}'_c)'$ ,

$$\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{K}), \quad \mathbf{w} \sim N(0, \sigma_w^2 \mathbf{I}_3 \otimes \mathbf{K}), \quad \mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I})$$

where  $\mathbf{y}$  is the vector of a trait at locations a, b and c (corresponding to Ubiaja, Mokwa and Ibadan),  $\boldsymbol{\beta}$  is the vector of fixed effects with the design matrix  $\mathbf{X}$  (relating observations to fixed effects as in Model 1);  $\mathbf{u}$  is the vector of random additive genomic effects with the design matrix  $\mathbf{Z}_1$  (relating trait values to clones),  $\mathbf{w}$  is the vector of random clone-by-location interaction effects with the design matrix  $\mathbf{Z}_2$ , which is  $\text{diag}(\mathbf{Z}_a, \mathbf{Z}_b, \mathbf{Z}_c)$  that relates records to clones in locations a, b and c, respectively. For the  $c$ th location, a column of  $\mathbf{Z}_c$  may be all 0s if the clone represented by the column was not evaluated in that location.  $\sigma_u^2 \mathbf{K}$  is the additive genomic relationship matrix generated from SNPs as above,  $\mathbf{I}$  is an identity matrix and  $\mathbf{I}_3$  is a  $3 \times 3$  identity matrix. In this model, the genomic value of a clone for the  $c$ th location was estimated as  $\hat{\mathbf{u}} + \hat{\mathbf{w}}_c$ . A more complete account of the error terms would have included clone-by-year and clone-by-location-by-year terms in the model. While such a model would have characterized the error in more detail, we believe that the obtained improvement of within-location estimation would have been marginal. Model 5 implies a compound symmetry structure [50] as described below.

#### The uE model defined as a compound symmetry (CS) covariance structure model

Using the same symbols as above, we defined the uE model with a CS covariance structure as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_2\mathbf{w} + \mathbf{e}, \quad (5B)$$

with  $\mathbf{y} = (\mathbf{y}'_a, \mathbf{y}'_b, \mathbf{y}'_c)'$  and  $\mathbf{e} = (\mathbf{e}'_a, \mathbf{e}'_b, \mathbf{e}'_c)'$ ,

$$\mathbf{w} \sim N\left(0, \left(\sigma_u^2 + \sigma_w^2\right) \Phi \otimes \mathbf{K}\right),$$

$$\Phi = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \quad \text{with} \quad \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_w^2},$$

$$\mathbf{V} = (\sigma_u^2 + \sigma_w^2)\mathbf{Z}_2(\Phi \otimes \mathbf{K})\mathbf{Z}_2' + \sigma_e^2\mathbf{I}.$$

The genomic effect from this CS model for the *c*th location  $\widehat{\mathbf{w}}_c$  is equal to  $\widehat{\mathbf{u}} + \widehat{\mathbf{w}}_c$  from the uE model. The  $\mathbf{Z}_2$  matrix is the same as in the uE model. Compared to the ME model described below, which replaces  $\Phi$  with an unstructured covariance matrix with nine parameters (six for genetic and three for error (co)variances, respectively), the CS model has three parameters including  $\sigma_{u+w}^2$  (equivalent to  $\sigma_u^2 + \sigma_w^2$  in the uE model),  $\sigma_e^2$  and  $\rho$ . For any trait for which the CS covariance structure best fits the data, it is expected that model uE will provide more accurate GEBV than the ME model, which will overfit the data. Furthermore, the uE model defined here assumes a homogeneous variance across locations a, b and c. Although a CS model with heterogeneous variances can be fit, this was not the case for the uE model. This assumption is incorrect if there are significant heterogeneous variances across these locations. In such a case, the ME model should provide more accurate breeding values.

**The multivariate multi-environment (ME) model**

We fit the ME model in a single-step procedure using the following model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{6}$$

with  $\mathbf{y} = (\mathbf{y}'_a, \mathbf{y}'_b, \mathbf{y}'_c)'$ ,  $\mathbf{u} = (\mathbf{u}'_a, \mathbf{u}'_b, \mathbf{u}'_c)'$  and  $\mathbf{e} = (\mathbf{e}'_a, \mathbf{e}'_b, \mathbf{e}'_c)'$ , where  $\mathbf{y}$  is the vector of a trait in locations a, b and c (corresponding to Ubiaja, Mokwa and Ibadan) recorded for *n* clones, the  $\mathbf{X}$  and  $\mathbf{Z}$  design matrices are block diagonal matrices represented as *diag*( $\mathbf{X}_a, \mathbf{X}_b, \mathbf{X}_c$ ) and *diag*( $\mathbf{Z}_a, \mathbf{Z}_b, \mathbf{Z}_c$ ), respectively allowing for missing clones and observations.  $\mathbf{X}$  is a design matrix for fixed effects  $\boldsymbol{\beta}$  (with components as in Model 1) and  $\mathbf{Z}$  is a design matrix for random genomic effects  $\mathbf{u}$ . Following a multivariate normal distribution ( $N_m$ ), the marginal density of  $\mathbf{y}$  is:

$$(\mathbf{y}|\boldsymbol{\beta}, \mathbf{R}, \mathbf{G}) \sim N_m(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \tag{7}$$

and  $\mathbf{V} = \mathbf{Z}(\mathbf{G} \otimes \mathbf{K})\mathbf{Z}^T + \mathbf{R} \otimes \mathbf{I}$ .

Given that *d* is the number of locations being analyzed,  $\mathbf{G}$  is a *d* × *d* symmetric and unstructured genomic covariance matrix, while  $\mathbf{R}$  is a *d*-dimensional diagonal error covariance matrix,  $\mathbf{K}$  remains the additive genomic

relationship matrix for *n* clones generated from SNPs as above, and  $\mathbf{I}$  is an identity matrix. In this model, the error covariance matrix  $\mathbf{R}$  is diagonal, thus allowing heterogeneous variances for a trait at different locations but the covariances are fixed to zero following the assumption that no mechanism generates error covariances of a trait measured in multiple environments.

Estimation of the parameters in Models (2), (3), (5) and (6) were performed using the average information (AI) REML procedure implemented in the *airemlf90* program [49] from which the best linear unbiased estimator (BLUE) of fixed effects and the BLUP of random effects were obtained by solving the mixed model equations (MME) [5, 6]. Custom R-scripts were used for cross-validation.

**Comparison of prediction accuracies**

We used a fivefold cross-validation scheme with 10 repeats for comparisons between the univariate and multivariate models. We used the same folds for the models in each scenario. Hereafter, we refer to predicted BLUP or genomic effects from these models as genomic EBV (GEBV). Prediction accuracies were calculated as a correlation of the validation fold GEBV to their corresponding EGV.

**Results**

**Scenario 1: MT versus uT model**

In Scenario 1, we observed that the prediction accuracies of the MT model were higher than those of the uT models for most traits and locations in our analysis (Table 2). On average (across traits and locations), the MT model yielded prediction accuracies that were 59% higher for VIGOR, 43% for RTNO, 27% for DM, 40% for MCMDs, 55% for FYLD and 18% for MCGM compared to the uT model. Averaged across traits and locations, the MT models were 40% more accurate than the uT models.

**Scenario 2: ME versus uE model**

In Scenario 2, we observed different patterns of prediction accuracies of the uE and ME models. The ME model yielded higher prediction accuracies for DM and MCMDs at all locations. On average (across locations), the uE model resulted in prediction accuracies that were 2% better for VIGOR and 1% for RTNO, while the ME model resulted in prediction accuracies that were 32% better for DM, 24% for MCMDs, 5% for FYLD, and 4% for MCGM. Prediction accuracy of the ME model was 12% higher than that of the uE model averaged across all traits and locations in the model. Trait correlations from the ME model representing the expected correlated responses to selection ranged from 0.21 to 0.66 for VIGOR, 0.36 to 0.54 for RTNO, 0.57 to 0.81 for DM, 0.68

**Table 2 Cross-validation prediction accuracies for GS models in Scenarios 1 and 2**

Single-trait single environment (uT)				Multi-trait (MT)		
Trait	Ubiaja	Mokwa	Ibadan	Ubiaja	Mokwa	Ibadan
GS Scenario 1						
VIGOR	0.24 (0.02)	0.16 (0.03)	0.42 (0.02)	0.41 (0.02)	0.31 (0.03)	0.58 (0.01)
RTNO	0.32 (0.02)	0.17 (0.02)	0.37 (0.02)	0.46 (0.02)	0.24 (0.03)	0.53 (0.02)
DM	0.60 (0.01)	0.33 (0.02)	0.51 (0.01)	0.72 (0.01)	0.46 (0.02)	0.64 (0.02)
MCMDS	0.49 (0.01)	0.37 (0.03)	0.59 (0.01)	0.69 (0.02)	0.60 (0.04)	0.74 (0.01)
FYLD	0.41 (0.02)	0.11 (0.03)	0.40 (0.01)	0.58 (0.02)	0.30 (0.03)	0.55 (0.02)
MCGM	0.38 (0.01)	0.50 (0.02)	0.58 (0.01)	0.48 (0.01)	0.56 (0.02)	0.69 (0.01)
Single-trait multi-environment (uE)				Multi-environment (ME)		
Trait	Ubiaja	Mokwa	Ibadan	Ubiaja	Mokwa	Ibadan
GS Scenario 2						
VIGOR	0.22 (0.01)	0.10 (0.01)	0.37 (0.01)	0.24 (0.01)	0.12 (0.02)	0.32 (0.01)
RTNO	0.29 (0.01)	0.11 (0.01)	0.34 (0.01)	0.27 (0.02)	0.13 (0.01)	0.34 (0.02)
DM	0.49 (0.01)	0.20 (0.02)	0.40 (0.01)	0.60 (0.01)	0.35 (0.01)	0.50 (0.01)
MCMDS	0.40 (0.01)	0.23 (0.01)	0.53 (0.01)	0.48 (0.01)	0.39 (0.02)	0.57 (0.01)
FYLD	0.38 (0.01)	0.10 (0.02)	0.35 (0.01)	0.37 (0.01)	0.12 (0.03)	0.36 (0.02)
MCGM	0.31 (0.01)	0.48 (0.01)	0.56 (0.01)	0.38 (0.02)	0.47 (0.01)	0.55 (0.01)

Prediction accuracies for MT and uT models (GS Scenario 1) and for ME and uE models (GS Scenario 2)

The numbers in brackets are standard deviations for cross-validation repeat cycles

VIGOR seedling vigor, RTNO number of storage roots per plot at harvest, FYLD fresh weight of harvested roots in tons per hectare, DM percentage dry matter of storage roots, MCMDS plot mean cassava mosaic disease severity, MCGM plot mean cassava green mites severity

to 0.87 for MCMDS, 0.31 to 0.52 for FYLD and 0.24 to 0.53 for MCGM. Thus, genetic effects for MCMDS and DM were more consistent across locations than those for the other traits.

**Discussion**

**Scenario 1: MT versus uT model**

Some studies reported comparisons between MT and uT genomic prediction models using simulated data or real datasets [51–53]. Based on simulated datasets, Guo et al. [53] and Calus et al. [52] reported similar accuracies between MT and uT models with accuracies of the MT models for lowly heritable traits being slightly higher when the genomic correlations between the traits increased. Using Holstein and Jersey breed datasets from the US Dairy National evaluation program, VanRaden et al. [51] also reported similar accuracies between MT and uT models for all the traits analyzed. However, for several traits, they obtained accuracies that were slightly higher with the uT model than with the MT model. For highly heritable traits and especially if complete phenotypic data are available for these, accuracies obtained by the MT model are not clearly better than those obtained by the uT model [53]. Improvement in prediction accuracies with the MT model is accrued mostly for lowly heritable traits when they are analyzed jointly with highly

heritable traits that have medium to high genetic correlations and low residual correlations [52, 53]. Our results were consistent with those of other studies [52, 53] since, in our analysis, our MT model yielded higher accuracies for most traits and locations and resulted from the joint analysis of low and high heritability traits. Most of the genetic correlations between traits at all locations in the MT models were significant (substantial) with low error correlations (not shown). These contributed to the increased prediction accuracies observed for the MT models compared to those of the uT models. Substantial increases in prediction accuracies of the MT models were observed for VIGOR, RTNO and FYLD, which had mostly moderate to high genetic correlations with other traits at all locations although their heritabilities were mostly low.

For parental selections in specific locations, we recommend the use of MT models but to confirm this conclusion, further studies on the selection gains obtained by applying these models are necessary.

**Scenario 2: ME versus uE model**

Comparisons between different ME and uE genomic prediction models have been reported in plant breeding literature [54–56]. Burgueno et al. [29] conducted extensive modeling for multi-environment trials using

pedigree and genomic markers and incorporated many covariance structures including diagonal, factor analytic (FA), identity and unstructured covariances for both the genomic and error components in their models. They observed that prediction accuracies of a genomic ME model with a diagonal genomic covariance structure and a diagonal error covariance structure ( $ME_{D-D}$ ) were higher than that of a genomic ME model with a FA genomic covariance structure and diagonal error covariance ( $ME_{FA-D}$ ) for most of the locations measured in their analysis based on a cross-validation scheme (CV1) [29]. This  $ME_{D-D}$  is a univariate model with fewer parameters but it can be compared to our uE model. Although the uE model in our study assumed identical genomic and error variances for all locations analyzed, total phenotypic variance was partitioned into direct clonal genomic, clone-by-location interaction and error variance components. Hence, effects due to clones and clone-by-location interaction were combined to generate location-specific GEBV, which can be compared to location-specific GEBV obtained in the  $ME_{D-D}$  model. Our results were in line with this study for the traits VIGOR and RTNO at all locations with the uE model yielding higher prediction accuracies than the ME models and differed for the traits DM and MCMDS at all locations with the ME models yielding higher prediction accuracies. However, on average across locations and traits, prediction accuracies of the ME models were higher.

To further understand the strength of the impact of GxE interaction on the cassava core traits analyzed in our study, we used information from the proportion of total variation explained by clone and clone-by-location effects from the uE model (Table 3). From the total variation explained by SNPs, the effect of clone-by-location interaction was approximately 30% for VIGOR, 48% for RTNO, 12% for DM, 15% for MCMDS, 56% for FYLD

**Table 3 Proportion of explained variance by clone and clone-by-location effects**

Trait	Variance explained by effect (%)	
	Clone-by-location	Clone
VIGOR	4.52	10.70
RTNO	8.70	9.54
DM	4.59	33.52
MCMDS	10.00	58.90
FYLD	13.01	10.11
MCGM	7.00	8.33

These estimates are from effects based on genotyping by sequencing markers from the uE model

VIGOR seedling vigor, RTNO number of storage roots per plot at harvest, FYLD fresh weight of harvested roots in tons per hectare, DM percentage dry matter of storage roots, MCMDS plot mean cassava mosaic disease severity, MCGM plot mean cassava green mites severity

and 46% for MCGM. These proportions show strong clone-by-location interactions for FYLD, RTNO, MCGM and VIGOR but weak interactions for DM and MCMDS. In addition, the genetic correlations between the three locations for DM and MCMDS were relatively high ranging from 57 to 81 and 68 to 87%, respectively (Table 4), supporting our findings. These high correlations revealed that cassava DM and MCMDS were repeatable across the locations in our study, which suggests that genotypes selected for these traits will perform comparably across locations. From the genetic correlations in Table 4, improvement for RTNO and FYLD at Ubiaja will result in a correlated response of about 50% for these traits at Mokwa and about 35% at Ibadan. The low predicted correlated responses confirm that the environment had a higher impact on RTNO, FYLD, VIGOR and MCGM, thus improving these traits is more challenging. This makes a case for decentralized breeding especially for yield component traits. Breeding for good varieties that combine these core traits may be targeted towards specific locations or groups of locations with specific genotypes selected for these locations.

The ME model exploits the positive genomic correlations captured in its G matrix for prediction. The major difference between the prediction accuracies obtained by the ME and uE models were mainly due to the fact that the ME model accounted for genetic covariances when generating GEBV since genetic variances from both models were similar. Genetic covariances from the ME models are a reflection of the GxE interactions for the trait of interest and ME breeding values capture both additive

**Table 4 Genetic correlations from the multi-environment analysis**

Trait	Location	Ubiaja	Mokwa
VIGOR	Mokwa	0.39 (0.02)	
	Ibadan	0.66 (0.02)	0.21 (0.03)
RTNO	Mokwa	0.54 (0.08)	
	Ibadan	0.36 (0.08)	0.38 (0.08)
DM	Mokwa	0.57 (0.00)	
	Ibadan	0.81 (0.00)	0.77 (0.00)
MCMDS	Mokwa	0.80 (0.03)	
	Ibadan	0.87 (0.05)	0.68 (0.04)
FYLD	Mokwa	0.52 (0.02)	
	Ibadan	0.31 (0.02)	0.33 (0.04)
MCGM	Mokwa	0.34 (0.01)	
	Ibadan	0.24 (0.01)	0.53 (0.01)

The genetic correlations are from the ME model (standard error of estimates in brackets)

VIGOR seedling vigor, RTNO number of storage roots per plot at harvest, FYLD fresh weight of harvested roots in tons per hectare, DM percentage dry matter of storage roots, MCMDS plot mean cassava mosaic disease severity, MCGM plot mean cassava green mites severity

genotypic and additive genotype-by-environment effects. However, the lack of information from between-trait correlations (which are captured by MT models) in ME breeding values represents a challenge when selection decisions based on information from the interconnection between multiple trait and multiple location data are required. The interconnection between these data may be useful for understanding GxE and for selection on traits that are highly influenced by the environment. Therefore, there is an opportunity for interconnection between information from a valuable single environment and MET data which are readily available in plant breeding programs.

Another potential use of the ME models is for clustering environments into target populations of environments (TPE). If correlated responses to selection of target traits are similar for certain locations based on genetic correlations from the ME model, then these locations can be grouped into a TPE. Regional breeding can begin within this TPE and all multi-location trials are carried out within this TPE. For example, for the traits VIGOR, DM and MCMDS that have correlated responses to selection ranging from 66 to 87% (Table 4), Ubiaja and Ibadan can belong to same TPE.

#### Parameter estimates and implications for cassava breeding

The estimates of genomic correlations and heritabilities in Table 5 have interesting implications for cassava genetic improvement. Genetic correlations between RTNO and FYLD estimated with the MT model were high and positive for all locations (ranging from 0.65 to 0.8), whereas those between RTNO and DM and between FYLD and DM were close to zero (ranging from  $-0.02$  to 0.20). The genetic correlations for these core production traits (DM, RTNO and FYLD) indicate that concurrent improvement of these traits is achievable. However, more replication in trials targeting these production traits will help reduce error variances and improve the accuracy of parental selections given the low heritabilities for FYLD and RTNO. VIGOR can also be improved concurrently with these production traits since it is mostly positively correlated with these (Table 5). The disease trait (MCMDS) showed moderate to strong negative genomic correlations with VIGOR and the production traits, which is favorable for cassava breeding in Africa especially where the cassava mosaic disease (CMD) pressure is high. Consequently, cassava breeders have selected for CMD resistance genes over time [57, 58]. With the favorable genetic correlations between these target traits in mind, the

**Table 5 Genetic correlations and heritabilities for the traits analyzed**

Trait	VIGOR	RTNO	DM	MCMDS	FYLD	MCGM
Ubiaja						
VIGOR	0.16					
RTNO	0.63 (0.01)	0.21				
DM	0.27 (0.01)	0.19 (0.01)	0.42			
MCMDS	$-0.67$ (0.02)	$-0.53$ (0.01)	$-0.22$ (0.03)	0.62		
FYLD	0.62 (0.01)	0.80 (0.01)	0.11 (0.01)	$-0.42$ (0.02)	0.26	
MCGM	0.05 (0.01)	$-0.03$ (0.00)	$-0.17$ (0.01)	0.22 (0.01)	$-0.08$ (0.01)	0.1
Mokwa						
VIGOR	0.06					
RTNO	$-0.11$ (0.01)	0.16				
DM	0.12 (0.02)	$-0.003$ (0.01)	0.31			
MCMDS	$-0.03$ (0.02)	$-0.35$ (0.01)	$-0.14$ (0.02)	0.64		
FYLD	0.04 (0.01)	0.65 (0.01)	$-0.15$ (0.01)	$-0.18$ (0.01)	0.21	
MCGM	0.32 (0.01)	$-0.15$ (0.01)	$-0.02$ (0.01)	$-0.03$ (0.01)	$-0.1$ (0.01)	0.26
Ibadan						
VIGOR	0.19					
RTNO	0.46 (0.01)	0.26				
DM	0.18 (0.02)	0.20 (0.02)	0.37			
MCMDS	$-0.64$ (0.03)	$-0.52$ (0.03)	$-0.13$ (0.03)	0.77		
FYLD	0.34 (0.02)	0.77 (0.02)	$-0.02$ (0.02)	$-0.44$ (0.04)	0.35	
MCGM	$-0.14$ (0.01)	0.16 (0.01)	$-0.08$ (0.01)	0.11 (0.03)	0.11 (0.02)	0.22

Plot-based heritabilities on the diagonal, genetic correlations from the MT model on the off-diagonal (standard error of estimates in brackets)

VIGOR seedling vigor, RTNO number of storage roots per plot at harvest, FYLD fresh weight of harvested roots in tons per hectare, DM percentage dry matter of storage roots, MCMDS plot mean cassava mosaic disease severity, MCGM plot mean cassava green mites severity



merit index from MT breeding values should be efficient since it takes genetic correlations into account.

We would like to make it clear here that fitting MT and ME models are computationally expensive since, in our case, they required the estimation of 90 and 36 additional covariance parameters for the MT and ME models, respectively, compared to the uT and uE models. We had a few thousand records to estimate these parameters accurately for our target traits as shown by the standard errors of these estimates in Tables 4 and 5. When these correlations are not significant, breeding values from univariate models are sufficient because MT models are not expected to result in improved prediction accuracies [59].

## Conclusions

The effectiveness of a breeding program is evaluated by its ability to provide adapted and productive varieties to the farming community in the target environments it serves. To achieve this goal for the cassava breeding program at IITA, we recommend a decentralized breeding strategy for the different agro-ecological zones in Nigeria using total merit indices based on MT breeding values. Further studies should be conducted to understand how much selection gain can be achieved by using this strategy. ME models provided less improvement in prediction accuracy but were useful for understanding GxE interactions.

## Authors' contributions

UGO designed, carried out study and drafted manuscript, DA provided statistical assistance and advice, JLJ supervised the study and revised the manuscript, IR and PK supervised data generation for this study and revised the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Section of Plant Breeding and Genetics, School of Integrative Plant Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853, USA. <sup>2</sup> International Institute of Tropical Agriculture (IITA), PMB 5320, Oyo Road, Ibadan, Nigeria. <sup>3</sup> Robert W. Holley Center for Agriculture and Health, USDA-ARS, Tower Road, Ithaca, NY 14853-2901, USA.

## Acknowledgements

We acknowledge the Bill & Melinda Gates Foundation and UKaid (Grant 1048542; <http://www.gatesfoundation.org>) and support from the CGIAR Research Program on Roots, Tubers and Bananas (<http://www.rtb.cgiar.org>). We give special thanks to AGO Dixon for his development of many of the breeding lines and historical data we analyzed. Thanks also to AI Smith and technical teams at IITA for collection of other phenotypic data and to A Agbona and M Wolfe for data curation. We give special thanks to the editors and the anonymous reviewers who helped improve the quality of this paper.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

Phenotypic and genotypic data and analyses scripts used in this study can be downloaded via ftp at (<ftp://ftp.cassavabase.org/manuscripts/>).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 6 April 2017 Accepted: 16 November 2017

Published online: 04 December 2017

## References

- The International Plant Names Index (IPNI). <http://www.ipni.org/ipni/idPlantNameSearch.do?id=351790-1>. Accessed 31 Oct 2015.
- Taylor N, Chavarriaga P, Raemakers K, Sirtunga D, Zhang P. Development and application of transgenic technologies in cassava. *Plant Mol Biol*. 2004;56:671–88.
- Moorthy SN. Physicochemical and functional properties of tropical tuber starches: a review. *Starch*. 2002;54:559–92.
- Balogopalan C. Cassava utilization in food, feed and industry. In: Hillocks RJ, Thresh JM, Bellotti AC, editors. *Cassava: biology, production and utilization*. Wallingford: CAB International; 2002. p. 301–18.
- Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 1975;31:423–47.
- Henderson CR. Estimation of variances and covariances under multiple trait models. *J Dairy Sci*. 1984;67:1581–9.
- Smith HF. A discriminant function of plant selection. *Ann Hum Genet*. 1936;7:240–50.
- Hazel LN. The genetic basis for constructing selection indexes. *Genetics*. 1943;28:476–90.
- Ducrocq V, Wiggins G. Genetic improvement of cattle. In: Garrick DJ, Ruvinsky A, editors. *The genetics of cattle*. 2nd ed. Wallingford: CAB; 2015. p. 371–96.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome wide dense marker maps. *Genetics*. 2001;157:1819–29.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92:433–43.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6:e19379.
- Segelke D, Reinhardt F, Liu Z, Thaller G. Prediction of expected genetic variation within groups of offspring for innovative mating schemes. *Genet Sel Evol*. 2014;46:42.
- Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 2008;177:2389–97.
- VanRaden PM, Wiggins GR. Derivation, calculation, and use of national animal model information. *J Dairy Sci*. 1991;74:2737–46.
- Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136:245–57.
- van der Werf J, van Arendonk JAM, De Vries AG. Improving selection of pigs using correlated characters. In: *Proceedings of the 43rd annual meeting of the European federation of animal science (EAAP)*, 14–17 September 1992, Madrid. 1992.
- Ducrocq V. Multiple trait prediction: principles and problems. In: *Proceedings of the 5th world congress on genetics applied to livestock production*, 7–12 August 1994, Guelph. 1994.
- Colleau JJ, Ducrocq V, Boichard D, Larroque H. Approximate multi-trait BLUP evaluation to combine functional traits information. *Interbull Bull*. 1999;23:1–9.
- Schaeffer LR. Multiple-country comparison of dairy sires. *J Dairy Sci*. 1994;77:2671–8.
- Schaeffer LR. Multiple trait international bull comparisons. *Livest Prod Sci*. 2001;69:145–53.
- Thompson R, Meyer K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest Prod Sci*. 1986;15:299–313.
- Smith A, Cullis B, Thompson R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*. 2001;57:1138–47.
- Falconer DS. The problem of environment and selection. *Am Nat*. 1952;86:293–8.
- Malosetti M, Ribaut JM, van Eeuwijk FA. The statistical analysis of multi-environment data: modelling genotype-by-environment interaction and its genetic basis. In: Monneveux P, Ribaut JM, Okono A, editors. *Drought phenotyping in crops: from theory to practice*, vol 4. Lausanne, Switzerland: Frontiers Media, SA; 2014. p. 53–69.

26. Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A. Modeling genotype  $\times$  environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop Sci*. 2006;46:1722–33.
27. Burgueño J, Crossa J, Cornelius PL, Yang RC. Using factor analytic models for joining environments and genotypes without crossover genotype  $\times$  environment interaction. *Crop Sci*. 2008;48:1291–305.
28. Burgueño J, Crossa J, Miguel Cotes J, San Vicente F, Das B. Prediction assessment of linear mixed models for multi-environment trials. *Crop Sci*. 2011;51:944–54.
29. Burgueño J, de los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotype  $\times$  environment interaction using pedigree and dense molecular markers. *Crop Sci*. 2012;52:707–19.
30. de los Campos G, Gianola D. Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. *Genet Sel Evol*. 2007;39:481–94.
31. Meyer K. Factor-analytic models for genotype  $\times$  environment type problems and structured covariance matrices. *Genet Sel Evol*. 2009;41:21.
32. Tsuruta S, Lourenco DAL, Misztal I, Lawlor TJ. Genotype by environment interactions on culling rates and 305-day milk yield of Holstein cows in 3 US regions. *J Dairy Sci*. 2015;98:5796–805.
33. Kolmodin R, Strandberg F, Madsen P, Jensen J, Jorjani H. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agric Scand Anim Sci*. 2002;52:11–24.
34. Windig JJ, Calus MPL, Veerkamp RF. Influence of herd environment on health and fertility and their relationship with milk production. *J Dairy Sci*. 2005;88:335–47.
35. Cooper M, DeLacy IH. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theor Appl Genet*. 1994;88:561–72.
36. Ducrocq V, Boichard D, Barbat A, Larroque H. Implementation of an approximate multi-trait BLUP evaluation to combine production traits and functional traits into a total merit index. In: Proceedings of the 52nd annual meeting of the European association for animal production, 26–29 August 2001, Budapest; 2001.
37. Lassen J, Sørensen MK, Madsen P, Ducrocq V. A stochastic simulation study on validation of an approximate multitrait model using preadjusted data for prediction of breeding values. *J Dairy Sci*. 2007;90:3002–11.
38. Lassen J, Sørensen MK, Madsen P, Ducrocq V. An approximate multi-trait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends. *Genet Sel Evol*. 2007;39:353–67.
39. Maziya-Dixon B, Dixon AGO, Adebawale ARA. Targeting different end uses of cassava: genotypic variations for cyanogenic potentials and pasting properties. *Int J Food Sci Technol*. 2007;42:969–76.
40. Okechukwu RU, Dixon AGO. Genetic gains from 30 years of cassava breeding in Nigeria for storage root yield and disease resistance in elite cassava genotypes. *J Crop Improv*. 2008;22:181–208.
41. Ly D, Hamblin M, Rabbi I, Melaku G, Bakare M, Gauch HG Jr, et al. Relatedness and genotype  $\times$  environment interaction affect prediction accuracies in genomic selection: a study in cassava. *Crop Sci*. 2013;53:1312–25.
42. Glaubitz J, Casstevens T, Elshire R, Harriman J, Buckler ES. TASSEL 3.0 genotyping by sequencing (GBS) pipeline documentation. Edward S. Buckler. Ithaca: USDA-ARS. 2012. <http://www.maizegenetics.net/tassel/docs/TasselPipelineGBS.pdf>. Accessed 3 Jan 2014.
43. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:D1178–86.
44. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
45. Bates D, Maechler M, Bolker B, Walker S. lme4: linear mixed-effects models using Eigen and S4. R package version 1.0-5. <http://CRAN.R-project.org/package=lme4>. 2013.
46. Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome*. 2012;5:51–61.
47. Wolfe MD, Kulakow P, Rabbi IY, Jannink JL. Marker-based estimates reveal significant non-additive effects in clonally propagated cassava (*Manihot esculenta*): implications for the prediction of total genetic value and the selection of varieties. *G3 (Bethesda)*. 2016;3:3497–506.
48. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
49. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee D. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th world congress on genetics applied to livestock production, 19–23 August 2002, Montpellier. CD-ROM Communication 28:07; 2002.
50. Piepho HP, Pillen K. Mixed modelling for QTL  $\times$  environment interaction analysis. *Euphytica*. 2004;137:147–53.
51. VanRaden PM, Tooker ME, Wright JR, Sun C, Hutchison JL. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J Dairy Sci*. 2014;97:7952–62.
52. Calus MP, Veerkamp RF. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol*. 2011;43:26.
53. Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet*. 2014;15:30.
54. Oakey H, Cullis B, Thompson R, Comadran J, Halpin C, Waugh R. Genomic selection in multi-environment crop trials. *G3 (Bethesda)*. 2016;6:1313–26.
55. Cuevas J, Crossa J, Soberanis V, Pérez-Elizalde S, Pérez-Rodríguez P, de los Campos G, et al. Genomic prediction of genotype  $\times$  environment interaction kernel regression models. *Plant Genome*. 2016;9:3.
56. Cuevas J, Crossa J, Montesinos-López OA, Burgueño J, Pérez-Rodríguez P, de los Campos G. Bayesian genomic prediction with genotype  $\times$  environment interaction kernel models. *G3 (Bethesda)*. 2017;7:41–53.
57. Jennings DL. Breeding for resistance to African cassava mosaic disease: progress and prospects. In: Nestel, editor. African cassava mosaic. Report of an interdisciplinary workshop, 19–22 February 1976, Muguga; 1976.
58. Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P, et al. Genome-wide association and prediction reveals genetic architecture of Cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome*. 2016;9:2.
59. Schaeffer LR. Sire and cow evaluation under multiple trait models. *J Dairy Sci*. 1984;67:1567–80.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

