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

Marnin D. Wolfe[*], Peter Kulakow[†], Ismail Y. Rabbi[†], Jean-Luc Jannink[*,‡]

[*] Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA
[†] International Institute for Tropical Agriculture (IITA), Ibadan, Oyo, Nigeria
[‡] USDA-ARS, R.W. Holley Center for Agriculture and Health, Ithaca, NY, USA

Running Title: Marker-based estimation of non-additive effects in cassava

Key words: genomic selection, non-additive effects, cassava

Corresponding Author:
Marnin D. Wolfe
Cornell University
Department of Plant Breeding and Genetics
417 Bradfield Hall
306 Tower Road
Ithaca, NY 14853
Email: wolfemd@gmail.com
Phone: 239-595-5081

ABSTRACT

In clonally propagated crops, non-additive genetic effects can be effectively exploited by the identification of superior genetic individuals as varieties. Cassava (*Manihot esculenta* Crantz) is a clonally propagated staple food crop that feeds hundreds of millions. We quantified the amount and nature of non-additive genetic variation for key traits in a breeding population of cassava from sub-Saharan Africa using additive and non-additive genome-wide marker-based relationship matrices. We then assessed the accuracy of genomic prediction of additive compared to total (additive plus non-additive) genetic value. We confirmed previous findings based on diallel populations, that non-additive genetic variation is significant, especially for yield traits. Further, we show that we total genetic value correlated more strongly to observed phenotypes than did additive value, although this is constrained by low broad-sense heritability and is not beneficial for traits with already high heritability. We address the implication of these results for cassava breeding and put our work in the context of previous results in cassava, and other plant and animal species.

INTRODUCTION

Understanding genetic architecture requires the decomposition of genetic effects into additive, dominance, and epistatic components (Fisher 1918; Cockerham 1954; Kempthorne 1954). However, partitioning genetic variance components is notoriously difficult, requiring specialized breeding designs (e.g. diallel crosses) and pedigree information (Lynch and Walsh 1998) often limiting the genetic diversity that can be sampled in any one given study. Genome-wide molecular marker data now enable the accurate measurement of relatedness in the form of genomic realized relationship matrices (GRMs) (VanRaden 2008; Heffner *et al.* 2009; Lorenz *et al.* 2011b). GRMs provide more accurate relatedness information than pedigrees because they directly measure Mendelian sampling (causing variation in relatedness within relatedness classes such as full-siblings) (Heffner *et al.* 2009). Further, GRMs can measure relationships even in diverse, nominally unrelated samples expanding the potential for studying inheritance in natural and breeding populations (Lorenz *et al.* 2011a).

Estimation of narrow-sense heritability and prediction of breeding values in genomic selection programs is becoming increasingly common using additive formulations of GRMs (Visscher *et al.* 2008). Several recent studies have described dominance and epistatic GRMs for the partitioning of non-additive genetic variance using genome-wide SNP markers (Su *et al.* 2012; Vitezica *et al.* 2013; Muñoz *et al.* 2014; Wang *et al.* 2014). Models using these new formulations have been shown to provide improved partitioning of genetic variances relative to pedigree-based approaches (Su *et al.* 2012; Muñoz *et al.* 2014). These new models can be used not only to estimate genetic

variances but also for genomic prediction of additive and total genetic value in genomic selection breeding programs (Su *et al.* 2012; Vitezica *et al.* 2013; Muñoz *et al.* 2014; Wang *et al.* 2014).

Cassava is a vegetatively propagated, staple food crop that is high in starch and feeds half a billion people worldwide (http://faostat.fao.org). Efforts to improve cassava genetically with cutting edge methodologies including transgenic and genomic selection (GS) approaches are underway thanks to new genomic resources (Prochnik *et al.* 2012; (ICGMC) 2014). Prediction of additive genetic value has recently been evaluated (Oliveira *et al.* 2012; Ly *et al.* 2013) and genomic selection using standard models is currently being tested **(**http://www.nextgencassava.org**)**. Vegetatively propagated crop (e.g. cassava) breeding can exploit non-additive genetic effects by identifying superior clones as varieties (Ceballos *et al.* 2015).

Diallel studies in cassava indicate that non-additive genetic effects (i.e. specific combining ability) are strong, particularly for root yield traits (Cach *et al.* 2005, 2006; Calle *et al.* 2005; Jaramillo *et al.* 2005; Perez *et al.* 2005; Pérez *et al.* 2005; Zacarias and Labuschagne 2010; Kulembeka *et al.* 2012; Tumuhimbise *et al.* 2014; Ceballos *et al.* 2015; Chalwe *et al.* 2015). If the limited number of parents tested thus far represents the broader cassava breeding germplasm, genetic gains, especially for already low-heritability root yield traits will be slow regardless of the breeding scheme employed (e.g. phenotypic vs. pedigree vs. genomic selection). Breeding gains have indeed been slow in cassava (Ceballos *et al.* 2012) and low accuracies have been reported for genomic prediction of yield compared with cassava mosaic disease (CMD) resistance and dry matter (DM) content (Oliveira *et al.* 2012; Ly *et al.* 2013). However, cassava varieties are

evaluated and disseminated to farmers by clonal propagation, meaning that accurate prediction of total (additive plus non-additive) genetic value could contribute to variety selection.

In this study, we test the hypothesis that certain cassava traits, especially root yield have relatively large non-additive genetic variance that account for low genomic prediction accuracies previously observed. We estimate additive and non-additive variance components using genomic relationship matrices in two populations of cassava from the International Institute of Tropical Agriculture's (IITA) genomic selection breeding program. Further, we assess the accuracy of predicting additive and total genetic value using the additive and non-additive models. We discuss the origin of non-additive genetic variance in cassava, its potential effect on cassava breeding, and its role in genomic selection strategies for cassava improvement in the future.

METHODS

Germplasm and Phenotyping Trials

We examined additive and non-additive effects in two populations of cassava that have been genotyped and phenotyped as part of the Next Generation Cassava Breeding Program at IITA, Nigeria (http://www.nextgencassava.org). The IITA's Genetic Gain (GG) population contains 694 historically important clones, most of which are advanced breeding lines although some are classified as superior landraces. These lines have been selected and maintained clonally since 1970 (Okechukwu and Dixon 2008; Ly *et al.* 2013). Most of these materials are derived from the cassava gene pool from West Africa as well as parents derived from the breeding program at Amani Station in Tanzania and

hybrids of germplasm introduced from Latin America. Available information on the GG accessions included in our analyses is provided in Table S1.

IITA's Genetic Gain trials were conducted in seven locations over 14 years (2000 to 2014) in Nigeria for a total of 24,373 observations. Each GG trial comprises a randomized, incomplete block design replicated one or two times per location and year. Since materials have been occasionally lost and new, selected materials are continuously added to the GG, the number of clones trialed in a given year changes gradually across years, generally increasing. The sample sizes, number of replicates and number of clones from the GG in each of the trials (location-year combinations) is provided in Table S2.

Theory suggests that founding events and truncation selection can both lead to a conversion of non-additive genetic variation into additive variance. This can happen because of the induction of linkage disequilibrium and reduction in allele frequency (or fixation of alleles) at some loci relative to others (Goodnight 1988; Turelli and Barton 2006; Hallander and Waldmann 2007). Consequently, our results might depend on the population examined. We therefore analyzed an additional population: a collection of 2187 clones that are the direct descendants of truncation selection on the GG. Briefly, in 2012 the GG and all available historical phenotype data was used as a reference population to obtain genome estimated breeding values (GEBVs) using the genomic BLUP (GBLUP) model (VanRaden 2008; Heffner *et al.* 2009). Selection was based on an index that included mean cassava mosaic disease severity (MCMDS), mean cassava bacterial blight disease severity (MCBBS), dry matter content (DM), harvest index (HI) and fresh root weight (RTWT). This index of GEBVs was used to select 83 members of

the GG to cross and generated a population of 137 full-sib families, which we refer to as the GS Cycle 1 (C1). The pedigree of the C1 is available in Table S3.

Cycle 1 progenies were evaluated in a single clonal evaluation trial during the 2013-2014 field season across three-locations (Ibadan, Ikenne, and Mokwa). For the C1 clonal trial, planting material was only available for one plot of five stands per clone, so each clone was only planted in one of the three locations (Table S2). Clones were assigned to each location so as to equally represent each family in every environment.

For both populations, we analyzed the five traits described above (MCMDS, MCBBS, DM, HI, RTWT) plus sprouting ability (SPROUT). MCMDS and MCBBS were scored on a scale of 1 (no symptoms) to 5 (severe symptoms). Most GG trials measured dry matter (DM) by the oven drying method although some trials used the specific gravity method. DM is expressed as a percentage of the fresh weight of roots. Fresh root weight (RTWT) is measured in kilograms per plot and harvest index (HI) is the percent of total biomass per plot (roots plus shoots) that is RTWT. Sprouting ability (SPROUT) was expressed as the percent of planted stakes sprouting at one month after planting.

Genotype data

We used genotyping-by-sequencing (GBS) to obtain genome-wide SNP marker data (Elshire *et al.* 2011). We used the ApeKI restriction enzyme as recommended by (Hamblin and Rabbi 2014). SNPs were called using the TASSEL GBS pipeline V4 (Glaubitz *et al.* 2014) and aligned to the cassava reference genome, version 5, which is available on Phytozome (http://phytozome.jgi.doe.gov) and described by the International Cassava Genetic Map Consortium ((ICGMC) 2014). We removed individuals with more

than 80% and markers with >60% missing genotype calls. Also, markers with extreme deviation from Hardy-Weinberg equilibrium (Chi-square > 20) were removed. If there were not at least two reads at a given locus for a given clone, the genotype was set to missing and imputed. SNP marker data was converted to the dosage format (0, 1, 2) and missing data were imputed with the glmnet algorithm in R (http://cran.r-project.org/web/packages/glmnet/index.html) as described in (Wong *et al.* 2014). We used 114,922 markers that passed these filters with a minor allele frequency greater than 1% to calculate genomic relationship matrices as described below.

Genomic Relationship Matrices

We measured the realized additive, dominance and epistatic relationships in our population using functions that have been described previously (Su *et al.* 2012; Muñoz *et al.* 2014). We used the additive realized relationship matrix, **A** as described by Van Raden (VanRaden 2008): $\mathbf{A} = \frac{WW'}{2\sum_i p_i q_i}$. Here W is a matrix of dimension *n* individuals by *m* SNP markers. The elements of W are the marker dosages $(0 - 2p_i)$ for aa, $(1 - 2p_i)$ for Aa and $(2 - 2p_i)$ for AA, where p_i is the frequency of the second allele and q_i is the frequency of the first allele, at the *i*th locus. The a (or 0) allele refers to the reference genome allele. The **A** matrix was calculated using the *A.mat* function in the *rrBLUP* package (Endelman 2011).

As described by Su et al. (Su *et al.* 2012) the dominance relationship matrix, **D** is $\mathbf{D} = \frac{HH'}{\sum_i 2p_i q_i (1-2p_i q_i)}$ (also see (Muñoz *et al.* 2014)). Where H has the same dimensions as W, heterozygotes are scored $(1 - 2p_iq_i)$ and homozygotes are $(0 - 2p_iq_i)$. We made a custom modification (available upon request) to the *A.mat* function to produce the **D**

181 matrix. Relationship matrices that capture epistasis can be calculated by taking the
182 hadamard product (element-by-element multiplication; denoted #) of two or more
183 matrices (Henderson 1985). For simplicity, we only explored additive-by-additive (**A#A)**
184 and additive-by-dominance **(A#D)** relationships in this study.
185
186 **Variance component and heritability models**
187 *Single-step, Multi-environment*
188 We used several approaches to estimate the relative importance of additive and non-
189 additive effects in the Genetic Gain and Cycle 1 populations. In the first analysis, we
190 analyzed the multi-year, multi-location GG data with a single random effects model.
191 Since the entire historical phenotype dataset is large (24,373 observations) and was
192 relatively unbalanced in sample size across years and locations, we only analyzed data
193 from trials with >400 individuals. This filter resulted in a dataset of 7745 observations
194 from three locations (Ibadan, Ubiaja, Mokwa) and eight years (2006-2014, excluding
195 2012). All 694 genotyped GG clones were represented in this dataset (Table S2).
196 The models we fit were similar to those described in Ly et al. (Ly *et al.* 2013). The
197 full model was specified as follows: $y = X\beta + Z_{loc.year}l + Z_{rep}r + Z_{add}a + Z_{dom}d +$
198 $Z_{epi}i + \varepsilon$. Here, y represents raw phenotypic observations. In our data, the only fixed
199 effect (β) was an intercept for all traits except RTWT, which contained a covariate
200 accounting for variation in the number of plants harvested per plot. The random effects
201 terms for experimental design terms included a unique intercept for each trial (i.e.
202 location-year combination), $l \sim N(0, \mathrm{I}\sigma_l^2)$, where I is the identity matrix and σ_l^2 is the

associated variance component as well as a replication effect, nested in location-year combination, $r \sim N(0, \mathrm{I}\sigma_r^2)$.

The genetic variance component terms included $a \sim N(0, \mathrm{A}\sigma_a^2)$, where A is the additive relationship matrix and σ_a^2 is the additive genetic variance component. Similarly, $d \sim N(0, \mathrm{D}\sigma_d^2)$, is the dominance effect with covariance D equal to the dominance relationship matrix and σ_d^2 equal to the dominance variance. The epistatic term $i \sim N(0, \mathrm{E}\sigma_i^2)$ where the covariance matrix E took the form either of the A#A matrix (additive-by-additive) or the A#D matrix (additive-by-dominance) and the epistatic variance σ_i^2 was correspondingly either $\sigma_{A\#A}^2$ or $\sigma_{A\#D}^2$. The final term, ε is the residual variance, assumed to be random and distributed $N(0, {\sigma_\varepsilon}^2)$. The terms X, $Z_{loc.year}$, Z_{rep}, Z_{add}, Z_{dom} and Z_{epi} are incidence matrices relating observations to the levels of each factor. We list the different models fit in Table 1, each of which are variations on the full model described above.

The formulation described above was used to fit the GG historical data in a single model. For the C1 progenies only a single trial was available and therefore we fit all data together. Since the C1 trial was conducted across three locations but with no replications we fit the same model for C1 as GG excluding the replication term. The models described above were fit using the *regress* package in R (R Core Team 2015). The *regress* function finds REML solutions to mixed models using the Newton-Raphson algorithm.

For each trait, in both the C1 and GG we identified a "best fit" model among the models listed in Table 1 based on the lowest Akaike Information Criterion (AIC; 2*k – 2*log(likelihood), where k = number parameters estimated). We also examined the log-likelihood of each model and the total genetic variance explained (H^2). The precision of

variance component estimates and the dependency among estimates was examined using the asymptotic variance-covariance matrix of estimated parameters **(V)**. Specifically, we report standard errors for each variance component,defined as the square root of the diagonal of **V**. We also converted **V** into a correlation matrix (**F,** as in (Muñoz *et al.* 2014)), where **F** is defined as $\mathbf{L}^{-1/2}\mathbf{V}\mathbf{L}^{-1/2}$ and **L** is a diagonal matrix containing one over the square root of the diagonal of **V**. We use **F** to assess the dependency of variance components estimates, especially for comparing results among traits and populations.

Within-trial analyses

We used only a subset of the GG trials to estimate variance components in a single multi-environment model. In addition, we leveraged the entire historical GG data by analyzing each trial (N=47, unique location-year combinations) separately. This provided us with 47 estimates of additive, dominance and epistatic variance and we examine the distribution of variance components estimates. As in the multi-environment models, within-trial models were fit with *regress* in R.

Genomic prediction and cross-validation

We assessed the influence that modeling non-additive genetic variance components have on genomic prediction using a cross-validation strategy. Because single-step multi-environment models are computationally intensive, we used a two-step approach here. In the first step, we combined data from all available GG and C1 trials using the following mixed model: $y = X\beta + Z_{rep}r + Z_{clone}g + \varepsilon$. In this model, β included a fixed effect for the population mean, the location-year combination and for

RTWT only, the number of plants harvested per plot. As in the single-step, multi-environment model for GG, we included the random replication effect $r \sim N(0, \mathrm{I}\sigma_r^2)$. In contrast to the previous model, we did not at this stage include a genomic relationship matrix, instead we fit a random effect for clone, $g \sim N\left(0, \mathrm{I}\sigma_g^2\right)$, where the covariance structure was the identity matrix (I). The BLUP ($\hat{g}$) for the clone effect therefore represents an estimate of the total genetic value for each individual. The mixed model above was solved using the *lmer* function of the *lme4* R package.

In our data, the number of observations per clone ranges from one to 131 with median of two and mean of 5.97 excluding the checks TMEB1 and I30572 which had 941 and 902 observations respectively. Pooling information from multiple years and locations, especially when there is so much variation in numbers of observations can introduce bias. Much theoretical development, particularly in animal breeding has been done to address this issue, and we followed the approach recommended by Garrick et al. (Garrick *et al.* 2009). Briefly, BLUPs for clone were deregressed according to $\frac{\hat{g}}{r^2}$ where r^2 is the reliability $(1 - \frac{PEV}{\sigma_g^2})$ and PEV is the prediction error variances of the BLUP. In the second step of analysis, where deregressed BLUPs are used as response variables, weights are applied to the diagonal of the error variance-covariance matrix **R**. Weights are calculated as $\frac{1-h^2}{0.1+\frac{1-r^2}{r^2}h^2}$, where h^2 is the proportion of the total variance explained by the clonal variance component, σ_g^2 (Garrick *et al.* 2009).

We implemented a 5-fold cross-validation scheme replicated 25 times to test the accuracy of genomic prediction using the genomic relationship matrices and models described above (Table 1). In this scheme, for each replication, we randomly divided the

dataset into five equally sized parts (i.e. folds). We used each fold in turn for validation by removing its phenotypes from the training population and then predicting them. We calculated accuracy as the Pearson correlation between the genomic prediction and the BLUP (ĝ, not-deregressed) from the first step. For each model, we calculated accuracy both of the prediction from the additive kernel (where present) and the total genetic value prediction, defined as the sum of the predictions from all available kernels (e.g. additive + dominance + epistasis). Genomic predictions were made using the *EMMREML* R package.

All raw genotype and phenotype data are available on www.cassavabase.org [an exact link to be provided upon acceptance of manuscript]. Custom code used for analysis is available upon request.

RESULTS

Single-step, Multi-environment Variance Component Models

Our first assessments of non-additive genetic effects in cassava were single-step multi-environment models implemented in both the Genetic Gain (GG) and the Cycle 1 (C1) populations. The five models (Table 1) were initially compared using the AIC. Tables 2 and 3 show the model results including AIC and variance components for the GG and C1, respectively. For HI, MCMDS, and SPROUT the best models were A + D, A + D + AD and A + D + AA, respectively. For these three traits, the best model according to AIC was the same between the GG and C1. For DM the additive only model was best in the GG but an A + D model was selected in C1. For RTWT, in the GG an A + D + AA model was selected but in the C1 the dominance only model was selected. Finally, for

MCBBS the additive only model was best in the GG but A + D + AD was selected in the C1.

For every trait, when comparing the model achieving the highest broad-sense heritability (H^2), we saw that the H^2 increased from GG to C1. This can be seen most easily in Figure 1, which shows how total explainable genetic variance (H^2) is partitioned among variance components in the C1 and GG (also see Tables 2 & 3). In GG, the additive only model had the highest H^2 for all traits, but in C1 additive + non-additive models always had at least slightly higher H^2.

For DM, the additive component explained more variance across all models in the C1 compared to the GG, e.g. 0.33 (GG, Additive) and 0.51 (C1, Additive). For DM, the A + D + AA model actually had the highest H^2 in the C1 but the variance component for A#A epistasis was not distinguishable from zero (2.05±2.59).

For HI, the additive only model was very similar between populations (h^2 = 0.34 in GG and 0.32 in C1). But for the best model, A + D, more of the total genetic variance is partitioned to C1 (0.28) than in the GG (0.07). Like for DM, the three component models (A + D + AA and A + D + AD) explained the most variance for the C1 (H^2 = 0.49), but the epistatic components had large standard errors (26.4±18.7 and 24.4±17.6 respectively).

RTWT was strongly non-additive in both populations. In the GG the A + D + AA model, genetics explained 25% of the phenotypic variance with non-additive variances (D + AA) explained over half of that amount (16% collectively) and the epistatic term was significantly different from zero (0.033 ± 0.014). In the C1, 31% of the phenotypic variance was explainable by dominance alone.

MCBBS had a low heritability overall (H^2 = 0.04 in the GG for the additive only model, and 0.25 in the C1 for the A + D + AD model). While the limited apparent genetic variance in the GG was best explained additively, in the C1 92% of total genetic variance was non-additive.

For MCMDS, the additive model had the highest H^2 (0.79) in the GG but the AIC best model, A + D + AD explained most (0.89) in the C1. The AIC best model for both C1 and GG was the same. In both cases, significant AxD epistasis was detected, explaining 44% of the variance in the GG and 18% in the C1. The additive variance increased from GG (0.25) to C1 (0.64).

SPROUT was best modeled with A + D + AA in both populations. Broad-sense heritability increased from 0.22 in GG to 0.39 in C1 for this trait. The dominance component was not significant (5.8 ± 8.6) in the GG but was in the C1 (31.7 ± 26.7).

We examined the asymptotic correlation matrices of parameter estimates (**F)** to ascertain the dependency of variance component estimation. Correlation matrices for every trait + model combination are provided for GG in Tables S3-S6 and for the C1 in Tables S7-S9. The correlation between genetic variance components was always negative and was, in general, higher in the GG compared to the C1. Correlations between additive and dominance components were highest in the A+D models (range -0.81 to -0.83 in the GG and -0.5 to -0.61 in the C1). Correlations between A and D components dropped in models with epistasis (range -0.42 to -0.63, GG and -0.26 to -0.58, C1). Correlations between additive and AxA epistatic variances (range -0.09 to -0.29) and AxD epistasis (range -0.07 to -0.22) were low. Correlations between dominance components and

339 epistasis were higher ranging from -0.28 to -0.64 with AxA epistasis and -0.36 to -0.69
340 with AxD epistasis.
341
342 *Within-trial analyses*
343 We also examined variance partitioning within each of 47 GG trials for the 5
344 models described in Table 1. The mean and variability of model parameters (variance
345 components, heritability, etc.) across these trials are summarized in Table S10. Figure 2
346 provides a visual summary of the proportion of phenotypic variability explained by each
347 genetic variance component on average across the trials. We also compared the mean
348 AIC across trials and found them to agree overall with the results of the one-step multi-
349 environment models (Table 2, Table S10). Specifically, the models that fit best in the
350 one-step models were best on average in the within trial analyses for the following traits:
351 DM (additive), HI (A+D), and MCMDS (A+D+AD). However, for RTWT (A+D instead
352 of A+D+AD) and MCBBS (D instead of A) the within trial AIC-best models were
353 different on average from the one-step multi-environment models.
354
355 *Genomic Prediction of Additive and Total Genetic Value*
356 We used cross-validation to assess the accuracy of genomic prediction of additive and
357 total genetic value for the five models (Table 1) in both populations. For additive
358 prediction, the additive only model had higher accuracy for every trait and every
359 population (Figure 3, Table S11-S12). The prediction accuracy for the additive kernel
360 was higher on average in the C1 for DM, MCBBS, MCMDS, and SPROUT compared to
361 the GG but lower for HI and RTWT. Additive accuracies by trait (across models) were

362 highest for DM (range: 0.54-0.60), followed by MCMDS (range: 0.44-0.54) and HI
363 (range: 0.32-0.45). The rest of the traits had similar additive accuracies dependent on the
364 population and model. RTWT had the overall lowest accuracy (range: 0.09-0.36).
365 Across all multi-kernel analyses conducted, prediction of total genetic value was on
366 average 42% more accurate compared to the prediction of the additive kernel in the same
367 model. Compared to the single-kernel additive prediction however, total genetic value
368 predictions were an average of only 6% better (maximum of 26% improvement; Figure 4,
369 Tables S11-S12). By model, improvements in the correlation between total value and
370 phenotype over the additive only model were 5%, 7% and 8% for A+D, A+D+AA and
371 A+D+AD respectively. The additive only model predictions were on average 11% less
372 accuracy in the C1 than in the GG. Additive kernel predictions from models with non-
373 additive components were 7% lower in C1 compared to GG. Total genetic value
374 predictions were also less accurate by 12% in the C1 relative to GG.
375 The models we fit for multi-kernel genomic prediction involved the estimation of
376 weight parameters corresponding to the partitioning of genetic variance among the
377 kernels. Overall, there was a tendency for non-additive kernels (mean of 0.40 for
378 dominance, 0.40 AxA epistasis and 0.46 AxD epistasis) to get more weight than additive
379 (mean of 0.32). Additive prediction accuracies were positively correlated (r = 0.81) with
380 weight placed on the additive kernel. The accuracy of total genetic value prediction was
381 negatively correlated (r = -0.64) with weight placed on non-additive kernels (Tables S11-
382 S12).
383
384

DISCUSSION

In clonally propagated crops, non-additive genetic effects can be effectively exploited by the identification of superior genetic individuals as varieties. For this reason, we quantified the amount and nature of non-additive genetic variation for key traits in a genomic selection breeding population of cassava from sub-Saharan African. Then we assessed the accuracy of genomic prediction of additive compared to total (additive plus non-additive) genetic value. Using several approaches and datasets based on genome-wide marker data, we confirmed previous findings in cassava based on diallel populations, that non-additive genetic variation is significant, especially for yield traits (Cach *et al.* 2005, 2006; Calle *et al.* 2005; Jaramillo *et al.* 2005; Perez *et al.* 2005; Pérez *et al.* 2005; Zacarias and Labuschagne 2010; Kulembeka *et al.* 2012; Tumuhimbise *et al.* 2014; Ceballos *et al.* 2015; Chalwe *et al.* 2015). Further, we found that total genetic value predicted observed phenotype more accurately than additive value, although this is constrained by low broad-sense heritability and is not beneficial for traits with already high heritability (DM, MCMDS). We address the implication of these results for cassava breeding and put our work in the context of previous results in cassava, other plant and animal species below.

Our results indicate strong non-additive variance for root yields and mostly additive inheritance of root dry matter content. These findings confirm the conclusions of numerous diallel studies conducted with both Latin American (Cach *et al.* 2005, 2006; Calle *et al.* 2005; Jaramillo *et al.* 2005; Perez *et al.* 2005; Pérez *et al.* 2005) and African cassava (Zacarias and Labuschagne 2010; Kulembeka *et al.* 2012; Tumuhimbise *et al.* 2014; Chalwe *et al.* 2015) germplasm (see also (Ceballos *et al.* 2015)). In agreement with

the findings of Ly et al. (Ly *et al.* 2013), we found cassava mosaic disease severity (MCMDS) to be well predicted with an additive only model. However, we found significant dominance and epistatic components in both populations analyzed. This result is in line with previous diallel studies indicating significant SCA (Tumuhimbise *et al.* 2014; Chalwe *et al.* 2015) and genetic mapping studies that identified a single major effect QTL with a dominant CMD resistance effect (Akano *et al.* 2002; Okogbenin *et al.* 2012; Rabbi *et al.* 2014). In addition, a recent genome-wide association and prediction study using non-additive genomic relationship matrices (GRMs) found that dominance and especially epistasis explain most of the variance in the region of a large-effect QTL, suggesting multiple interacting loci in the region **(Wolfe et al. *in review*).** Nearly every study (but see (Zacarias and Labuschagne 2010)) indicates that Harvest Index is mostly explained by additive variance, in agreement with our findings for the GG but in contrast to our results for the C1. Our study is the first to report on the modes of inheritance of CBB and SPROUT.

The importance of non-additive genetic variance in evolution by natural and artificial selection is controversial (Hill *et al.* 2008; Crow 2010; Hansen 2013). Nevertheless, numerous studies have found and exploited dominance and epistasis in animal breeding, including dairy (Ahlborn-Breier and Hohenboken 1991; Fuerst and Sölkner 1994; Varona *et al.* 1998; Van Tassell *et al.* 2000; Palucci *et al.* 2007) and beef (Rodriguezalmeida *et al.* 1995) cattle. Diallel studies have indicated significant SCA for maize grain yield (Doerksen *et al.* 2003; Wardyn *et al.* 2007). Aside from cassava, breeding of other non-inbred, clonally propagated species also identify and make use of non-additive effects, including potato (Killick 1977), Eucalyptus (Costa E Silva *et al.*

2004) and loblolly pine (Muñoz *et al.* 2014). More recently, marker-based and GRM-based models have identified significant non-additive effects in pigs (Su *et al.* 2012; Nishio and Satoh 2014), mice (Vitezica *et al.* 2013), beef cattle (Bolormaa *et al.* 2015), dairy cows (Morota *et al.* 2014), maize (Dudley and Johnson 2009), soy (Hu *et al.* 2011), loblolly pine (Muñoz *et al.* 2014) and apple (Satish Kumar, Claire Molloy, Patricio Muñoz, Hans Daetwyler, David Chagné 2015). Results from the present study suggest that the combination of genomic selection and hybrid breeding strategies should increase the rate of genetic gain for complex traits such as yield. However, initial investment in identification of complementary heterotic groups with good specific combining ability is required.

One of the more interesting aspects of our study relative to previous ones is the comparison between a parental generation (the Genetic Gain) and their offspring (Cycle 1), a collection of full- and half-sib families. From GG to C1, the H^2 generally increased. For RTWT, MCBBS and HI this is largely attributable to increased non-additive variance and CMD for which non-additive variance dropped in C1 relative to GG. In contrast to our result, theory suggests that reduction (or fixation) of allele frequencies at some loci relative to others in populations undergoing bottlenecks (Goodnight 1988), inbreeding (Turelli and Barton 2006) or truncation selection (Hallander and Waldmann 2007) should cause a conversion of non-additive (where present) to additive variance. These results have, however, been based on models with finite numbers of loci in linkage equilibrium. Based on the mean diagonal of the kinship matrix, C1 (0.66) does not appear notably more inbred than GG (0.64). We also calculated mean pairwise LD (GG = 0.27, C1 =

453 0.29) and mean LD block size (21.7 kb in GG and 23.1 kb in C1) using PLINK (version
454 1.9, https://www.cog-genomics.org/plink2) and found the two generations to be similar.
455 Probably the strongest explanation for the difference in genetic variance
456 components between GG and C1 is the family structure (137 full-sib families from 83
457 outbred parents). In a population of half-sibs ¾ of the dominance variance is expressed
458 within families and all of it for full-sib populations (Hallauer *et al.* 2010; Ceballos *et al.*
459 2015). Indeed, increasing the number of full-sib relationships is known to increase the
460 non-additive genetic variance detectable in a population (Varona *et al.* 1998; Van Tassel
461 *et al.* 2010).
462 It is also conceivable that maternal plant effects could increase apparent non-
463 additive effects in C1. The C1 clones in contrast to the GG clones are new, and were
464 derived from stem cuttings of seedling plants germinated in the previous field season
465 (2012-2013). The suggestion is therefore that the quality and vigor of the seedling plant,
466 giving rise to the C1 clones may influence their performance in the 2013-2014 trial. We
467 further caution that comparison of GG and C1 may be biased by the disproportionate
468 amount of data available for the GG.
469 In our study, when additive and non-additive kernels were used together, the
470 ability of the additive kernel to predict the phenotype decreased, suggesting that the
471 additive kernel by itself was absorbing non-additive variance. Estimates of additive
472 genetic variance have previously been shown to capture some non-additive effects (Lu *et*
473 *al.* 1999; Su *et al.* 2012; Zuk *et al.* 2012; Muñoz *et al.* 2014). Predictions models that do
474 not explicitly incorporate non-additive components may therefore achieve gains in the
475 short-term that break down over the long-term (Cockerham and Tachida 1988; Walsh

476 2005; Hansen 2013). Including non-additive GRMs when estimating additive genetic (i.e.
477 breeding) values may therefore provide a less biased, more accurate selection of parents
478 for genomic selection (Costa E Silva *et al.* 2004; Palucci *et al.* 2007). We note that our
479 predictions of total genetic value were focused on parametric models based in
480 quantitative genetic theory. However, many non-parametric and non-linear approaches
481 are available (e.g. RKHS and random forests) that may capture even more non-additive
482 variation than found in our study.

483 Non-additive variation is prevalent in cassava, especially for low heritability traits.
484 This has many important implications for cassava breeding. It explains, in part, why
485 genetic gains have been slow (Ceballos *et al.* 2012). Inbreeding to convert dominance
486 variance to additive and better control epistatic combinations, as in maize, has been
487 suggested as a solution to non-additive genetics (Ceballos *et al.* 2015). Even for low h^2
488 traits and without inbred cassava, using the kinds of models presented in this paper, good
489 parents can be selected based on additive predictions and total genetic value can be
490 simultaneously estimated for the identification of potential commercial varieties, all
491 based on the combination of marker and preliminary field trial data(Heslot and Mark
492 2015). This approach has been previously advocated for plant breeding (Oakey *et al.*
493 2007; Heslot and Mark 2015) and has proven effective in animal breeding, e.g. (Ahlborn-
494 Breier and Hohenboken 1991; Palucci *et al.* 2007; Su *et al.* 2012; Nishio and Satoh 2014).
495 Non-additive models using genomic relationship matrices can thus improve the efficiency
496 and productivity of variety selection pipelines that are the most labor and time intensive
497 part of selecting good cassava clones after crossing.

498
499

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 A. G. O. Dixon for his development of many of the breeding lines and historical data we analyzed. Thanks also to A. I. Smith and technical teams at IITA for collection of phenotypic data and to A. Agbona and P. Peteti for data curation.

Literature Cited

(ICGMC), I. C. G. M. C., 2014 High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (Manihot esculenta Crantz) from Ten Populations. G3 Genes| Genomes| Genet.

Ahlborn-Breier, G., and W. D. Hohenboken, 1991 Additive and nonadditive genetic effects on milk production in dairy cattle: evidence for major individual heterosis. J. Dairy Sci. 74: 592–602.

Akano, O., O. Dixon, C. Mba, E. Barrera, and M. Fregene, 2002 Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor. Appl. Genet. 105: 521–525.

Bolormaa, S., J. E. Pryce, Y. Zhang, A. Reverter, W. Barendse *et al.*, 2015 Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. Genet. Sel. Evol. 47: 1–12.

Cach, N. T., J. I. Lenis, J. C. Perez, N. Morante, F. Calle *et al.*, 2006 Inheritance of useful traits in cassava grown in subhumid conditions. Plant Breed. 125: 177–182.

Cach, N. T., J. C. Perez, J. I. Lenis, F. Calle, N. Morante *et al.*, 2005 Epistasis in the expression of relevant traits in cassava (Manihot esculenta Crantz) for subhumid conditions. J. Hered. 96: 586–592.

Calle, F., J. C. Perez, W. Gaitán, N. Morante, H. Ceballos *et al.*, 2005 Diallel inheritance of relevant traits in cassava (Manihot esculenta Crantz) adapted to acid-soil savannas. Euphytica 144: 177–186.

529 Ceballos, H., R. S. Kawuki, V. E. Gracen, G. C. Yencho, and C. H. Hershey, 2015
530 Conventional breeding, marker-assisted selection, genomic selection and inbreeding
531 in clonally propagated crops: a case study for cassava. Theor. Appl. Genet.

532 Ceballos, H., P. Kulakow, and C. Hershey, 2012 Cassava Breeding: Current Status,
533 Bottlenecks and the Potential of Biotechnology Tools. Trop. Plant Biol. 5: 73–87.

534 Chalwe, A., R. Melis, P. Shanahan, and M. Chiona, 2015 Inheritance of resistance to
535 cassava green mite and other useful agronomic traits in cassava grown in Zambia.
536 Euphytica.

537 Cockerham, C. C., 1954 An Extension of the Concept of Partitioning Hereditary Variance
538 for Analysis of Covariances among Relatives When Epistasis Is Present. Genetics
539 39: 859–882.

540 Cockerham, C. C., and H. Tachida, 1988 Permanency of response to selection for
541 quantitative characters in finite populations. Proc. Natl. Acad. Sci. U. S. A. 85:
542 1563–5.

543 Costa E Silva, J., N. M. G. Borralho, and B. M. Potts, 2004 Additive and non-additive
544 genetic parameters from clonally replicated and seedling progenies of Eucalyptus
545 globulus. Theor. Appl. Genet. 108: 1113–1119.

546 Crow, J. F., 2010 On epistasis: why it is unimportant in polygenic directional selection.
547 Philos. Trans. R. Soc. Lond. B. Biol. Sci. 365: 1241–1244.

548 Doerksen, T. K., L. W. Kannenberg, and E. a Lee, 2003 Effect of Recurrent Selection on
549 Combining Ability in Maize Breeding Populations. Crop Sci. 43: 1652–1658.

550 Dudley, J. W., and G. R. Johnson, 2009 Epistatic models improve prediction of
551 performance in corn. Crop Sci. 49: 763–770.

552 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. a Poland, K. Kawamoto *et al.*, 2011 A robust,
553 simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS
554 One 6: e19379.

555 Endelman, J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with
556 R Package rrBLUP. Plant Genome J. 4: 250.

557 Fisher, R. A., 1918 The Correlation between Relatives on the Supposition of Mendelian
558 Inheritance. Trans. R. Soc. Edinburgh 52: 399–433.

559 Fuerst, C., and J. Sölkner, 1994 Additive and nonadditive genetic variances for milk yield,
560 fertility, and lifetime performance traits of dairy cattle. J. Dairy Sci. 77: 1114–1125.

561 Garrick, D. J., J. F. Taylor, and R. L. Fernando, 2009 Deregressing estimated breeding

values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41: 55.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9: e90346.

Goodnight, C. J., 1988 Epistasis and the Effect of Founder Events on the Additive Genetic Variance. Evolution (N. Y). 42: 441–454.

Hallander, J., and P. Waldmann, 2007 The effect of non-additive genetic interactions on selection in multi-locus genetic models. Heredity (Edinb). 98: 349–359.

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho, 2010 *Quantitative Genetics in Maize Breeding*. Springer-Verlag, New York.

Hamblin, M. T., and I. Y. Rabbi, 2014 The Effects of Restriction-Enzyme Choice on Properties of Genotyping-by-Sequencing Libraries: A Study in Cassava (). Crop Sci. 54: 2603.

Hansen, T. F., 2013 Why epistasis is important for selection and adaptation. Evolution (N. Y). 67: 3501–3511.

Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009 Genomic Selection for Crop Improvement. Crop Sci. 49: 1.

Henderson, C. R., 1985 Best Linear Unbiased Prediction of Nonadditive Genetic Merits in Noninbred Populations. J. Anim. Sci. 60: 111–117.

Heslot, N., and J. J. Mark, 2015 Perspectives for genomic selection applications and research in plants. Crop Sci. 55: 1–12.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4.:

Hu, Z., Y. Li, X. Song, Y. Han, X. Cai *et al.*, 2011 Genomic value prediction for quantitative traits under the epistatic model. BMC Genet. 12: 15.

Jaramillo, G., N. Morante, J. C. Pérez, F. Calle, H. Ceballos *et al.*, 2005 Diallel analysis in cassava adapted to the midaltitude valleys environment. Crop Sci. 45: 1058–1063.

Kempthorne, O., 1954 The Correlation between Relatives in a Random Mating Population. Proc. R. Soc. B Biol. Sci. 143: 103–113.

Killick, R., 1977 Genetic analysis of several traits in potatoes by means of a diallel cross. Ann. Appl. Biol. 86: 279–286.

Kulembeka, H. P., M. Ferguson, L. Herselman, E. Kanju, G. Mkamilo *et al.*, 2012 Diallel analysis of field resistance to brown streak disease in cassava (Manihot esculenta Crantz) landraces from Tanzania. Euphytica 187: 277–288.

Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi *et al.*, 2011a *Genomic Selection in Plant Breeding□: Knowledge and Prospects*.

Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi *et al.*, 2011b Genomic Selection in Plant Breeding: Knowledge and Prospects.

Lu, P. X., D. A. Huber, and T. L. White, 1999 Potential biases of incomplete linear models in heritability estimation and breeding value prediction. Can. J. For. Res. 29: 724–736.

Ly, D., M. Hamblin, I. Rabbi, G. Melaku, M. Bakare *et al.*, 2013 Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. Crop Sci. 53: 1312.

Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*.

Morota, G., R. Abdollahi-Arpanahi, A. Kranis, and D. Gianola, 2014 Genome-enabled prediction of quantitative traits in chickens using genomic annotation. BMC Genomics 15: 109.

Muñoz, P. R., M. F. R. Resende, S. a Gezan, M. Deon, and V. Resende, 2014 Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. Genetics 198: 1759–1768.

Nishio, M., and M. Satoh, 2014 Impacts of genotyping strategies on long-term genetic response in genomic selection. Anim. Sci. J. 2–7.

Oakey, H., A. P. Verbyla, B. R. Cullis, X. Wei, and W. S. Pitchford, 2007 Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. Theor. Appl. Genet. 114: 1319–1332.

Okechukwu, R. U., and a. G. O. Dixon, 2008 Genetic Gains from 30 Years of Cassava Breeding in Nigeria for Storage Root Yield and Disease Resistance in Elite Cassava Genotypes. J. Crop Improv. 22: 181–208.

Okogbenin, E., C. Egesi, B. Olasanmi, O. Ogundapo, S. Kahya *et al.*, 2012 Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. Crop Sci. 52: 2576–2586.

Oliveira, E. J., M. D. V. Resende, V. Silva Santos, C. F. Ferreira, G. A. F. Oliveira *et al.*, 2012 Genome-wide selection in cassava. Euphytica 187: 263–276.

627 Palucci, V., L. R. Schaeffer, F. Miglior, and V. Osborne, 2007 Non-additive genetic
628 effects for fertility traits in Canadian Holstein cattle (Open Access publication).
629 Genet. Sel. Evol. 39: 181–193.

630 Perez, J. C., H. Ceballos, F. Calle, N. Morante, W. Gaitán *et al.*, 2005 Within-family
631 genetic variation and epistasis in cassava (Manihot esculenta Crantz) adapted to the
632 acid-soils environment. Euphytica 145: 77–85.

633 Pérez, J. C., H. Ceballos, G. Jaramillo, N. Morante, F. Calle *et al.*, 2005 Epistasis in
634 cassava adapted to midaltitude valley environments. Crop Sci. 45: 1491–1496.

635 Prochnik, S., P. R. Marri, B. Desany, P. D. Rabinowicz, C. Kodira *et al.*, 2012 The
636 Cassava Genome: Current Progress, Future Directions. Trop. Plant Biol. 5: 88–94.

637 R Core Team, 2015 *R: A language and environment for statistical computing*. R
638 Foundation for Statistical Computing, Vienna, Austria.

639 Rabbi, I. Y., M. T. Hamblin, P. L. Kumar, M. a Gedil, A. S. Ikpan *et al.*, 2014 High-
640 resolution mapping of resistance to cassava mosaic geminiviruses in cassava using
641 genotyping-by-sequencing and its implications for breeding. Virus Res.

642 Rodriguezalmeida, F. a, L. D. Vanvleck, R. L. Willham, and S. L. Northcutt, 1995
643 Estimation of Nonadditive Genetic Variances in 3 Synthetic Lines of Beef-Cattle
644 Using an Animal-Model. J. Anim. Sci. 73: 1002–1011.

645 Satish Kumar, Claire Molloy, Patricio Muñoz, Hans Daetwyler, David Chagné, R. V.,
646 2015 Genome-enabled estimates of additive and non-additive genetic variances and
647 prediction of apple phenotypes across environments. G3 (Bethesda).

648 Su, G., O. F. Christensen, T. Ostersen, M. Henryon, and M. S. Lund, 2012 Estimating
649 Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using
650 Genome-Wide Dense Single Nucleotide Polymorphism Markers. PLoS One 7: 1–7.

651 Van Tassel, D. L., L. R. DeHaan, and T. S. Cox, 2010 Missing domesticated plant forms:
652 can artificial selection fill the gap? Evol. Appl. 3: 434–452.

653 Van Tassell, C. P., I. Misztal, and L. Varona, 2000 Method R estimates of additive
654 genetic, dominance genetic, and permanent environmental fraction of variance for
655 yield and health traits of Holsteins. J. Dairy Sci. 83: 1873–1877.

656 Tumuhimbise, R., R. Melis, and P. Shanahan, 2014 Diallel analysis of early storage root
657 yield and disease resistance traits in cassava (Manihot esculenta Crantz). F. Crop.
658 Res. 167: 86–93.

659 Turelli, M., and N. H. Barton, 2006 Will population bottlenecks and multilocus epistasis
660 increase additive genetic variance? Evolution 60: 1763–1776.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–23.

Varona, L., I. Misztal, J. K. Bertrand, and T. J. Lawlor, 1998 Effect of full sibs on additive breeding values under the dominance model for stature in United States Holsteins. J. Dairy Sci. 81: 1126–1135.

Visscher, P. M., W. G. Hill, and N. R. Wray, 2008 Heritability in the genomics era--concepts and misconceptions. Nat. Rev. Genet. 9: 255–266.

Vitezica, Z. G., L. Varona, and A. Legarra, 2013 On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195: 1223–1230.

Walsh, B., 2005 The struggle to exploit non-additive variation. Aust. J. Agric. Res. 56: 873–881.

Wang, C., D. Prakapenka, S. Wang, S. Pulugurta, H. B. Runesha *et al.*, 2014 GVCBLUP: a computer package for genomic prediction and variance component estimation of additive and dominance effects. BMC Bioinformatics 15: 270.

Wardyn, B. M., J. W. Edwards, and K. R. Lamkey, 2007 The genetic structure of a maize population: The role of dominance. Crop Sci. 47: 467–476.

Wong, W. W. L., J. Griesman, and Z. Z. Feng, 2014 Imputing genotypes using regularized generalized linear regression models. Stat. Appl. Genet. Mol. Biol. 13: 519–529.

Zacarias, a. M., and M. T. Labuschagne, 2010 Diallel analysis of cassava brown streak disease, yield and yield related characteristics in Mozambique. Euphytica 176: 309–320.

Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander, 2012 The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. 109: 1193–1198.

Figure & Table Legends

Table 1. Additive plus non-additive effects models tested and their abbreviations.

Table 2. Single-step multi-environment model results for the Genetic Gain population. Variance components (± standard errors), narrow-sense heritabilities (h^2), proportion of the total phenotypic variance explained by dominance (d^2), additive-by-additive epistasis ($i^2_{A\#A}$), additive-by-dominance epistasis ($i^2_{A\#D}$) and broad-sense heritability (H^2) are provided. Model log-likelihoods and Akaike Information Criterion (AIC) are also given. The best model for each trait (lowest AIC) is highlighted in grey.

Table 3. Single-step multi-environment model results for the Cycle 1 population. Variance components (± standard errors), narrow-sense heritabilities (h^2), proportion of the total phenotypic variance explained by dominance (d^2), additive-by-additive epistasis ($i^2_{A\#A}$), additive-by-dominance epistasis ($i^2_{A\#D}$) and broad-sense heritability (H^2) are provided. Model log-likelihoods and Akaike Information Criterion (AIC) are also given. The best model for each trait (lowest AIC) is highlighted in grey.

Figure 1. Partitioning of broad-sense heritability among variance components for single-step multi-environment models in the Genetic Gain and Cycle 1 populations for each trait. Results from each of five models are shown in each panel broken down by trait (rows) and population (columns). Models include additive only (Additive), dominance only (Dominance), Additive plus Dominance (Add_Dom), Additive plus dominance plus either AxA epistasis (Add_Add_Epistasis) or AxD epistasis (Add_Dom_Epistasis).

Figure 2. Partitioning of broad-sense heritability among variance components for the within-trial models in the Genetic Gain population. Three models were fitted for each trait in each of 47 Genetic Gain trials. Each panel contains boxplots showing the distribution proportions of the phenotypic variability explained by a corresponding variance components, including the broad-sense heritability (H^2) across each trial. Red horizontal lines are the median narrow-sense heritability (h^2) from the additive only model. Traits are on columns and three models are on the rows: additive plus dominance (Add_Dom), additive plus dominance plus AxA epistasis (AxA_epi) and additive plus dominance plus AxD epistasis (AxD_epi).

Figure 3. Accuracy of additive genetic value prediction in the Genetic Gain and Cycle 1 populations. Boxplots showing the distribution over 25 replicates of 5-fold cross-validation of the prediction accuracy of the additive component from four different models are shown in each panel. The accuracy within the Genetic Gain (red) and Cycle 1 (blue) are shown. Traits are in the columns. Accuracy is defined as the correlation between the prediction from the additive kernel in the model and the BLUP from the first stage of analysis where location, year and replicate variability were removed. Models included are: additive only (Add), additive plus dominance (Add_Dom), additive plus dominance plus AxA epistasis (AxA_epi) and additive plus dominance plus AxD epistasis (AxD_epi).

Figure 4. Accuracy of total genetic value prediction in the Genetic Gain and Cycle 1 populations. Boxplots showing the distribution over 25 replicates of 5-fold cross-validation of the prediction accuracy of the total genetic value from five different models are shown in each panel. The accuracy within the Genetic Gain (red) and Cycle 1 (blue) are shown. Traits are in the columns. Accuracy is defined as the correlation between the sum of predictions from all genetic variance components in the model and the BLUP from the first stage of analysis where location, year and replicate variability were removed. Models included are: additive only (Add), dominance only (Dom), additive plus dominance (Add_Dom), additive plus dominance plus AxA epistasis (AxA_epi) and additive plus dominance plus AxD epistasis (AxD_epi).

Supplementary Table 1. Pedigree and related information for the IITA: Genetic Gain population, Nigeria.

Supplementary Table 2. Sample size (Nobs), replication number (Nreps), clone number (Nclones) and whether a trial (LOC.YEAR) was included in one-step multi-environment models for both populations analyzed.

Supplementary Table 3. Pedigree information for the IITA: GS Cycle 1 population, Nigeria.

Supplementary Table 4. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance model in the Genetic Gain population, by trait.

Supplementary Table 5. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance plus additive-by-additive epistasis model in the Genetic Gain population, by trait.

Supplementary Table 6. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance plus additive-by-dominance epistasis model in the Genetic Gain population, by trait.

Supplementary Table 7. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance model in the Cycle 1 population, by trait.

Supplementary Table 8. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance plus additive-by-additive epistasis model in the Cycle 1 population, by trait.

Supplementary Table 9. Asymptotic correlation matrices of parameter estimates representing the dependency of variance component estimation for the additive plus dominance plus additive-by-dominance epistasis model in the Cycle 1 population, by trait.

Supplementary Table 10. Within-trial model results for the Genetic Gain population. The mean (± standard errors) across 47 trials for each trait and model fitted is given for the following model parameters: variance components, narrow-sense heritabilities (h^2), proportion of the total phenotypic variance explained by dominance (d^2), additive-by-additive epistasis ($i^2_{A\#A}$), additive-by-dominance epistasis ($i^2_{A\#D}$) and broad-sense heritability (H^2), trial sample size (N), model log-likelihoods and Akaike Information Criterion (AIC) are also given. The best model for each trait (lowest AIC) is highlighted in grey.

Supplementary Table 11. Results from 25 replicates of 5-fold cross-validation in the Genetic Gain population are given. Mean (± standard errors) across the 25 replicates are given for prediction accuracy of each kernel plus total genetic value (sum across all kernels), variance components (V_g and V_e) and kernel weights.

Supplementary Table 12. Results from 25 replicates of 5-fold cross-validation in the Cycle 1 population are given. Mean (± standard errors) across the 25 replicates are given for prediction accuracy of each kernel plus total genetic value (sum across all kernels), variance components (V_g and V_e) and kernel weights.

803 Table 1.

Model	Relationship Matrices / Variance Components
A	Additive
D	Dominance
A + D	Additive + Dominance
A + D + AA	Additive + Dominance + A#A Epistasis
A + D + AD	Additive + Dominance + A#D Epistasis

804 Table 2

Trait	Model	$\sigma^2_{loc.year}$	σ^2_{rep}	σ^2_{add}	σ^2_{dom}	$\sigma^2_{A\#A}$	$\sigma^2_{A\#D}$	σ^2_{error}	h^2	d^2	$i^2_{A\#A}$	$i^2_{A\#D}$	H^2	loglik	AIC
DM	Additive	0.025 ± 0.48	6.16 ± 5.44	10.1 ± 1.00				15.36 ± 0.33	0.33				0.33	-9457	18921.9
	Dominance	0.017 ± 0.48	6.17 ± 5.44		7.2 ± 0.65			15.29 ± 0.33		0.25			0.25	-9470	18947.7
	Add + Dom	0.010 ± 0.48	6.17 ± 5.44	8.5 ± 1.64	1.1 ± 0.90			15.33 ± 0.33	0.27	0.04			0.31	-9456	18932.4
	Add + Dom + AxA Epi	0.007 ± 0.48	6.17 ± 5.44	8.2 ± 1.64	0.2 ± 1.10	1.7 ± 1.58		15.31 ± 0.33	0.26	0.01	0.05		0.32	-9455	18933.0
	Add + Dom + AxD Epi	0.006 ± 0.48	6.18 ± 5.44	8.1 ± 1.62	0.0 ± 1.19		1.5 ± 1.27	15.30 ± 0.33	0.26	0.00		0.05	0.31	-9455	18932.6
HI	Additive	48.9 ± 30.5	11.8 ± 9.88	87.1 ± 7.40				10.8 ± 2.03	0.3				0.3	10656	-21305.1
	Dominance	48.7 ± 30.4	11.8 ± 9.88		58.8 ± 4.77			10.8 ± 2.03		0.3			0.3	10646	-21285.0
	Add + Dom	48.9 ± 30.4	11.8 ± 9.88	57.3 ± 12.60	17.6 ± 6.99			10.8 ± 2.03	0.2	0.1			0.3	10659	-21309.1
	Add + Dom + AxA Epi	48.9 ± 30.4	11.8 ± 9.88	55.6 ± 12.61	1.1 ± 8.99	8.3 ± 14.04		10.8 ± 2.03	0.2	0.1	0.00		0.3	10659	-21307.4
	Add + Dom + AxD Epi	48.9 ± 30.4	11.8 ± 9.88	54.1 ± 11.82	7.7 ± 9.39		2.0 ± 9.44	10.8 ± 2.03	0.2	0.0		0.01	0.3	10660	-21308.5
logRTWT	Additive	0.055 ± 0.03	0.014 ± 0.01	0.095 ± 0.09				0.17 ± 0.03	0.28				0.28	2362.1	-4716.1
	Dominance	0.056 ± 0.03	0.014 ± 0.01		0.064 ± 0.05			0.17 ± 0.03		0.21			0.21	2369.5	-4731.0
	Add + Dom	0.056 ± 0.03	0.014 ± 0.01	0.036 ± 0.13	0.037 ± 0.09			0.17 ± 0.03	0.11	0.12			0.23	2375.4	-4740.8
	Add + Dom + AxA Epi	0.056 ± 0.03	0.014 ± 0.01	0.029 ± 0.12	0.020 ± 0.11	0.033 ± 0.14		0.17 ± 0.03	0.09	0.06	0.0		0.25	2378.1	-4744.2
	Add + Dom + AxD Epi	0.056 ± 0.03	0.014 ± 0.01	0.031 ± 0.12	0.017 ± 0.12		0.025 ± 0.12	0.17 ± 0.03	0.10	0.06		0.08	0.23	2377.8	-4743.6
MCBS	Additive	0.073 ± 0.23	0.020 ± 0.01	0.029 ± 0.04				0.19 ± 0.03	0.04				0.04	2383.9	-4759.77
	Dominance	0.074 ± 0.23	0.020 ± 0.01		0.022 ± 0.03			0.19 ± 0.03		0.03			0.03	2374.9	-4741.84
	Add + Dom	0.074 ± 0.23	0.020 ± 0.01	0.023 ± 0.06	0.005 ± 0.04			0.19 ± 0.03	0.03	0.01			0.04	2384.4	-4759.56
	Add + Dom + AxA Epi	0.074 ± 0.23	0.020 ± 0.01	0.022 ± 0.07	0.004 ± 0.05	0.003 ± 0.07		0.19 ± 0.03	0.03	0.01	0.0		0.04	2384.7	-4757.78

	Add + Dom	0.4		0.		0.0		0.		0.0		0.0		0.0		0.0					0.0		0.0	0.1		0.0		0.	0.		0.0	0.	2384.	-4757
	+ AxD Epi	74	±	23		20	±	01		23	±	07		05	±	06					00	±	06	9	±	03		03	01		0	04	78	.56
MCM		0.0		0.		0.0		0.		1.4		0.0												0.3		0.0		0.				0.		1255
DS	Additive	51	±	02		00	±	00		4	±	9												4	±	06		79				79	-624	.7
		0.0		0.		0.0		0.						0.7		0.0								0.3		0.0			0.			0.		1207
	Dominance	51	±	02		00	±	00						76	±	5								4	±	06			66			66	-600	.6
		0.0		0.		0.0		0.		0.3		0.1		0.5		0.0								0.3		0.0		0.	0.			0.		1202
	Add + Dom	51	±	02		00	±	00		0	±	2		67	±	8								4	±	06		24	45			69	-596	.1
	Add + Dom	0.0		0.		0.0		0.		0.2		0.1		0.1		0.0		0.5		0.1				0.3		0.0		0.	0.	0.4		0.		1180
	+ AxA Epi	51	±	02		00	±	00		8	±	1		55	±	9		89	±	2				4	±	06		20	11	2		72	-584	.7
	Add + Dom	0.0		0.		0.0		0.		0.3		0.1		0.0		0.0					0.5		0.0	0.3		0.0		0.	0.		0.4	0.		1173
	+ AxD Epi	51	±	02		00	±	00		2	±	0		00	±	8					56	±	9	4	±	06		25	00		4	69	-581	.0
SPRO		28.		13				0.		70.														22				0.				0.	10384	-2076
UT	Additive	2	±	.6		0.3	±	5		4	±	7.4												8.2	±	3.8		2				2	.9	1.9
		28.		13				0.						48.										22					0.			0.	10397	-2078
	Dominance	2	±	.6		0.3	±	5						1	±	4.8								6.5	±	3.8			2			2	.4	6.8
		28.		13				0.		19.		10.		34.										22				0.	0.			0.	10399	-2078
	Add + Dom	2	±	.6		0.3	±	5		3	±	8		2	±	8.2								6.5	±	3.8		1	1			2	.5	9.1
	Add + Dom	28.		13				0.		10.								55.		12.				22				0.	0.			0.	10410	-2080
	+ AxA Epi	1	±	.5		0.3	±	5		9	±	9.1		5.8	±	8.6		9	±	6				5.4	±	3.8		0	0	0.2		2	.4	8.7
	Add + Dom	28.		13				0.		12.											36.		10.	22				0.	0.			0.	10406	-2080
	+ AxD Epi	1	±	.5		0.3	±	5		2	±	9.4		8.4	±	9.8					5	±	4	5.6	±	3.8		0	0		0.1	2	.0	0.0

805

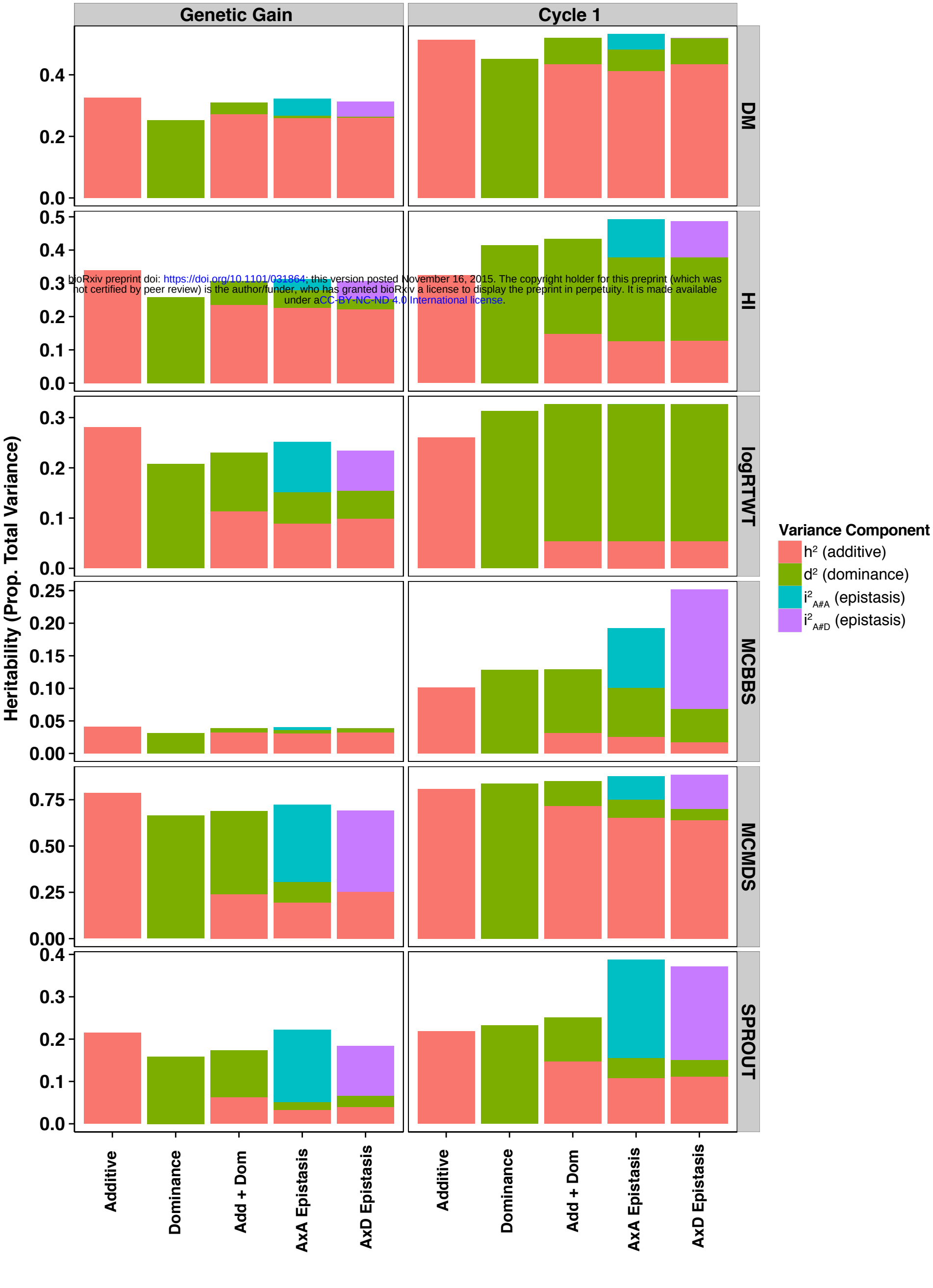
806

807 Table 3

Trait	Model	$\sigma^2_{loc.year}$	σ^2_{add}	σ^2_{dom}	$\sigma^2_{A\#A}$	$\sigma^2_{A\#D}$	σ^2_{error}	h^2	d^2	$i^2_{A\#A}$	$i^2_{A\#D}$	H^2	loglik	AIC
DM	Additive	8.32 ± 8.3	20.9 ± 2.3				11.5 ± 0.6	0.51				0.51	-3544.1	7094.316
	Dominance	8.68 ± 8.7		16.0 ± 1.6			10.6 ± 0.7		0.45			0.45	-3585.3	7176.634
	Add + Dom	8.38 ± 8.4	17.3 ± 2.5	3.4 ± 1.5			10.7 ± 0.7	0.43	0.08			0.52	-3537.9	7083.907
	Add + Dom + AxA Epi	8.42 ± 8.4	16.6 ± 2.5	2.8 ± 1.6	2.050 ± 91		10.4 ± 0.8	0.41	0.07	0.05		0.53	-3537.5	7084.999
	Add + Dom + AxD Epi	8.38 ± 8.4	17.3 ± 2.6	3.4 ± 1.7		0.001 ± 40	10.7 ± 0.9	0.43	0.08		0.02	0.52	-3537.9	7085.908
HI	Additive	19.5 ± 7	69.2 ± 6				124.7 ± 4.7	0.32				0.32	-3653.6	7301.267
	Dominance	19.1 ± 3		86.6 ± 2			103.1 ± 5.2		0.41			0.41	-3666.1	7326.248
	Add + Dom	19.2 ± 4	32.3 ± 3	8.61 ± 0			103.8 ± 5.2	0.13	0.25			0.38	-3672.2	7336.512
	Add + Dom + AxA Epi	19.2 ± 3	29.2 ± 4	57.8 ± 6	4.07 ± 7		97.4 ± 6.3	0.13	0.25	0.11		0.49	-3673.1	7336.275
	Add + Dom + AxD Epi	19.2 ± 4	28.8 ± 4	56.4 ± 1		4.001 ± 6	96.0 ± 6.9	0.13	0.25		0.12	0.49	-3673.0	7336.183
logRT WT	Additive	0.007 ± 08	0.101 ± 18				0.28 ± 10	0.06				0.06	-160.58	315.160
	Dominance	0.006 ± 07		0.116 ± 18			0.25 ± 11		0.31			0.31	-183.50	361.018
	Add + Dom	0.007 ± 07	0.020 ± 19	0.003 ± 22			0.25 ± 11	0.05	0.07			0.13	-184.19	360.385
	Add + Dom + AxA Epi	0.007 ± 07	0.020 ± 20	0.003 ± 24	0.000 ± 34		0.25 ± 14	0.05	0.07	0.00		0.13	-184.19	358.385
	Add + Dom + AxD Epi	0.007 ± 07	0.020 ± 20	0.003 ± 25		0.000 ± 35	0.25 ± 15	0.05	0.07		0.00	0.13	-184.19	358.385
MCBBS	Additive	0.047 ± 5	0.026 ± 07				0.19 ± 06	0.10				0.10	-751.23	1496.477
	Dominance	0.048 ± 5		0.033 ± 08			0.18 ± 07		0.13			0.13	-754.60	1503.206
	Add + Dom	0.048 ± 5	0.008 ± 09	0.025 ± 10			0.18 ± 07	0.10	0.03			0.13	-754.98	1501.92

																															64
	Add + Dom + AxA Epi	0.0 47	±	0.0 5	0.0 07	±	0.0 09	0.0 20	±	0.0 10	0.0 25	±	0.0 20							0.1 7	±	0.0 08	0.0 3	0.0 8	0.0 9		0.1 9	755.66 6	-1501.3 31
	Add + Dom + AxD Epi	0.0 47	±	0.0 5	0.0 05	±	0.0 09	0.0 14	±	0.0 10				0.0 51	±	0.0 21				0.1 6	±	0.0 10	0.0 2	0.0 5		0.1 8	0.2 5	757.78 3	-1505.5 66
MCM DS	Additive	0.0 47	±	0.0 5	2.1 99	±	0.1 59													0.4 8	±	0.0 26	0.8 1				0.8 1	-1205.9 13	2417.8 26
	Dominance	0.0 55	±	0.0 6				1.9 93	±	0.1 05										0.3 3	±	0.0 26		0.8 4			0.8 4	-1370.1 89	2746.3 78
	Add + Dom	0.0 49	±	0.0 5	1.9 84	±	0.1 79	0.3 71	±	0.0 80										0.3 7	±	0.0 26	0.7 2	0.1 3			0.8 5	-1194.1 53	2396.3 06
	Add + Dom + AxA Epi	0.0 52	±	0.0 53	1.8 91	±	0.1 83	0.2 84	±	0.0 86	0.3 61	±	0.1 20							0.3 1	±	0.0 25	0.6 5	0.1 0	0.1 2		0.8 8	-1190.8 57	2391.7 14
	Add + Dom + AxD Epi	0.0 54	±	0.0 55	1.7 80	±	0.1 78	0.1 72	±	0.0 82				0.5 14	±	0.1 01				0.2 6	±	0.0 23	0.6 4	0.0 6		0.1 8	0.8 9	-1184.4 64	2378.9 28
SPRO UT	Additive	25. 3	±	25. 8	130 .0	±	25. 0													438 .8	±	14. 8	0.2 2				0.2 2	-2462.2 02	4918.4 04
	Dominance	25. 8	±	26. 4				133 .3	±	23. 5										414 .5	±	16. 5		0.2 3			0.2 3	-2459.6 33	4913.2 66
	Add + Dom	25. 5	±	26. 1	87. 3	±	30. 3	61. 1	±	27. 5										418 .4	±	16. 4	0.1 5	0.1 0			0.2 5	-2467.0 84	4926.1 67
	Add + Dom + AxA Epi	26. 5	±	27. 1	72. 2	±	29. 4	31. 7	±	26. 7	154 .5	±	58. 9							382 .0	±	21. 1	0.1 1	0.0 5	0.2 3		0.3 9	-2470.6 44	4931.2 89
	Add + Dom + AxD Epi	26. 9	±	27. 5	71. 1	±	29. 3	25. 7	±	26. 7				140 .4	±	55. 1				373 .9	±	23. 7	0.1 1	0.0 4		0.2 2	0.3 7	-2470.4 71	4930.9 42

808



DM
HI
logRTWT
MCBBS
MCMDS
SPROUT
Add_Dom
AxA_epi
AxD_epi
Heritability (Prop. Total Var.)
1.00
0.75
0.50
0.25
0.00
h2.add
h2.dom
h2.epi
H2
Variance Component
Prop. Explained
h2.add
h2.dom
h2.epi
H2

DM
HI
logRTWT
MCBBS
MCMDS
SPROUT
0.6
0.4
0.2
Accuracy (Additive Value)
ADD
ADD_DOM
ADD_DOM_AxA
ADD_DOM_AxD
Model
Population
Genetic_Gain
Cycle1

DM
HI
logRTWT
MCBBS
MCMDS
SPROUT
Accuracy (Tot. Gen. Value)
0.6
0.5
0.4
0.3
0.2
ADD
DOM
ADD_DOM
ADD_DOM_AxA
ADD_DOM_AxD
Model
Population
Genetic_Gain
Cycle1