Epistasis detection and modeling for genomic selection in cowpea (*Vigna unguiculata.* L. Walp.)

Marcus O. Olatoye[1], Zhenbin Hu[2], and Peter O. Aikpokpodion*[3]

[1]Dept. of Crop Sciences, University of Illinois, Urbana-Champaign, IL, [2]Dept. of Agronomy, Kansas State University, Manhattan, KS. [3]Department of Genetics and Biotechnology, University of Calabar, Nigeria

Corresponding Author: Peter O. Aikpokpodion

Address: Department of Genetics and Biotechnology, University of Calabar, Nigeria

Email: paikpokpodion@unical.edu.ng; paikpokpodion@gmail.com

ORCID iD: 0000-0002-4553-8299

Abstract

Genetic architecture reflects the pattern of effects and interaction of genes underling phenotypic variation. Most mapping and breeding approaches generally consider the additive part of variation but offer limited knowledge on the benefits of epistasis which explains in part the variation observed in traits. In this study, the cowpea multiparent advanced generation inter-cross (MAGIC) population was used to characterize the epistatic genetic architecture of flowering time, maturity, and seed size. In addition, considerations for epistatic genetic architecture in genomic-enabled breeding (GEB) was investigated using parametric, semi-parametric, and non-parametric genomic selection (GS) models. Our results showed that large and moderate effect sized two-way epistatic interactions underlie the traits examined. Flowering time QTL colocalized with cowpea putative orthologs of *Arabidopsis thaliana* and *Glycine max* genes like *PHYTOCLOCK1* (*PCL1* [Vigun11g157600]) and *PHYTOCHROME A* (*PHY A* [Vigun01g205500]). Flowering time adaptation to long and short photoperiod was found to be controlled by distinct and common main and epistatic loci. Parametric and semi-parametric GS models outperformed non-parametric GS model. Using known QTL as fixed effects in GS models improved prediction accuracy when traits were controlled by both large and moderate effect QTL. In general, our study demonstrated that prior understanding the genetic architecture of a trait can help make informed decisions in GEB. This is the first report to characterize epistasis and provide insights into the underpinnings of GS versus marker assisted selection in cowpea.

Keywords: Cowpea, Genetic architecture, Epistasis, QTL, Genomic-enabled breeding, Genomic selection, flowering time, and photoperiod.

Introduction

Asymmetric transgressive variation in quantitative traits is usuallly controlled by non-additive gene action known as epistasis (Rieseberg, Archer and Wayne, 1999). Epistasis has been defined as the interaction of alleles at multiple loci (Mathew *et al.*, 2018). The joint effect of the alleles at these loci may be lower or higher than the total effects of these loci (Johnson, 2008). In selfing species, epistasis is common due to high level of homozygousity (Volis *et al.*, 2010) and epistatic interactions have been found among loci underlying flowering time in barley (Mathew *et al.*, 2018), rice (Chen *et al.*, 2015; M. Chen *et al.*, 2018), and sorghum (Li *et al.*, 2018). However, the direct quantification of the importance of epistasis for breeding purposes has not been fully realized due to the fact that most of the current statistical models cannot efficiently characterize or account for epistasis (Mackay, 2001; Moore and Williams, 2009; Sun, Ma and Mumm, 2012; Mathew *et al.*, 2018). Common quantitative traits mapping approaches are often single-locus analysis techniques. These techniques focus on the additive contribution of genomic loci (H.Barton and D.Keightley, 2002) which may explain a fraction of the genetic variation; thus leading to missing heritability.

Regardless of the limitations of genomic mapping approaches, characterization of the genetic basis of complex agronomic traits has been beneficial for breeding purposes. For example, markers tagging quantitative trait loci (QTL) have been used in marker-assisted selection (MAS) in breeding programs (Zhang *et al.* 2003; Pan *et al.* 2006; Saghai Maroof *et al.* 2008; Foolad and Panthee 2012; Massman *et al.* 2013; Mohamed *et al.* 2014; Zhao *et al.* 2014). However, the efficiency of QTL based MAS approach in breeding is limited. First, the small sample size of bi-parental populations where QTL are detected often result in overestimation of the respective QTL effect sizes; a phenomenon known as Beavis effect (Utz, Melchinger and Schön, 2000; Xu, 2003; King and Long, 2017). Second, genetic diversity is limited to the two parents forming the bi-parental population, thus QTL may not reflect the entire variation responsible for the trait and may not be transferable to other genetic backgrounds (Xu *et al.*, 2017). Third, linkage mapping is limited in power to detect small effect loci, thus only the available large effects loci are used for MAS (Ben-Ari and Lavi 2012). Notably, MAS is more efficient with traits controlled by few genomic loci and not polygenic traits (Bernardo, 2008). In contrast, genomic selection (GS) that employs genome wide markers has been found to be more suited for complex traits, and also having higher response to selection than MAS (Bernardo and Yu, 2007; Wong and Bernardo, 2008; Cerrudo *et al.*, 2018).

In GS, a set of genotyped and phenotyped individuals are first used to train a model that estimates the genomic estimated breeding values (GEBVs) of un-phenotyped but genotyped individuals (Jannink, Lorenz and Iwata, 2010). GS models often vary in performance with the genetic architecture of traits. Parametric GS models are known to capture additive genetic effects but not efficient with epistatic effects due to the computational burden of high-order interactions (Moore and Williams, 2009; Howard, Carriquiry and Beavis, 2014). Parametric GS models with incorporated kernels (marker based relationship matrix) for epistasis have recently been developed (Covarrubias-pazaran, 2017). Semi-parametric and non-parametric GS models capturing epistatic interactions have been developed and implemented in plant breeding (Gianola, Fernando and Stella, 2006; Gianola and de los Campos, 2008; De Los Campos *et al.*, 2010). Semi-parametric models as Reproducing Kernel Hilbert Space (RKHS) reduces parametric space dimensions to efficiently capture epistatic interactions among markers (Jiang and Reif, 2015; de Oliveira Couto *et al.*, 2017). Using simulated data, Howard *et al.* 2014 showed that semi-parametric and non-parametric GS models can improve prediction accuracies under epistatic genetic architectures. In general, GS has been widely studied in and applied to major crop species including both cereals and legumes. However, in orphan crop species, applications of genomic-enabled breeding (GEB) methods is still limited (Varshney *et al.*, 2012).

Cowpea (*Vigna unguiculata* L. Walp) is a widely adapted warm-season orphan herbaceous leguminous annual crop and an important source of protein in developing countries (Muchero *et al.*, 2009; Varshney *et al.*, 2012; Boukar *et al.*, 2018; Huynh *et al.*, 2018). Cowpea is cultivated over 12.5 million hectares in tropical and sub-tropical zones of the world including Sub-Saharan Africa, Asia, South America, Central America, the Caribbean, United States of America and around the Mediterranean Sea. However, more than 95 *per cent* of cultivation takes place in Sub-Saharan Africa (Boukar *et al.*, 2018). It is the most economically important African leguminous crop and of vital importance to the livelihood of several millions of people. Due to its flexibility as a “hungry season crop” (Langyintuo *et al.*, 2003), cowpea is part of the rural families’ coping strategies to mitigate the effect of changing climatic conditions.

Cowpea’s nitrogen fixing and drought tolerance capabilities make it a valuable crop for low-input and smallholder farming systems (Hall *et al.*, 2003; Boukar *et al.*, 2018). Breeding efforts using classical approaches have been made to improve cowpea’s tolerance to both biotic (disease and pest) and abiotic (drought and heat) stressors (Hall *et al.*, 2003; Hall, 2004). Advances in applications of next generation sequencing (NGS) and development

of genomic resources (consensus map, draft genome, and multiparent population) in cowpea have provided the opportunity for the exploration for GEB (Muchero *et al.*, 2009; Boukar *et al.*, 2018; Huynh *et al.*, 2018). MAS and GS have improved genetic gain in soybean (*Glycine max*) (Jarquin, Specht and Lorenz, 2016; Kurek, 2018; Matei *et al.*, 2018) and common bean (*Phaseolus vulgaris*) (Schneider, Brothers and Kelly, 1997; Yu, Park and Poysa, 2000; Wen *et al.*, 2019). However, cowpea still lags behind major legumes in the area of GEB applications. GEB has the potential of enabling expedited cowpea breeding to ensure food security in developing countries where national breeding programs still depend on labor-intensive and time-consuming classical breeding approaches.

In this study, the cowpea multiparent advanced generation inter-cross (MAGIC) population was used to explore MAS and GS. The cowpea MAGIC population was derived from intercrossing among eight founder lines (Huynh *et al.*, 2018) and offers greater genetic diversity than bi-parentals to identify higher-order epistatic interactions (Mathew *et al.*, 2018). Although, theoretical models and empirical studies involving simulations have suggested the significant role for epistasis in breeding (Melchinger *et al.*, 2007; Volis *et al.*, 2010; Messina *et al.*, 2011; Howard, Carriquiry and Beavis, 2014); empirical evidence from practical breeding are limited. Therefore, the epistatic genetic architecture of three traits in cowpea was evaluated alongside its considerations in genomic enabled breeding using parametric, semi-parametric, and non-parametric GS models.

Materials and Methods

Plant genetic resource and phenotypic evaluation

This study was performed using publicly available cowpea MAGIC population's phenotypic and genotypic data (Huynh *et al.*, 2018). The MAGIC population was derived from intercross between eight founders. The F_1s were derived from eight-way intercross between the founders and were subsequently selfed through single seed descent for eight generations. The F_8 RILs were later genotyped with 51,128 SNPs using the Illumina Cowpea Consortium Array. A core set of 305 MAGIC RILs were selected and phenotyped (Huynh *et al.*, 2018). The RILs were evaluated under two irrigation regimes.

In this study, the flowering time (FLT), maturity (MAT), and seed size (SS) data were analyzed for environment-by-environment correlations and best linear unbiased predictions (BLUPs). The traits analyzed in this study are; FTFILD (flowering time under full irrigation and long day), FTRILD (flowering time under restricted irrigation and long day), FTFISD (flowering time under full irrigation and short day), FTRISD (flowering time under restricted

irrigation and short day), FLT_BLUP (BLUP of flowering time across environments), MFISD (maturity under full irrigation and short day), MRISD (maturity under restricted irrigation and short day), MAT_BLUP (BLUP of maturity across environments), SSFISD (seed size under full irrigation and short day), SSRISD (seed size under restricted irrigation and short day), SS_BLUP(BLUP of seed size across environments). In addition, using both genomic and phenotypic data, narrow sense heritability was estimated using RRBLUP package in R (Endelman, 2011).

QTL and Epistasis Mapping

QTL mapping was performed for all traits using the stepwise regression model implemented in TASSEL 5.0 standalone version (Bradbury *et al.*, 2007). The approach implements both forward inclusion and backward elimination steps. The model accounts for major effect loci and reduces collinearity among markers. The model was designed for multi-parental populations and no family term was used in the model since MAGIC population development involved several steps of intercross that reshuffles the genome and minimizes phenotype-genotype covariance. A total of 32,130 SNPs across 305 RILs were used in the analysis. A permutation of 1000 was used in the analysis.

To characterize the epistatic genetic architecture underlying flowering time, maturity, and seed size, the Stepwise Procedure for constructing an Additive and Epistatic Multi-Locus model (SPAEML; (Chen *et al.*, 2018)) epistasis pipeline implemented in TASSEL 5.0 was used to perform epistasis mapping for phenotypic traits (FTFILD, FTRILD, FTFISD, FTRISD, FT_BLUP, MFISD, MRISD, MT_BLUP, SSFISD, SSRISD, and SS_BLUP). One critical advantage of SPAEML that led to its consideration for this study is its ability to correctly distinguish between additive and epistatic QTL. SPAEML source code is available at https://bitbucket.org/wdmetcalf/tassel-5-threaded-model-fitter. The minor allele frequency of each QTL was estimated using a custom R script from http://evachan.org/rscripts.html. The proportion of phenotypic variation explained (PVE) by each QTL from both QTL and Epistasis mapping was estimated by multiplying the R^2 obtained from fitting a regression between the QTL and the trait of interest by 100. The regression model for estimating PVE is;

$$y_{ij} = \mu + \gamma_i + + \varepsilon_{ij} \qquad [1]$$

where y_{ij} is the phenotype, μ is the overall mean, γ_i is the term for QTL, and ε_{ij} is the residual term.

A set of *a priori* genes (n=100; Data S1) was developed from *Arabidopsis thaliana* and *Glycine max* flowering time and seed size genes obtained from literature and https://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes. The cowpea orthologs of these genes were obtained by blasting the *A. thaliana* and *G. max* sequence of the *a priori* genes on the new *Vigna* genome assembly *v.1* on Phytozome (Goodstein *et al.*, 2012). The corresponding cowpea gene with the highest score was selected as a putative ortholog. Colocalizations between the cowpea putative orthologs and QTL were identified using a custom R script.

Marker Assisted Selection Pipeline

In order to evaluate the performance of MAS in cowpea, a custom pipeline was developed in R. First, using subbagging approach, 80% of the 305 RILs randomly sampled without replacement was used as the training population; followed by performing a Multi-locus GWAS (Multi-locus Mixed Model, MLMM) (Segura *et al.*, 2012) on both genomic and phenotypic data of the training population. The MLMM approach implements stepwise regression involving both forward and backward regressions. This model accounts for major effect loci and reduces the effect of allelic heterogeneity. A K-only model that accounts for a random polygenic term (kinship relationship matrix) was used in the MLMM model. No term for population structure was used in the model since MAGIC population development involved several steps of intercross that reshuffles the genome and minimizes phenotype-genotype covariance. A total of 32130 SNPs across 305 RILs were used in the GWAS analysis and coded as -1 and 1 for homozygous SNPs and 0 for heterozygous SNPs. Bonferroni correction with $\alpha = 0.05$ was used to determine the cut-off threshold for each trait association (α/total number of markers = 1.6 e-06).

$$y = X\beta + Zu + e \qquad [2]$$

where y is the vector of phenotypic data, β is a vector of fixed effects other than SNP or population structure effects; u is an unknown vector of random additive genetic effects from multiple background QTL for RILs. X and Z are incident matrices of 1s and 0s relating y to β and u (Yu *et al.*, 2006).

Second, top three most significant associations were then selected from the genomic data of the training population to train a regression model by fitting the SNPs in a regression analysis with the phenotypic information. This training model was later used alongside the predict function in R to predict the phenotypic information of the validation population (20% that remained after sub-setting the training population). The prediction accuracy of MAS was

obtained as the correlation between this predicted phenotypic information and the observed phenotypic information for the validation data.

Genomic Selection Pipeline

In order to evaluate the performance of using known QTL as fixed effects in GS models and to compare the performance of parametric, semi-parametric and non-parametric GS models; a custom GS pipeline was developed in R. The GS pipeline was made up of four GS models, which were named as FxRRBLUP (Ridge Regression BLUP where markers were fitted as both fixed and random effects; parametric), RRBLUP (RRBLUP where markers were only fitted as random effects; parametric), Reproducing Kernel Hilbert Space (RKHS; semi-parametric), and Support Vector Regression (SVR; non-parametric). First, using subagging approach, 80% of the RILs were randomly sampled without replacement (training population) followed by running MLMM GWAS and selecting the three most significant associations, which were used as fixed effects in the FxRRBLUP. These three SNPs were removed from the rest of SNPs that were fitted as random effects in the FxRRBLUP model. The RRBLUP, RKHS, and SVR models were fitted simultaneously in the same cycle as FxRRBLUP to ensure unbiased comparison of GS models. Likewise, in order to ensure unbiased comparison between GS and MAS approaches; similar seed numbers were used for the subagging sampling of training populations across 100 cycles for GS and MAS. The validation set was composed of the remaining 20% of the RILs after sampling the 80% (training set). Prediction accuracy in GS was estimated as the Pearson correlation between measured phenotype and genomic estimated breeding values of the validation population. Also, for flowering time, each environment was used as a training population to predict the other three environments.

Ridge Regression BLUP (RRBLUP)

The RRBLUP models without and with fixed effects can be described as;

$$\boldsymbol{y} = \mu + \sum_{m=1}^{p} \boldsymbol{Z}_m u_m + \boldsymbol{e} \qquad [3]$$

$$\boldsymbol{y} = \mu + \sum_{k=1}^{q} \boldsymbol{X}_k \alpha_k + \sum_{m=1}^{p} \boldsymbol{Z}_m u_m + \boldsymbol{e} \qquad [4]$$

where **y** is the vector (n x 1) of observations (simulated phenotypic data), μ is the vector of the general mean, q is the number of selected significant associated markers (q=3), $\boldsymbol{X_k}$ is the k^{th} column of the design matrix $\boldsymbol{X}$, α is the fixed additive effect associated with markers k . . .

q, $\boldsymbol{u}$ random effects term, with $E(u_m) = 0$, $Var(u_m) = \sigma^2_{u_m}$(variance of marker effect), *p* is the marker number (*p* > *n*), $\boldsymbol{Z_m}$ is the m^{th} column of the design matrix $\boldsymbol{Z}$, *u* is the vector of random marker effects associated with markers *m* . . . *p.* In the model, $\boldsymbol{u}$ random effects term, with $E(u_m) = 0$, $Var(u_m) = \sigma^2_{u_m}$(variance of marker effect), $Var(\mathbf{e}) = \sigma^2$ (residual variance), $Cov(\mathrm{u}, \mathbf{e}) = 0$, and the ridge parameter λ equals σ_e^2 / σ_u^2 (Meuwissen, Hayes and Goddard, 2001; Endelman, 2011; Howard, Carriquiry and Beavis, 2014). In this study RRBLUP with and without fixed effects were implemented using the *mixed.solve* function in *rrBLUP* R package (Endelman, 2011).

Reproducing Kernel Hilbert Space (RKHS)

Semi-parametric models are known to capture interactions among loci. The semi-parametric GS approach used in this study was implemented as Bayesian RKHS in *BLGR* package in R (Perez, 2014), and described as follows:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{u} + \boldsymbol{\varepsilon} \qquad [5]$$

where $\boldsymbol{y}$ is the vector of phenotype; $\mathbf{1}$ is a vector of 1's; μ is the mean; $\boldsymbol{u}$ is vector of random effects ~MVN ($\mathbf{0}$, $\mathbf{K_h}\sigma_u^2$); and $\boldsymbol{\varepsilon}$ is the random residual vector ~ MVN ($\mathbf{0}$, $\boldsymbol{I}\sigma_\varepsilon^2$). In Bayesian RKHS, the priors $p(\mu, \boldsymbol{u}, \boldsymbol{\varepsilon})$ are proportional to the product of density functions MVN ($\mathbf{0}$, $\mathbf{K_h}\sigma_u^2$) and MVN ($\mathbf{0}$, $\boldsymbol{I}\sigma_\varepsilon^2$). The kernel entries matrix ($\mathbf{K_h}$) with a Gaussian kernel uses the squared Euclidean distance between marker genotypes to estimate the degree of relatedness between individuals, and a smoothing parameter (*h*) multiplies each entry in $\mathbf{K_h}$ by a constant. In the implementation of RKHS a default smoothing parameter *h* of 0.5 was used alongside 1,000 burns and 2,500 iterations.

Support Vector Regression (SVR)

Support vector regression method (Vapnik, 1995; Maenhout *et al.*, 2007; Long *et al.*, 2011) was used to implement non-parametric GS approach in this study. The aim of the SVR method is to minimize prediction error by implementing models that minimizes large residuals (Long *et al.*, 2011). Thus, it is also referred to as the "ε-intensive" method. It was implemented in this study using the normal radial function kernel (*rbfdot*) in the *ksvm* function of *kernlab* R package (Karatzoglou *et al.*, 2004).

Parameters evaluated in GS and MAS

Additional parameters were estimated to further evaluate the performance of GS and MAS models. A regression model was fitted between observed phenotypic information and

GEBV of the validation set to obtain both intercept and slope for both GS and MAS in each cycle of prediction. The estimates of the intercept and slope of the regression of the observed phenotypic information on GEBVs are valuable since their deviations from expected values can provide insight into deficiencies in the GS and MAS models (Daetwyler *et al.*, 2013). The bias estimate (slope and intercept) signify how the range of values in measured and predicted traits differ from each other. In addition, the coincidence index between the observed and GEBVs for both GS and MAS models was evaluated. The coincidence index (Fernandes *et al.*, 2018) evaluates the proportion of individuals with highest trait values (20%) that overlapped between the measured phenotypes and predicted phenotypic trait values for the validation population.

Evaluation of the effect of marker density and training population size

The effect of marker density and training population size on GS performance were evaluated. GS was performed using 20% (6426 SNPs), 40% (12852), 60% (19278), and 80% (25704) of the total number of markers available in this study (32130). Each proportion of the aforementioned marker densities was randomly sampled without replacement and used for training GS models and predict in the validation set and repeated for 100 cycles. Furthermore, to evaluate the effect of training population size on prediction accuracy, four levels (20% (61 RILs), 40% (122 RILs), 60% (183 RILs), and 80% (244 RILs)) of total population size (305 RILs) were used to train GS model and validate only 20% (61 RILs) of the total population size (305 RILs). Subagging approach was used to subset the training and validation sets at a time and repeated for 100 cycles.

Results

Phenotypic and genotypic variation in cowpea

Results showed variation between number of days to 50% flowering under long-day photoperiod and short-day photoperiod. Days to flowering time were higher for RILs under long-day than short-day (Figure 1). Results showed positive correlations between environments for each trait (Table S1 and S2). Furthermore, genomic heritability were moderate for the traits ranging between 0.41 under long day photoperiod to 0.48 for flowering time under short-day photoperiod , 0.21 under restricted irrigation to 0.30 under full irrigation for maturity, and 0.39 under restricted irrigation to 0.47 under full irrigation for seed size (Table S1 and S2).

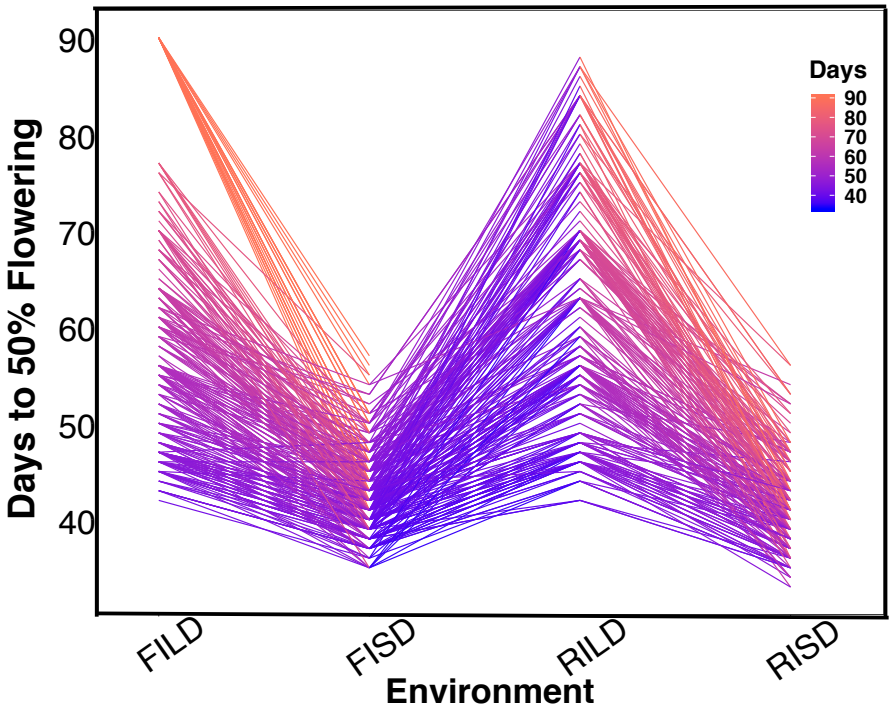


Figure 1:The norm of reaction plot for flowering time variation under long-day and short-day periods. Evaluation environments are represented on the x-axis (full irrigation and long day [FILD], full irrigation and short day [FISD], restricted irrigation and long day [RILD], and restricted irrigation and short day [RISD]). The number of days to 50% flowering is represented on the y-axis.

Genetic architecture of traits

Main effect QTL

The cowpea MAGIC population facilitated the characterization of the genetic architecture of flowering time, maturity and seed size. In this study QTL associated with flowering time, maturity, and seed size were identified using stepwise regression analysis (Table S3, Data S2). Results showed that 32 QTL (22 unique) in total were associated with flowering time traits (FT_BLUP [8 QTL, explaining 73.2 % of phenotypic variation (PV)], FTFILD [5 QTL, explaining 66.2% of PV], FTRILD [5 QTL explaining 48.6% of PV], FTFISD [8 QTL explaining 52.1% of PV], and FTRISD [6 QTL explaining 43.9% of PV]). Each of the total QTL associated with flowering time traits explained between 2% to 28% of the phenotypic variation. QTL qVu9:23.36, qVu9:24.77, and qVu9:22.65 (MAF= 0.29, 0.28, and 0.49) explained the largest proportion of variation (28%, 24%, and 19%) with additive effects of 7, 7, and 6 days respectively. The minor allele frequency (MAF) of the flowering time QTL ranges from 0.13 to 0.50. For maturity traits, 13 QTL (11 unique QTL) in total were identified with five QTL (explaining 35.9% of PV) for MAT_BLUP, 4 QTL (explaining 24.5% of PV) for MFISD, and 4 QTL (explaining 27.9% of PV) for MRISD. All maturity

traits QTL explained between 4.5 to 10% of phenotypic variation and MAF ranges from 0.15 to 0.49.

Furthermore, for seed size traits, 10 QTL (7 unique QTL) in total were identified with 3 QTL (explaining 39.3% of PV) for SS_BLUP, 3 QTL (explaining 41% of PV) for SSFISD, and 4 QTL (explaining 39.4% of PV) for SSRISD. QTL qVu8:74.21, qVu8:74.29, qVu8:76.81 associated with SSFISD, SS_BLUP, and SSRISD explained the largest PV (29%, 25%, and 20%). All seed size trait QTL explained between 3 to 29% of PV and with MAF range between 0.21 and 0.49. A pleiotropic QTL qVu8:74.21 (MAF=0.24) was associated with both MRISD and SSRISD (explained 5% and 29% of PV respectively).

Two-way epistatic interaction QTL

Currently there is limited knowledge about what role epistasis plays in phenotypic variation in cowpea. Our results identified epistatic QTL underlying flowering time, maturity, and seed size (Table S4, Data S3). For flowering time traits, there were 42 two-way epistatic interactions at 84 epistatic loci (only 65 loci were unique). Among these are; 20 epistatic loci for FLT_BLUP, 18 epistatic or FTFILD, 12 epistatic loci for FTRILD, 14 epistatic loci for FTFISD, and 20 epistatic loci for FTRISD. Some large effect loci were involved in epistatic interactions in flowering time; examples include, QTL qVu9:25.39 (MAF=0.28, FT_BLUP PVE=23.5%, FTFILD PVE=24.5%, FTRILD PVE=26%) and QTL qVu9:3.46 (MAF=0.35, FLT_BLUP PVE=13.5%, FTRILD PVE=14.1%). For maturity, there were 17 pairwise epistatic interactions across 34 loci (of which 30 were unique). Among the maturity QTL, qVu9:8.37 had the largest effect explaining ~9% of the phenotypic variation. One epistatic interaction overlapped with both FTRISD, MRISD, and MT_BLUP (qVu2:48.05+ qVu9:8.37, MAF=0.30 and 0.39 respectively). For seed size, there were 13 interactions at 26 loci (19 were unique). Only one QTL (qVu8:74.29, MAF=0.25) had interactions with multiple QTL. The largest effect epistatic QTL associated with the three seed size traits (SS_BLUP, SSFISD, and SSRISD) is qVu8:74.29 (MAF0.25). Some QTL were found to overlap among main effect QTL and epistatic effect QTL for flowering time (nine QTL), maturity (three QTL), and seed size (three QTL) (Figure S1).

Main effect and epistatic QTL colocalized with *a priori* genes

Gene functions can be conserved across species (Huang *et al.*, 2017). In this study, a set of a priori genes was compiled from both *A. thaliana* and *G. max*. Both main effect QTL and epistatic QTL colocalized with putative cowpea orthologs of *A. thaliana* and *G. max* flowering time and seed size genes (Figure 2 - 5, Figure S2 - S11, Data S4). A putative cowpea ortholog (Vigun09g050600) of *A. thaliana* circadian clock gene *phytochrome E*

(*PHYE*; AT4G18130) (Aukerman and Sakai, 2003) colocalized with FTFILD QTL (qVu9:22.65; PVE=19.5%; main effect QTL) at the same genetic position. Also, a putative cowpea ortholog (Vigun07g241700) of *A. thaliana* circadian clock gene *TIME FOR COFFEE (TIC*; AT3G22380) (Hall *et al.*, 2003) colocalized at the same genetic position with FTFISD QTL (qVu7:86.92; PVE=2.6%; main effect QTL). The cowpea flowering time gene (*VuFT*; Vigun06g014600; CowpeaMine *v*.06) colocalized with an epistatic QTL (qVu6:0.68; PVE=3.5%) associated with FLT_BLUP and FTRILD at the same genetic position. The cowpea ortholog (Vigun11g157600) of *A. thaliana* circadian clock gene *PHYTOCLOCK1* (*PCL1*; AT3G46640) (Hazen *et al.*, 2005) colocalized with an epistatic QTL (qVu11:50.94; PVE=8-10%) associated with both FTFILD and FTRILD at the same genetic position. A putative cowpea ortholog (Vigun11g148700) of *A. thaliana* photoperiod gene *TARGET OF EAT2* (*TOE2*; AT5G60120) (Mathieu *et al.*, 2009) was found at a proximity of 0.6cM from a QTL (qVu11:49.06; PVE=7-11%; main effect QTL) associated with FTFILD, FTRILD, and FLT_BLUP. Some of the *a priori* genes colocalized with some QTL that are both main effect and epistatic QTL. For instance, the cowpea ortholog (Vigun01g205500) of *G. max* flowering time gene *phytochrome A* (*PHYA*; Glyma19g41210) (Tardivel *et al.*, 2014) colocalized with a FTFILD QTL (qVu1:66.57; PVE=5.3%; both main effect and epistatic QTL) at the same genetic position (Data S4). Lastly, a putative cowpea ortholog (Vigun08g217000) of *A. thaliana histidine kinase2* gene (*AHK2*; AT5G35750) (Orozco-Arroyo *et al.*, 2015) was found at a proximity of about 1-2cM from three QTL (qVu8:74.29, qVu8:74.21, qVu8:76.81; PVE=25%, 29.3%, and 20% respectively; main effect and epistatic QTL) associated with seed size traits SS_BLUP, SSFISD, and SSRISD).

In addition, some a priori genes were associated with multiple traits. The putative cowpea ortholog (Vigun05g024400) of *A. thaliana* circadian clock gene *CONSTANS* (*CO*; AT5G15840) (Wenkel *et al.*, 2006) colocalized at the same genetic position with a QTL (qVu5:8.5; PVE=6-8%; both main effect and epistatic QTL) associated with flowering time and maturity traits (FLT_BLUP, FTFISD, FTRILD, FTRISD, MAT_BLUP, and MFISD); The putative cowpea ortholog (Vigun09g025800) of *A. thaliana* circadian clock gene *ZEITLUPE* (*ZTL*; AT5G57360) (Somers *et al.*, 2000) colocalized at the same genetic position with a QTL (qVu9:8.37; PVE=9-11%; both main effect and epistatic QTL) associated with flowering time and maturity traits (FTFISD, FTRISD, and MRISD).

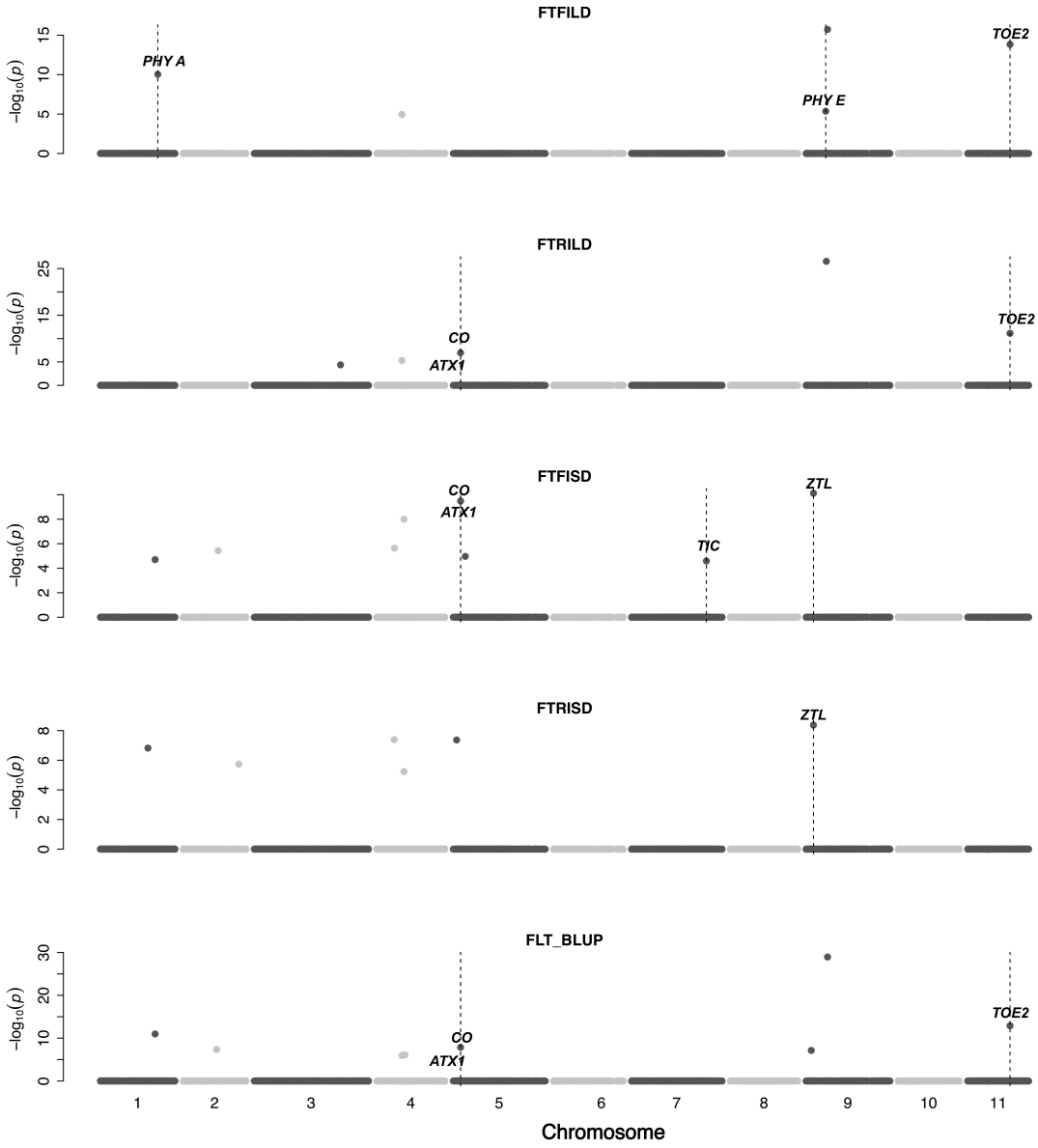
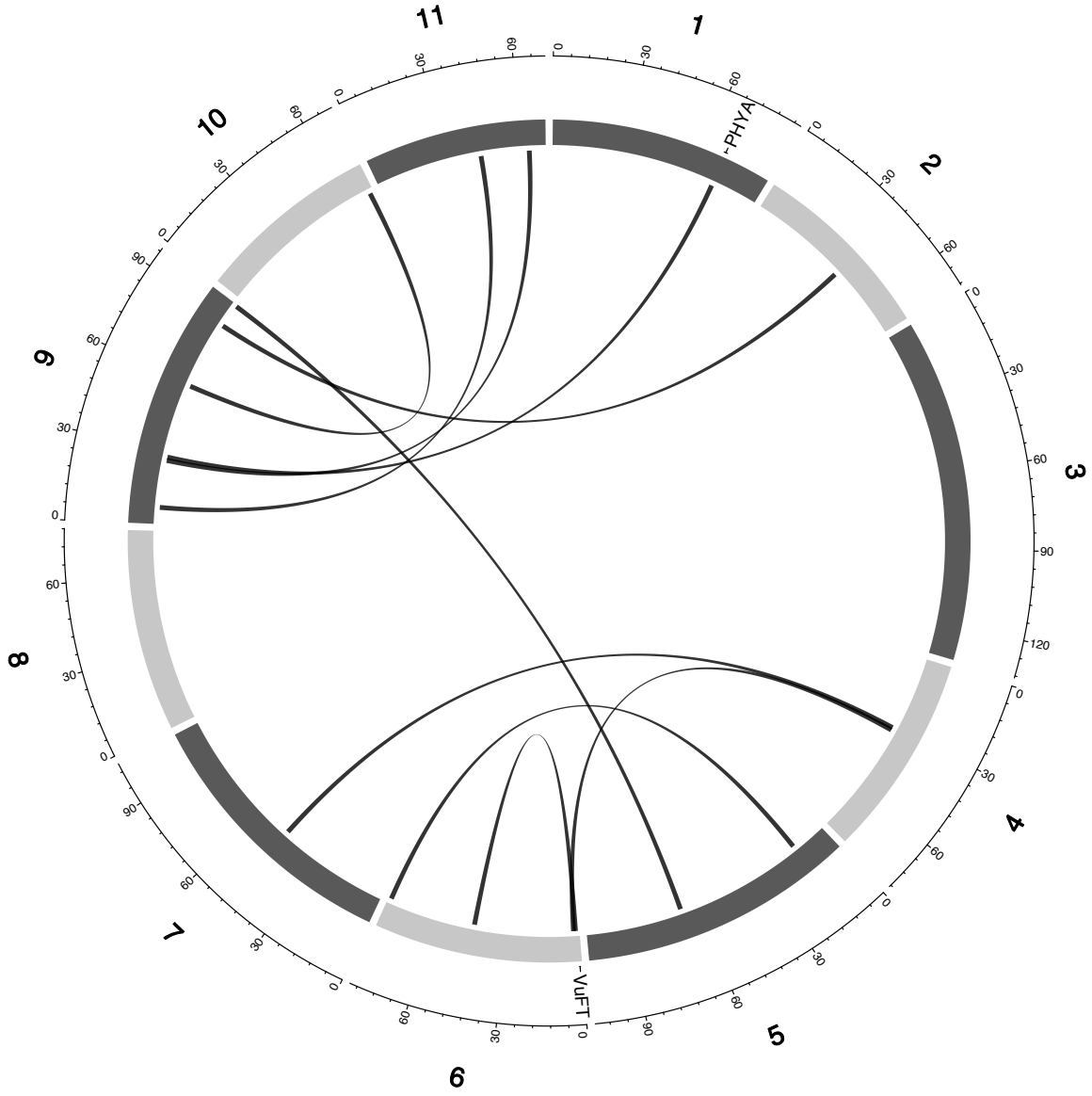


Figure 2: Main QTL plot for flowering time traits in the cowpea MAGIC population. QTL plots for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short day (FTRISD), and BLUPs of environments (FLT_BLUP). The chromosome numbers are located on the x-axis and the negative log of the *P*-values on the y-axis. The genetic position of the colocalization between QTL and *a priori* genes are indicated by broken vertical lines. The texts displayed on the vertical broken lines are the names of *a priori genes*.

397



398

399 **Figure 3: Epistatic QTL for FLT_BLUP for MAGIC population.** Chromosomes are shown in shades of
400 gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts
401 between chromosomes and genetic map.

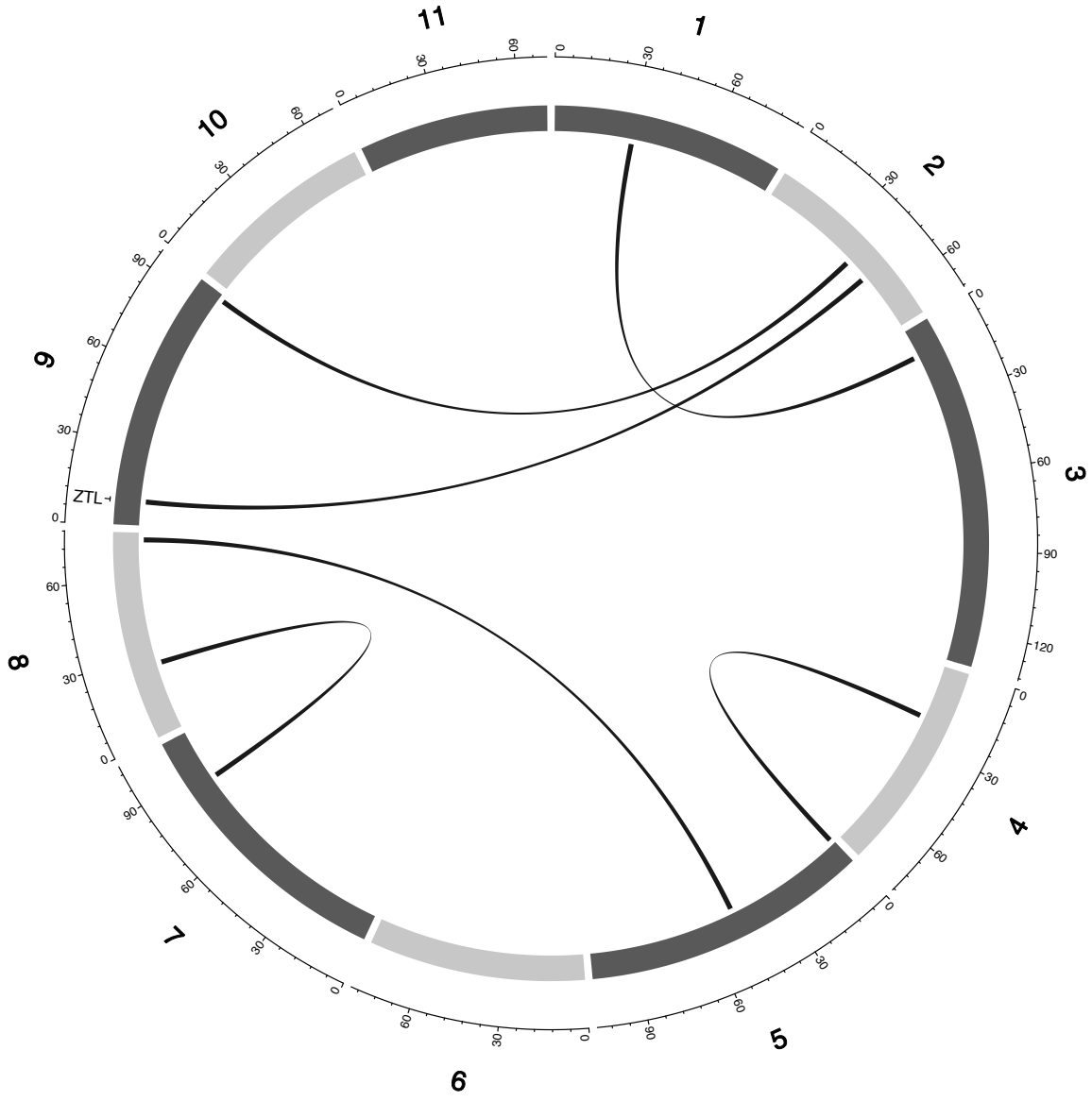


Figure 4: Epistatic QTL for MAT_BLUP in MAGIC population. Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

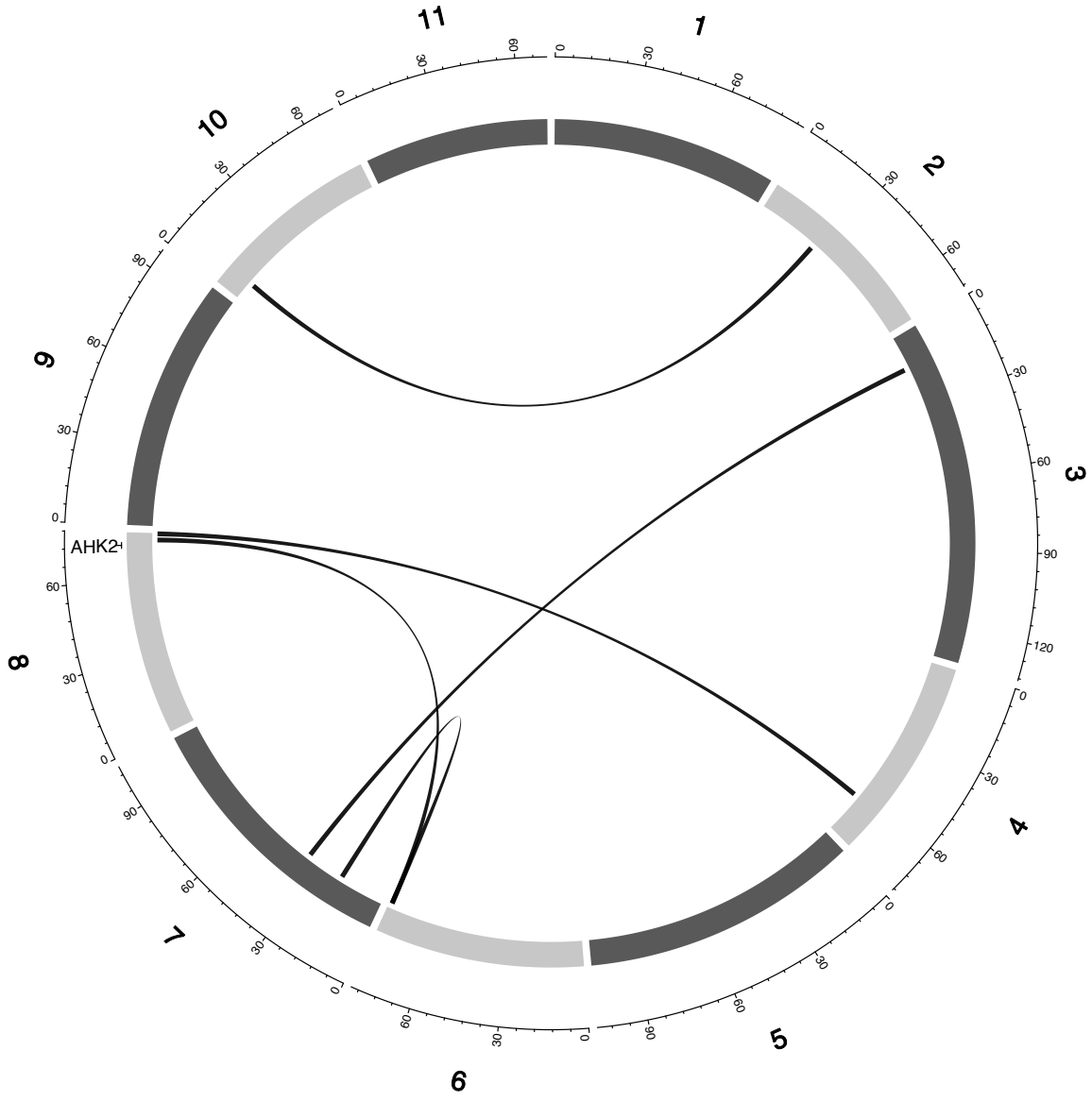


Figure 5: Epistatic QTL for MAT_BLUP in MAGIC population. Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

GS and MAS for flowering time

Prior knowledge about the genetic architecture of a trait can help make informed decisions in breeding. First to compare the performance of GS and MAS models for flowering time within each daylength results showed that under long day length (FTFILD and FTRILD); FxRRBLUP (mean prediction accuracy [mPA] = 0.68, 0.68; mean coincidence index [mCI]=0.49, 0.40) and MAS [mPA=0.64, 0.61; mCI=0.45, 0.37] outperformed RRBLUP [mPA=0.55, 0.58; mCI=0.37, 0.35], RKHS [mPA=0.55, 0.58; mCI=0.37, 0.36], and SVR [mPA=0.54, 0.50; mCI=0.35, 0.28] (Figures 6 and 7, Table S 3 and 4). For flowering time under long day, coincidence index values were higher under full irrigation

than under restricted irrigation. For flowering time under short day (FTFISD and FTRISD), all GS models outperformed MAS [mPA=0.33, 0.25; mCI=0.30, 0.26]. Among the GS models, RKHS and RRBLUP had the highest prediction accuracies. However, the coincidence index of FxRRBLUP was higher than the rest of the GS models for FTRISD. In general, the mean of the slope and intercept for the GS models except SVR were usually close to the expected (1 and 0) (Figure S12-S13). MAS also deviated away from the expected slope and intercept (1 and 0) more than the FxRRBLUP, RKHS, and RRBLUP for FTRISD (Figure S12-S13). Second to evaluate the effect of photoperiod and irrigation regime on the performance of training population, each environment (day length and irrigation regime combination) was used as a training population to predict the rest in a di-allele manner. Results showed that prediction accuracy between environments in the same photoperiod was higher than environments in different photoperiod (Figure S14). Also, when training populations were under full irrigation, their prediction accuracies were higher than when training populations were under restricted irrigation (Figure S14). For FT_BLUP, GS models outperformed MAS except SVR which had the same mPA [0.59] as MAS while FxRRBLUP had the highest mPA and mCIs among the GS models (Figure S 15 and 16).

GS and MAS for maturity and seed size

For maturity (MT_BLUP, MFISD, and MRISD), RKHS and RRBLUP had better performance (Figures 6 and 7; Table S4 and S5) than the rest of the models including MAS. All models deviated from the expected slope and intercept estimates, but RRBLUP had the least deviation for MRISD. For seed size, FxRRBLUP had the best performance followed by MAS compared to the rest of the GS models (RKHS, RRBLUP, and SVR) (Figures 6 and 7; Table S5 and S6). GS and MAS models had varying levels of deviation from the expected estimates of slope and intercept. RKHS and RRBLUP were closer to the expected than FxRRBLUP and MAS (Figure S12-S13). SVR had the highest deviation.

Effect of marker density and training population size

The effect of marker density and population size on GS in cowpea was investigated with the aim of making recommendations for resource limited national research centers in developing countries. For the effect of marker density on prediction accuracy, no significant relationship was observed between marker densities for MTBLUP while a significant increase in prediction accuracy was only observed between marker density 20% - 60% for FTBLUP, and between marker densities 40% - 60% and 40% - 80% for SSBLUP (Figure S19A). For the training population size effect, results revealed that prediction accuracy increased with increasing the size of the training set. All difference between training set sizes

were significantly increased with the training population size increase (Tukey test *P*-value < 0.001) (Figure S19B).

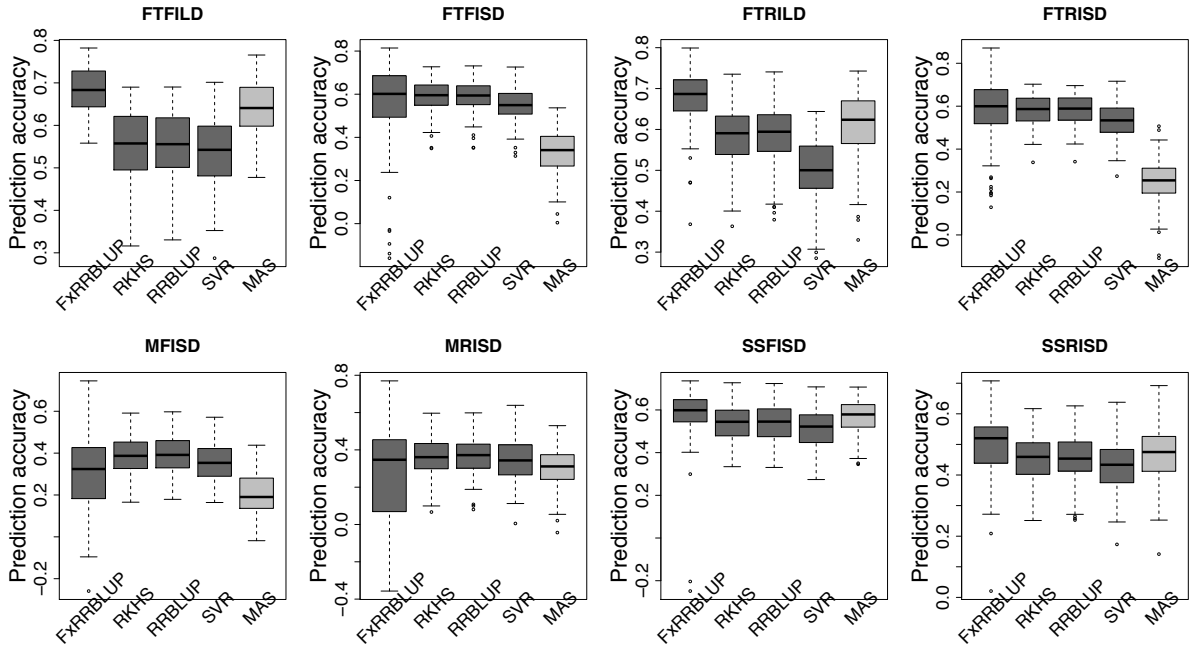


Figure 6: Comparison of prediction accuracy across GS and MAS models. Boxplots in each panel showed the distribution of prediction accuracy values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short day (FTRISD), maturity under full irrigation and short day (MFISD), maturity under restricted irrigation and short day, seed size under full irrigation and short day (SSFISD), and seed size under restricted irrigation and short day (SSRISD).

s

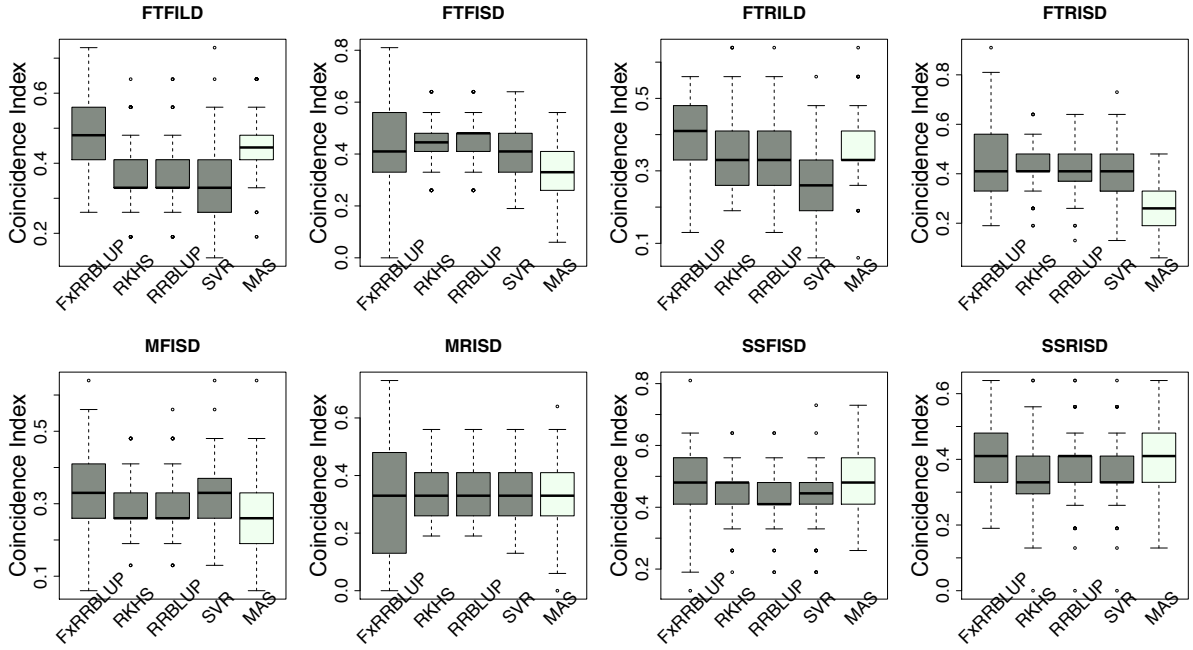


Figure 7: Comparison of coincidence index across GS and MAS models. Boxplots in each panel showed the distribution of coincidence index values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short day (FTRISD), maturity under full irrigation and short day (MFISD), maturity under restricted irrigation and short day, seed size under full irrigation and short day (SSFISD), and seed size under restricted irrigation and short day (SSRISD).

Discussion

Epistasis play important roles in determining the genetic architecture of agronomic traits in cowpea

Multi-parental populations have demonstrated ability to facilitate robust characterization of genetic architecture in terms of genetic effect size, pleiotropy, and epistasis (Buckler *et al.*, 2009; Brown *et al.*, 2011; Peiffer *et al.*, 2014; Bouchet *et al.*, 2017; Mathew *et al.*, 2018). Using the cowpea MAGIC population, this study showed that both additive main QTL and additive x additive epistatic QTL with large and (or) moderate effects underlie flowering time, maturity, and seed size in cowpea. Although most of the epistatic QTL identified were two-way interacting loci, results showed some of them were involved in interactions with more than one independent loci (Figure 3-5 and Figure S4-11). This implies the possibility of three-way epistatic interactions underlying some of the traits. Our inability to identify and discuss three-way epistatic interactions is due to the mapping approach used, which only mapped two-way epistatic interactions. Three-way epistatic interactions have been found to underlie flowering time in the selfing crop specie barley (Mathew *et al.*, 2018). Furthermore, overlaps between main and epistatic QTL (Figure S2) indicate these to be main QTL that are

494 involved in epistatic interactions with other loci. However, one caveat that may also be
495 responsible for some of the QTL among the overlaps is the false positive rate of SPEAML.
496 The SPEAML software used for epistasis mapping showed high false positive rate with a
497 sample size of 300 individuals (Chen *et al.*, 2018). It is possible that some of the overlapped
498 QTL are main QTL that were miscategorized as epistatic loci by SPEAML since our cowpea
499 MAGIC population had 305 RILs.

500 Flowering time is an important adaptive trait in breeding. In this study, our results
501 demonstrated that the flowering time variation in cowpea is due to large and moderate main
502 effects and epistatic loci (Table S3 and Table S4). Epistatic loci underlie flowering time in
503 both selfing (Huang *et al.*, 2013; Juenger *et al.*, 2005; Komeda, 2004;Mathew *et al.*, 2018)
504 (Chen *et al.*, 2018)(Li *et al.*, 2018) and outcrossing (Buckler *et al.*, 2009; Durand *et al.*, 2012)
505 species. In addition, the effect size of flowering time loci differs between selfing and out
506 crossing species as QTL effect sizes are large in the former (Lin, Schertz and Paterson, 1995;
507 Maurer *et al.*, 2015) and small in the later (Buckler *et al.*, 2009). In the present study, the
508 large effects (up to 25% PVE and additive effect of 7 days) flowering time loci were only
509 identified under long day photoperiod and not under short-day photoperiod (Table S3 and
510 Table S4). The loci detected under short day photoperiod were of moderate effects
511 (PVE=1%-10% and maximum additive effect size of 2 days). A trait's genetic architecture is
512 a reflection of its stability over evolutionary time and traits subjected to strong recent
513 selection were characterized with large effect loci (Brown *et al.*, 2011). Our result suggests
514 that cowpea flowering time adaptation to long-day photoperiod has undergone a recent
515 selection compared to flowering time under short-day photoperiod.

516 **Distinct and common genetic regulators underlie flowering time**

517 Conserved genetic pathways often underlie traits in plant species (Liu *et al.*, 2013;
518 Huang *et al.*, 2017). Examination of colocalizations between a priori genes and main effect
519 and epistatic QTL in this study identified putative cowpea orthologs of *A. thaliana* and *G.*
520 *max* flowering time and seed size genes that may be underlie phenotypic variation in cowpea.
521 Flowering time is affected by photoperiodicity and regulated by a network of genes (Sasaki,
522 Frommlet and Nordborg, 2017) involved in floral initiation, circadian clock regulation, and
523 photoreception (Lin, 2002). Photoperiod impacted days to flowering time as observed from
524 the norm of reaction plot for cowpea MAGIC flowering time data which showed drastic
525 reductions in days to flowering for RILs under short day compared to long days (Figure 1).
526 Our mapping results (main effect and epistatic) showed both unique and common loci
527 underlying flowering time under both long and short photoperiod (Figure 1; Figure S4-S8). In

addition, certain *a priori genes* were unique to either flowering time under long day or short day. For instance, cowpea putative orthologs of photoreceptors (*PHY A* [Vigun01g205500] and *PHY E* [Vigun09g050600]) and circadian clock gene *PHYTOCLOCK1* (*PCL1* [Vigun11g157600]) colocalized with only QTL associated with flowering time under long day, while cowpea putative orthologs of circadian clock genes (*Time for Coffee* [*TIC* (Vigun07g241700)] and *Zeitlupe* [*ZTL*]) colocalized with only QTL associated with flowering time under short day. However, the cowpea putative ortholog of photoperiod gene *CONSTANS* (*CO* [Vigun05g024400]) colocalized with QTL associated with flowering time under both long and short days. Thus, our study suggests that distinct and common genetic regulators control flowering time adaptation to both long and short-day photoperiod in cowpea. Further studies utilizing functional approaches will be helpful to decipher gene regulation patterns under both long and short photoperiod in cowpea.

Genetic architecture influenced GS and MAS performance

GS models differ in their efficiency to capture complex cryptic interactions among genetic markers (de Oliveira Couto *et al.*, 2017). The traits evaluated in this study are controlled by both main effect and epistatic loci. In this study, comparison among the GS models showed that parametric and semi-parametric GS models outperformed non-parametric GS model for all traits. SVR, a non-parametric model had the least prediction accuracy and coincidence index and also had the highest bias (Figure S12 and S13). Previous studies have shown that semi-parametric and non-parametric models increased prediction accuracy under epistatic genetic architecture (Howard, Carriquiry and Beavis, 2014; Jacquin, Cao and Ahmadi, 2016). In this study, none of semi-parametric and non-parametric models outperformed parametric models (Figure 6 and 7). Some of the studies comparing the performance of parametric, semi-parametric and non-parametric GS models were based on simulations of traits controlled solely by epistatic genetic architectures. Therefore, the performance of the models under simulated combined genetic effects (additive + epistasis) is not well understood. The comparable performance of RKHS to RRBLUP (parametric model) in this study in terms of prediction accuracy, coincidence index, and bias estimates, attests to RKHS ability to capture both additive and epistatic interactions (Gianola, Fernando and Stella, 2006; Gianola and Van Kaam, 2008; De Los Campos *et al.*, 2010; Gota and Gianola, 2014) for both prediction accuracy and selection of top performing lines. The performance of GS models' is often indistinguishable and RRBLUP has been recommended as an efficient parametric GS model (Heslot *et al.*, 2012; Lipka *et al.*, 2015). SVR had the worst performance with extremely high bias estimates.

Understanding the genetic architecture of agronomic traits can help improve predictions (Hayes *et al.*, 2010; Swami, 2010). Our study demonstrated that the effect size of QTL associated with a trait played a role in the performance of GS and MAS models. For instance, for traits controlled by both large and moderate effects loci (FTFILD, FTRILD, SSFISD, and SSRISD) parametric model with known loci as fixed effect (FxRRBLUP) followed by MAS outperformed the rest of the GS models (RRBLUP, RKHS, and SVR). The use of known QTL as fixed effects has been shown to increase prediction accuracy (Bernardo, 2014; Spindel *et al.*, 2016) in parametric GS models. For traits that were controlled by moderate effects loci (FTFISD, FTRISD, MFISD, and MTRISD), our results showed that the two parametric GS models (FxRRBLUP and RRBLUP) and semi-parametric (RKHS) had similar prediction accuracy, however FxRRBLUP had higher bias than RRBLUP and RKHS (Figure S12 - S13). Accuracy of prediction is influenced by genetic architecture (Hayes *et al.*, 2010). Furthermore, the performance of MAS in comparison to GS models in this study showed that large effects loci are important influencers of MAS (Bernardo, 2008). For small breeding programs in developing countries, MAS might be a prudent choice over GS for traits controlled only loci of large effects in cowpea since GS will require genotyping of more markers than MAS. The large effect QTL identified in this study can be transferred to different breeding populations because they were identified in a MAGIC population with wide genetic background (Chen *et al.*, 2018; Huynh *et al.*, 2018). Our study thus demonstrates that prior knowledge of the genetic architecture of a trait can help make informed decision about the best GEB method to employ in breeding.

Experimental design considerations for GS in cowpea

An important consideration in this study is to provide recommendations to breeders on resources needed for the implementation of GS in cowpea. First, this study demonstrated that genomic prediction within the same photoperiod is more efficient than across different photoperiod (Figure S14). Prediction between irrigation regimes had similar performance. The differences observed for GS between photoperiods showed that genotype x environment (GxE) interaction is an important factor to consider in cowpea flowering time GS. Increased genetic gains were observed in GS approaches that modeled GxE interactions (Lopez-Cruz *et al.*, 2015; Crossa *et al.*, 2016; de Oliveira Couto *et al.*, 2017). Second, our results showed that the size of the training population had an effect on prediction accuracy as prediction accuracy increased with increase in training population size. The size of a training population is an important factor influencing prediction accuracy (Liu *et al.*, 2018) and studies have shown increase in prediction accuracy with increase in training population size in several crop

596 species (Albrecht *et al.*, 2011; Spindel *et al.*, 2015). Third, increase in marker density only
597 significantly increased prediction between 20-60% for FLT_BLUP and 40-60% and 60-80%
598 for SS_BLUP (Figure S19). Though these differences were significant, the mean prediction
599 accuracy values were close to each other for all marker densities (Figure S19A). If using 20%
600 of markers (6424 SNPs) gave similar prediction accuracy as 32,130 SNPs; then it might be
601 more cost efficient for a breeder to use a small marker density. For instance, for flowering
602 time, 6424 SNPs gave a mean prediction accuracy of 0.665 and 32130 SNPs gave a
603 prediction accuracy of 0.671, then it might be logical and cost efficient to use ~6000 markers
604 for GS.

605 In summary, to the best of our knowledge, this is the first study that will characterize
606 epistasis and provide insights into the underpinnings of genomic selection *versus* marker
607 assisted selection in cowpea. Our study identified both main QTL and two-way epistatic loci
608 underlying flowering time, maturity, and seed size. We also found that flowering time is
609 under the control of both large and moderate effect loci similar to findings in other inbreeding
610 species. The large effect QTL and their colocalized *a priori* genes identified in this study will
611 serve as pedestal for future studies aimed at the molecular characterization of the genes
612 underlying flowering time and seed size in cowpea. We demonstrated that prior knowledge of
613 the genetic architecture of a trait can help make informed decision in GEB. Together, our
614 findings in this study are relevant for crop improvement in both developed and developing
615 countries.

616 **Acknowledgement**

617 We express our gratitude to Prof. Timothy Close, Prof. Philip Roberts, Dr. Bao-Lam
618 Huynh and their team at the University of California - Riverside, USA for their incredible
619 contributions to cowpea genomics and the privilege to use the cowpea MAGIC population
620 data for this study. The MAGIC population development, phenotyping, and genotyping was
621 supported in large part by grants from the Generation Challenge Program of the Consultative
622 Group on International Agricultural Research, with additional support from the USAID Feed
623 the Future Innovation Lab for Collaborative Research on Grain Legumes (Cooperative
624 Agreement EDH-A-00-07-00005), the USAID Feed the Future Innovation Lab for Climate
625 Resilient Cowpea (Cooperative Agreement AID-OAA-A-13-00070), and NSF-BREAD
626 (Advancing the Cowpea Genome for Food Security). We also thank Dr. Sandeep Marla, and
627 Fanna Maina for helping with the manuscript review.

628 **Authors' contributions**

629 M.O.O. obtained data from UCR; concept by M.O.O and Z.H; M.O.O. and Z.H. analyzed the
630 data; M.O.O, Z.H, and P.O.A wrote the manuscript. All authors read and approved the
631 manuscript.

632 **Supporting information**

633 All the R scripts used for analyses in the study are available at:
634 https://github.com/marcbios/Cowpea.git
635
636

References

Albrecht, T., Wimmer, V., Auinger, H.-J., Erbe, M., Knaak, C., Ouzunova, M., Simianer, H. and Schön, C.-C. (2011) 'Genome-based prediction of testcross values in maize', *Theoretical and Applied Genetics*, 123(2), pp. 339–350. doi: 10.1007/s00122-011-1587-7.

Aukerman, M. J. and Sakai, H. (2003) 'Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes.', *The Plant cell*. American Society of Plant Biologists, 15(11), pp. 2730–41. doi: 10.1105/tpc.016238.

Ben-Ari, G. and Lavi, U. (2012) 'Marker-assisted selection in plant breeding', in *Plant Biotechnology and Agriculture*. Elsevier, pp. 163–184. doi: 10.1016/B978-0-12-381466-1.00011-0.

Bernardo, R. (2008) 'Molecular markers and selection for complex traits in plants: Learning from the last 20 years', *Crop Science*. Crop Science Society of America, pp. 1649–1664. doi: 10.2135/cropsci2008.03.0131.

Bernardo, R. (2014) 'Genomewide Selection when Major Genes Are Known', *Crop Science*. The Crop Science Society of America, Inc., 54(1), p. 68. doi: 10.2135/cropsci2013.05.0315.

Bernardo, R. and Yu, J. (2007) 'Prospects for Genomewide Selection for Quantitative Traits in Maize', *Crop Science*. Crop Science Society of America, 47(3), p. 1082. doi: 10.2135/cropsci2006.11.0690.

Bouchet, S., Olatoye, M. O., Marla, S. R., Perumal, R., Tesso, T., Yu, J., Tuinstra, M. and Morris, G. P. (2017) 'Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population', *Genetics*, 206(2), pp. 573–585. doi: 10.1534/genetics.116.198499.

Boukar, O., Belko, N., Chamarthi, S., Togola, A., Batieno, J., Owusu, E., Haruna, M., Diallo, S., Umar, M. L., Olufajo, O. and Fatokun, C. (2018) 'Cowpea (Vigna unguiculata): Genetics, genomics and breeding', *Plant Breeding*. Edited by C. Ojiewo, 9 May. doi: 10.1111/pbr.12589.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. (2007) 'TASSEL: software for association mapping of complex traits in diverse samples.', *Bioinformatics (Oxford, England)*, 23(19), pp. 2633–5. doi: 10.1093/bioinformatics/btm308.

Brown, P. J., Upadyayula, N., Mahone, G. S., Tian, F., Bradbury, P. J., Myles, S., Holland, J. B., Flint-Garcia, S., McMullen, M. D., Buckler, E. S. and Rocheford, T. R. (2011) 'Distinct genetic architectures for male and female inflorescence traits of maize', *PLoS Genetics*. Public Library of Science, 7(11), p. e1002383. doi: 10.1371/journal.pgen.1002383.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., Goodman, M. M., Harjes, C., Guill, K., Kroon, D. E., Larsson, S., Lepak, N. K., Li, H., Mitchell, S. E., Pressoir, G., Peiffer, J. A., Rosas, M. O., Rocheford, T. R., Romay, M. C., Romero, S., Salvo, S., Villeda, H. S., Da Silva, H. S., Sun, Q., Tian, F., Upadyayula, N., Ware, D., Yates, H., Yu, J., Zhang, Z., Kresovich, S. and McMullen, M. D. (2009) 'The genetic architecture of maize flowering time', *Science*. American Association for the Advancement of Science, 325(5941), pp. 714–718. doi: 10.1126/science.1174276.

Cerrudo, D., Cao, S., Yuan, Y., Martinez, C., Suarez, E. A., Babu, R., Zhang, X. and Trachsel, S. (2018) 'Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments', *Frontiers in Plant Science*. Frontiers Media SA, 9, p. 366. doi: 10.3389/fpls.2018.00366.

Chen, A. H., Ge, W., Metcalf, W., Jakobsson, E., Mainzer, L. S. and Lipka, A. E. (2018) 'An assessment of true and false positive detection rates of stepwise epistatic model selection as a function of sample size and number of markers', *Heredity*, 15 November, p. 1. doi: 10.1038/s41437-018-0162-2.

Chen, J., Alqudah, A. M., Hirotsu, N., Swamy, B. P. M., Iris, G., Descalsota, L., Mallikarjuna Swamy, B. P., Zaw, H., Ann Inabangan-Asilo, M., Amparado, A., Mauleon, R., Chadha-

Mohanty, P., Arocena, E. C., Raghavan, C., Leung, H., Hernandez, J. E., Lalusin, A. B., Mendioro, M. S., Genaleen, M., Diaz, Q. and Reinke, R. (2018) 'Genome-Wide Association Mapping in a Rice MAGIC Plus Population Detects QTLs and Genes Useful for Biofortification'. doi: 10.3389/fpls.2018.01347.

Chen, J., Li, X., Cheng, C., Wang, Y., Qin, M., Zhu, H., Zeng, R., Fu, X., Liu, Z. and Zhang, G. (2015) 'Characterization of Epistatic Interaction of QTLs LH8 and EH3 Controlling Heading Date in Rice', *Scientific Reports*. Nature Publishing Group, 4(1), p. 4263. doi: 10.1038/srep04263.

Chen, M., Ahsan, A., Meng, X., Rahaman, M., Chen, H. and Monir, M. (2018) 'Identification epistasis loci underlying rice flowering time by controlling population stratification and polygenic effect', *DNA Research*. Edited by S. Isobe. doi: 10.1093/dnares/dsy043.

Covarrubias-pazaran, G. (2017) 'Quick start for the sommer package', pp. 1–12. Available at: https://cran.r-project.org/web/packages/sommer/vignettes/sommer.start.pdf (Accessed: 5 March 2018).

Crossa, J., De Los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J. and Pérez-Rodríguez, P. (2016) 'Extending the marker × Environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat', *Crop Science*, 56(5), pp. 2193–2209. doi: 10.2135/cropsci2015.04.0260.

Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G. and Hickey, J. M. (2013) 'Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking', *Genetics*, pp. 347–365. doi: 10.1534/genetics.112.147983.

Durand, E., Bouchet, S., Bertin, P., Ressayre, A., Jamin, P., Charcosset, A., Dillmann, C. and Tenaillon, M. I. (2012) 'Flowering Time in Maize: Linkage and Epistasis at a Major Effect Locus', *Genetics*, 190(4), pp. 1547–1562. doi: 10.1534/genetics.111.136903.

Endelman, J. B. (2011) 'Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP', *The Plant Genome Journal*. Crop Science Society of America, 4(3), p. 250. doi: 10.3835/plantgenome2011.08.0024.

Fernandes, S. B., Dias, K. O. G., Ferreira, D. F. and Brown, P. J. (2018) 'Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum', *Theoretical and Applied Genetics*. Springer Berlin Heidelberg, 131(3), pp. 747–755. doi: 10.1007/s00122-017-3033-y.

Foolad, M. R. and Panthee, D. R. (2012) 'Marker-Assisted Selection in Tomato Breeding', *Critical Reviews in Plant Sciences*. Taylor & Francis Group , 31(2), pp. 93–123. doi: 10.1080/07352689.2011.616057.

Gianola, D., Fernando, R. L. and Stella, A. (2006) 'Genomic-Assisted Prediction of Genetic Value with Semiparametric Procedures', *Genetics*. Genetics, 173(3), pp. 1761–1776. doi: 10.1534/genetics.105.049510.

Gianola, D. and Van Kaam, J. B. C. H. M. (2008) 'Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits', *Genetics*. Genetics, 178(4), pp. 2289–2303. doi: 10.1534/genetics.107.084285.

Gianola, D. and de los Campos, G. (2008) 'Inferring genetic values for quantitative traits non-parametrically', *Genetics Research*. Cambridge University Press, 90(6), pp. 525–540. doi: 10.1017/S0016672308009890.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D. S. (2012) 'Phytozome: a comparative platform for green plant genomics.', *Nucleic acids research*. Oxford University Press, 40(Database issue), pp. D1178-86. doi: 10.1093/nar/gkr944.

Gota, M. and Gianola, D. (2014) 'Kernel-based whole-genome prediction of complex traits: A review', *Frontiers in Genetics*, 5(OCT), pp. 1–13. doi: 10.3389/fgene.2014.00363.

H.Barton, N. and D.Keightley (2002) 'UNDERSTANDING QUANTITATIVE GENETIC VARIATION', *Nature Reviews Genetics*, 3, pp. 11–21. doi: 10.1038/nrg700.

Hall, A. E. (2004) 'Breeding for adaptation to drought and heat in cowpea', *European*

Journal of Agronomy, 21(4), pp. 447–454. doi: 10.1016/j.eja.2004.07.005.
Hall, A. E., Cisse, N., Thiaw, S., Elawad, H. O. A., Ehlers, J. D., Ismail, A. M., Fery, R. L., Roberts, P. A., Kitch, L. W., Murdock, L. L., Boukar, O., Phillips, R. D. and McWatters, K. H. (2003) 'Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP', *Field Crops Research*, 82(2–3), pp. 103–134. doi: 10.1016/S0378-4290(03)00033-9.
Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. and Goddard, M. E. (2010) 'Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits', *PLoS Genetics*. Edited by M. Georges. Public Library of Science, 6(9), p. e1001139. doi: 10.1371/journal.pgen.1001139.
Hazen, S. P., Schultz, T. F., Pruneda-Paz, J. L., Borevitz, J. O., Ecker, J. R. and Kay, S. A. (2005) 'LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(29), pp. 10387–92. doi: 10.1073/pnas.0503029102.
Heslot, N., Yang, H. P., Sorrells, M. E. and Jannink, J. L. (2012) 'Genomic selection in plant breeding: A comparison of models', *Crop Science*. The Crop Science Society of America, Inc., 52(1), pp. 146–160. doi: 10.2135/cropsci2011.06.0297.
Howard, R. (2016) 'Evaluation of Parametric and Nonparametric Statistical Methods in Genomic Prediction', *Graduate Theses and Dissertations*. Available at: https://lib.dr.iastate.edu/etd/15720 (Accessed: 10 April 2018).
Howard, R., Carriquiry, A. L. and Beavis, W. D. (2014) 'Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures', *G3: Genes|Genomes|Genetics*, 4(6), pp. 1027–1046. doi: 10.1534/g3.114.010298.
Huang, P., Jiang, H., Zhu, C., Barry, K., Jenkins, J., Sandor, L., Schmutz, J., Box, M. S., Kellogg, E. A. and Brutnell, T. P. (2017) 'Sparse panicle1 is required for inflorescence development in Setaria viridis and maize', *Nature Plants*, 3(5), p. 17054. doi: 10.1038/nplants.2017.54.
Huang, X., Ding, J., Effgen, S., Turck, F. and Koornneef, M. (2013) 'Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis', *New Phytologist*, 199, pp. 843–857. doi: 10.1111/nph.12306.
Huynh, B. L., Ehlers, J. D., Huang, B. E., Muñoz-Amatriaín, M., Lonardi, S., Santos, J. R. P., Ndeve, A., Batieno, B. J., Boukar, O., Cisse, N., Drabo, I., Fatokun, C., Kusi, F., Agyare, R. Y., Guo, Y. N., Herniter, I., Lo, S., Wanamaker, S. I., Xu, S., Close, T. J. and Roberts, P. A. (2018) 'A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (Vigna unguiculata L. Walp.)', *Plant Journal*. Wiley/Blackwell (10.1111), 93(6), pp. 1129–1142. doi: 10.1111/tpj.13827.
Jacquin, L., Cao, T.-V. and Ahmadi, N. (2016) 'A Unified and Comprehensible View of Parametric and Kernel Methods for Genomic Prediction with Application to Rice', *Frontiers in Genetics*. Frontiers, 7, p. 145. doi: 10.3389/fgene.2016.00145.
Jannink, J. L., Lorenz, A. J. and Iwata, H. (2010) 'Genomic selection in plant breeding: From theory to practice', *Briefings in Functional Genomics and Proteomics*. Oxford University Press, 9(2), pp. 166–177. doi: 10.1093/bfgp/elq001.
Jarquin, D., Specht, J. and Lorenz, A. (2016) 'Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical Data Creates Robust Models for Enhancing Selection of Accessions', *G3: Genes|Genomes|Genetics*. G3: Genes, Genomes, Genetics, 6(8), pp. 2329–2341. doi: 10.1534/g3.116.031443.
Jiang, Y. and Reif, J. C. (2015) *Modeling Epistasis in Genomic Selection*. Available at: www.genetics.org/lookup/suppl/ (Accessed: 27 July 2018).
Johnson, N. (2008) 'Sewall Wright and the Development of Shifting Balance Theory', *Nature Education*, 1(1), p. 52. doi: 10.1093/rfs/hhx028.
Juenger, T. E., Sen, S., Stowe, K. A. and Simms, E. L. (2005) 'Epistasis and genotype-

environment interaction for quantitative trait loci affecting flowering time in Arabidopsis thaliana', in *Genetics of Adaptation*. Berlin/Heidelberg: Springer-Verlag, pp. 87–105. doi: 10.1007/1-4020-3836-4_9.

Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A. (2004) 'kernlab - An S4 Package for Kernel Methods in R', *Journal of Statistical Software*, 11(9), pp. 1–20. doi: 10.18637/jss.v011.i09.

King, E. G. and Long, A. D. (2017) 'The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster.', *G3 (Bethesda, Md.)*. G3: Genes, Genomes, Genetics, 7(6), pp. 1643–1652. doi: 10.1534/g3.117.041426.

Komeda, Y. (2004) 'GENETIC REGULATION OF TIME TO FLOWER IN ARABIDOPSIS THALIANA', *Annu. Rev. Plant Biol*, 55, pp. 521–556. doi: 10.1146/annurev.arplant.55.031903.141644.

Kurek, A. (2018) *Phenotypic and genomic selection for multi-trait improvement in soybean line and variety development*. Iowa State University. Available at: https://lib.dr.iastate.edu/etd (Accessed: 4 February 2019).

Langyintuo, A. S., Lowenberg-DeBoer, J., Faye, M., Lambert, D., Ibro, G., Moussa, B., Kergna, A., Kushwaha, S., Musa, S. and Ntoukam, G. (2003) 'Cowpea supply and demand in West and Central Africa', *Field Crops Research*. Elsevier, 82(2–3), pp. 215–231. doi: 10.1016/S0378-4290(03)00039-X.

Li, X., Guo, T., Mu, Q., Li, X. and Yu, J. (2018) 'Genomic and environmental determinants and their interplay underlying phenotypic plasticity', *Proceedings of the National Academy of Sciences*, 115(26), pp. 6679–6684. doi: 10.1073/pnas.1718326115.

Lin, C. (2002) 'Photoreceptors and Regulation of Flowering Time', *Plant Physiology*, 123(1), pp. 39–50. doi: 10.1104/pp.123.1.39.

Lin, Y.-R., Schertz, K. F. and Paterson, A. H. (1995) 'Comparative Analysis of QTLs Affecting Plant Height and Maturity Across the Poaceae, in Reference to an Interspecific Sorghum Population', *Genetics*, 141, pp. 391–411. Available at: http://www.genetics.org/content/genetics/141/1/391.full.pdf (Accessed: 13 October 2017).

Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J. and Gore, M. A. (2015) 'From association to prediction: statistical methods for the dissection and selection of complex traits in plants', *Current Opinion in Plant Biology*, 24, pp. 110–118. doi: 10.1016/j.pbi.2015.02.010.

Liu, C., Teo, Z. W. N., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z. and Yu, H. (2013) 'A Conserved Genetic Pathway Determines Inflorescence Architecture in Arabidopsis and Rice', *Developmental Cell*, 24(6), pp. 612–622. doi: 10.1016/j.devcel.2013.02.013.

Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., Wang, S., Li, W.-X., Zou, C., Prasanna, B. M., Olsen, M. S., Huang, C. and Xu, Y. (2018) 'Factors affecting genomic selection revealed by empirical evidence in maize', *The Crop Journal*. Elsevier, 6(4), pp. 341–352. doi: 10.1016/J.CJ.2018.03.005.

Long, N., Gianola, D., Rosa, G. J. M. and Weigel, K. A. (2011) 'Application of support vector regression to genome-assisted prediction of quantitative traits', *Theoretical and Applied Genetics*, 123(7), pp. 1065–1074. doi: 10.1007/s00122-011-1648-y.

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., Singh, R. P., Autrique, E. and de los Campos, G. (2015) 'Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model.', *G3 (Bethesda, Md.)*, 5(4), pp. 569–82. doi: 10.1534/g3.114.016097.

De Los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A. and Crossa, J. (2010) 'Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods', *Genetics Research*. Cambridge University Press, 92(4), pp. 295–308. doi: 10.1017/S0016672310000285.

Mackay, T. F. C. (2001) 'The Genetic Architecture of Quantitative Traits', *Annual Review of Genetics*, 35(1), pp. 303–339. doi: 10.1146/annurev.genet.35.102401.090633.

Maenhout, S., De Baets, B., Haesaert, G. and Van Bockstaele, E. (2007) 'Support vector machine regression for the prediction of maize hybrid performance', *Theoretical and Applied Genetics*, 115(7), pp. 1003–1013. doi: 10.1007/s00122-007-0627-9.

Massman, J. M., Jung, H.-J. G. and Bernardo, R. (2013) 'Genomewide Selection versus Marker-assisted Recurrent Selection to Improve Grain Yield and Stover-quality Traits for Cellulosic Ethanol in Maize', *Crop Science*. The Crop Science Society of America, Inc., 53(1), p. 58. doi: 10.2135/cropsci2012.02.0112.

Matei, G., Woyann, L. G., Milioli, A. S., de Bem Oliveira, I., Zdziarski, A. D., Zanella, R., Coelho, A. S. G., Finatto, T. and Benin, G. (2018) 'Genomic selection in soybean: accuracy and time gain in relation to phenotypic selection', *Molecular Breeding*. Springer Netherlands, 38(9), p. 117. doi: 10.1007/s11032-018-0872-4.

Mathew, B., Léon, J., Sannemann, W. and Sillanpää, M. J. (2018) 'Detection of epistasis for flowering time using bayesian multilocus estimation in a barley MAGIC population', *Genetics*. Genetics, 208(2), pp. 525–536. doi: 10.1534/genetics.117.300546.

Mathieu, J., Yant, L. J., Mürdter, F., Küttner, F. and Schmid, M. (2009) 'Repression of Flowering by the miR172 Target SMZ', *PLoS Biology*. Edited by C. Dean. Public Library of Science, 7(7), p. e1000148. doi: 10.1371/journal.pbio.1000148.

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., Kilian, B., Reif, J. C. and Pillen, K. (2015) 'Modelling the genetic architecture of flowering time control in barley through nested association mapping.', *BMC genomics*. ???, 16(1), p. 290. doi: 10.1186/s12864-015-1459-7.

Melchinger, A. E., Utz, H. F., Piepho, H. P., Zeng, Z. B. and Schön, C. C. (2007) 'The role of epistasis in the manifestation of heterosis: A systems-oriented approach', *Genetics*. Genetics Society of America, 177(3), pp. 1815–1825. doi: 10.1534/genetics.107.077537.

Messina, C. D., Podlich, D., Dong, Z., Samples, M. and Cooper, M. (2011) 'Yield-trait performance landscapes: From theory to application in breeding maize for drought tolerance', *Journal of Experimental Botany*. Oxford University Press, pp. 855–868. doi: 10.1093/jxb/erq329.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001) 'Prediction of total genetic value using genome-wide dense marker maps', *Genetics*, 157(4), pp. 1819–1829. doi: 11290733.

Mohamed, A., Ali, R., Elhassan, O., Suliman, E., Mugoya, C., Masiga, C. W., Elhusien, A. and Hash, C. T. (2014) 'First products of DNA marker-assisted selection in sorghum released for cultivation by farmers in sub-saharan Africa', *Journal of Plant Science and Molecular Breeding*. Herbert Publications, 3(1), p. 3. doi: 10.7243/2050-2389-3-3.

Moore, J. H. and Williams, S. M. (2009) 'Epistasis and its implications for personal genetics.', *American journal of human genetics*. Elsevier, 85(3), pp. 309–20. doi: 10.1016/j.ajhg.2009.08.006.

Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., Hearne, S., Cisse, N., Fatokun, C., Ehlers, J. D., Roberts, P. A. and Close, T. J. (2009) 'A consensus genetic map of cowpea [Vigna unguiculata (L) Walp.] and synteny based on EST-derived SNPs.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(43), pp. 18159–64. doi: 10.1073/pnas.0905886106.

Okogbenin, E., Porto, M. C. M., Egesi, C., Mba, C., Espinosa, E., Santos, L. G., Ospina, C., Mar??n, J., Barrera, E., Guti??rrez, J., Ekanayake, I., Iglesias, C. and Fregene, M. A. (2007) 'Marker-assisted introgression of resistance to cassava mosaic disease into latin American germplasm for the genetic improvement of cassava in Africa', *Crop Science*, 47(5), pp. 1895–1904. doi: 10.2135/cropsci2006.10.0688.

de Oliveira Couto, E. G., Bandeira e Sousa, M., Jarquín, D., Burgueño, J., Crossa, J., Fritsche-Neto, R., Pérez-Rodríguez, P. and Cuevas, J. (2017) 'Genomic-Enabled Prediction in Maize Using Kernel Models with Genotype × Environment Interaction', *G3: Genes|Genomes|Genetics*, 7(6), pp. 1995–2014. doi: 10.1534/g3.117.042341.

Orozco-Arroyo, G., Paolo, D., Ezquer, I. and Colombo, L. (2015) 'Networks controlling seed size in Arabidopsis', *Plant Reproduction*. Springer Berlin Heidelberg, 28(1), pp. 17–32. doi: 10.1007/s00497-015-0255-5.

Pan, Y. B., Tew, T. L., Schnell, R. J., Viator, R. P., Richard, E. P., Grisham, M. P. and White, W. H. (2006) 'Microsatellite DNA marker-assisted selection of Saccharum spontaneum cytoplasm-derived germplasm', *Sugar Tech*. Springer India, 8(1), pp. 23–29. doi: 10.1007/BF02943737.

Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., Gardner, C. A. C., McMullen, M. D., Holland, J. B., Bradbury, P. J. and Buckler, E. S. (2014) 'The genetic architecture of maize height', *Genetics*, 196(4), pp. 1337–1356. doi: 10.1534/genetics.113.159152.

Perez, P. (2014) 'BGLR : A Statistical Package for Whole Genome Regression and Prediction', *Genetics*, 198(2), pp. 483–495. doi: 10.1534/genetics.114.164442.

Rieseberg, L. H., Archer, M. A. and Wayne, R. K. (1999) 'Transgressive segregation, adaptation and speciation', *Heredity*. Nature Publishing Group, 83(4), pp. 363–372. doi: 10.1038/sj.hdy.6886170.

Saghai Maroof, M. A., Jeong, S. C., Gunduz, I., Tucker, D. M., Buss, G. R. and Tolin, S. A. (2008) 'Pyramiding of Soybean Mosaic Virus Resistance Genes by Marker-Assisted Selection', *Crop Science*. Crop Science Society of America, 48(2), p. 517. doi: 10.2135/cropsci2007.08.0479.

Sasaki, E., Frommlet, F. and Nordborg, M. (2017) 'The genetic architecture of the network underlying flowering time variation in Arabidopsis thaliana'. doi: 10.1534/genetics.XXX.XXXXXX.

Schneider, K. A., Brothers, M. E. and Kelly, J. D. (1997) 'Marker-Assisted Selection to Improve Drought Resistance in Common Bean', *Crop Science*. Crop Science Society of America, 37(1), p. 51. doi: 10.2135/cropsci1997.0011183X003700010008x.

Somers, D. E., Schultz, T. F., Milnamow, M. and Kay, S. A. (2000) 'ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis.', *Cell*. Elsevier, 101(3), pp. 319–29. doi: 10.1016/S0092-8674(00)80841-7.

Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J.-L. and Mccouch, S. (2016) 'Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement', *Heredity*, 116, pp. 395–408. doi: 10.1038/hdy.2015.113.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J. L. and McCouch, S. R. (2015) 'Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines', *PLoS Genetics*. Edited by R. Mauricio. Public Library of Science, 11(2), pp. 1–25. doi: 10.1371/journal.pgen.1004982.

Sun, X., Ma, P. and Mumm, R. H. (2012) 'Nonparametric Method for Genomics-Based Prediction of Performance of Quantitative Traits Involving Epistasis in Plant Breeding', *PLoS ONE*. Edited by X. Wang. Public Library of Science, 7(11), p. e50604. doi: 10.1371/journal.pone.0050604.

Swami, M. (2010) 'Using genetic architecture to improve predictions', *Nature Reviews Genetics*, 11(11), pp. 748–748. doi: 10.1038/nrg2888.

Tardivel, A., Sonah, H., Belzile, F. and O'Donoughue, L. S. (2014) 'Rapid Identification of Alleles at the Soybean Maturity Gene E3 using genotyping by Sequencing and a Haplotype-Based Approach', *The Plant Genome*, 7(2), p. 0. doi: 10.3835/plantgenome2013.10.0034.

Utz, H. F., Melchinger, A. E. and Schön, C. C. (2000) 'Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples', *Genetics*, 154(4), pp. 1839–1849. doi: 10.2307/1403680.

Vapnik, V. N. (1995) *The Nature of Statistical Learning Theory*. 1st edn. Edited by V. N.

Michael Jordan, Steffen L. Lauritzen, Jerald F. Lawless. New York: Springer. doi: 10.1007/978-1-4757-3264-1.

Varshney, R. K., Ribaut, J. M., Buckler, E. S., Tuberosa, R., Rafalski, A. J. and Langridge, P. (2012) 'Can genomics boost productivity of orphan crops?', *Nature Biotechnology*, 30(12), pp. 1172–1176. doi: 10.1038/nbt.2440.

Volis, S., Shulgina, I., Zaretsky, M. and Koren, O. (2010) 'Epistasis in natural populations of a predominantly selfing plant', *Heredity*, 106, pp. 300–309. doi: 10.1038/hdy.2010.79.

Wen, L., Chang, H.-X., Brown, P. J., Domier, L. L. and Hartman, G. L. (2019) 'Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean Rhg1 locus', *Horticulture Research*. Nature Publishing Group, 6(1), p. 9. doi: 10.1038/s41438-018-0085-3.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Gourrierec, J. Le, Samach, A. and Coupland, G. (2006) 'The Plant Cell Online', *Plant Cell*. American Society of Plant Biologists, 12(12), pp. 2473–2484. doi: 10.1105/tpc.12.12.2473.

Wong, C. K. and Bernardo, R. (2008) 'Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations', *Theoretical and Applied Genetics*. Springer-Verlag, 116(6), pp. 815–824. doi: 10.1007/s00122-008-0715-5.

Xu, S. (2003) 'Theoretical Basis of the Beavis Effect', *Genetics*, 165(4), pp. 2259–2268. Available at: http://www.genetics.org/content/genetics/165/4/2259.full.pdf (Accessed: 15 November 2017).

Xu, Y., Li, P., Yang, Z. and Xu, C. (2017) 'Genetic mapping of quantitative trait loci in crops', *The Crop Journal*. Elsevier, 5(2), pp. 175–184. doi: 10.1016/J.CJ.2016.06.003.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S. and Buckler, E. S. (2006) 'A unified mixed-model method for association mapping that accounts for multiple levels of relatedness', *Nature Genetics*. Nature Publishing Group, 38(2), pp. 203–208. doi: 10.1038/ng1702.

Yu, K., Park, S. J. and Poysa, V. (2000) 'Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics', *Plant Breeding*. John Wiley & Sons, Ltd (10.1111), 119(5), pp. 411–415. doi: 10.1046/j.1439-0523.2000.00514.x.

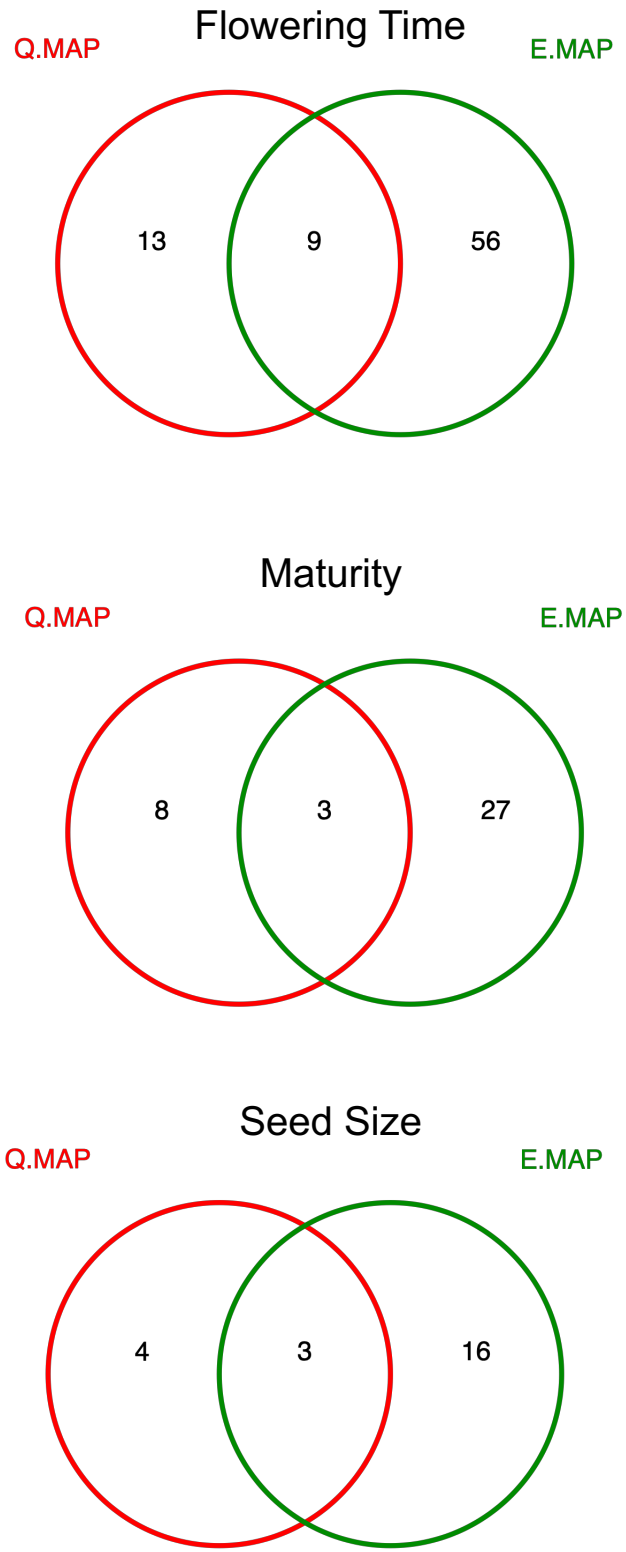
Zhang, T., Yuan, Y., Yu, J., Guo, W. and Kohel, R. J. (2003) 'Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection', *Theoretical and Applied Genetics*. Springer-Verlag, 106(2), pp. 262–268. doi: 10.1007/s00122-002-1101-3.

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H. and Reif, J. C. (2014) 'Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat', *Heredity*, 112(6), pp. 638–645. doi: 10.1038/hdy.2014.1.

986 **Tables**

987 **Supplementary figures**

988



989
990 **Figure S 1: Venn diagram of QTL overlap between main effect QTL mapping (Q.MAP) and epistasis**
991 **mapping (E.MAP) for flowering time, maturity, and seed size.**

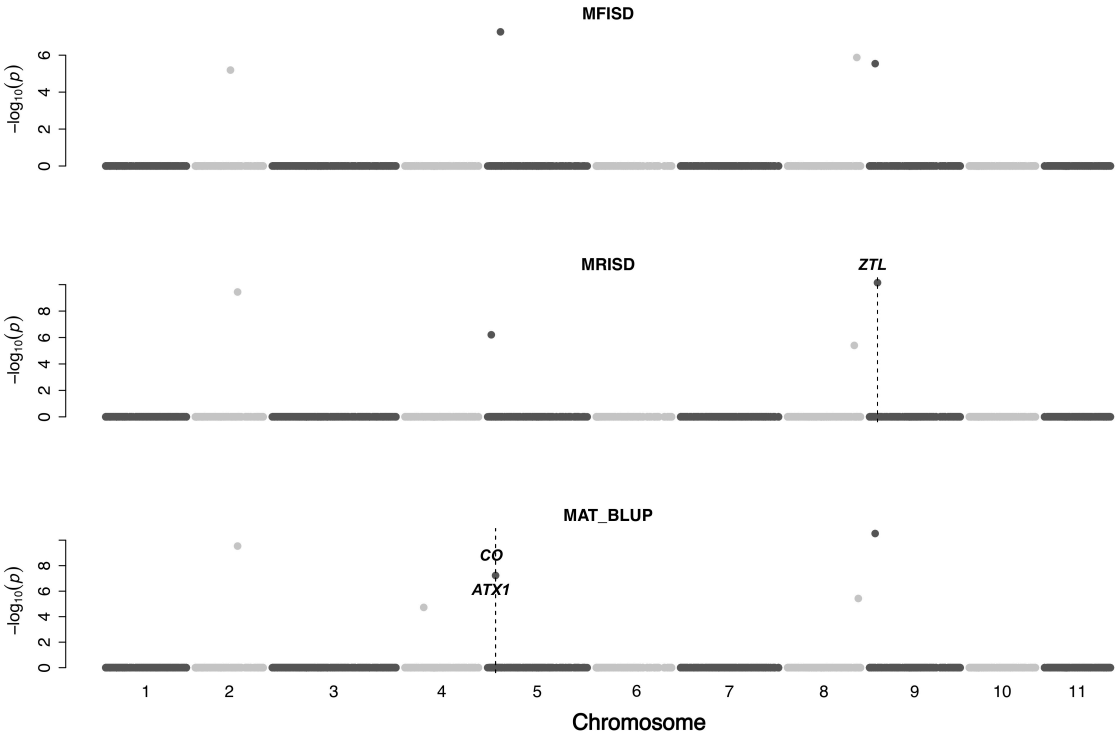


Figure S 2: QTL plot for maturity traits in the cowpea MAGIC population. QTL plots for maturity under full irrigation and short day (MFISD), maturity under restricted irrigation and short day (MRISD), and BLUPs of environments (MAT_BLUP). The chromosome numbers are located on the x-axis and the negative log of the *P*-values on the y-axis. The genetic position of the colocalization between QTL and *a priori* genes are indicated by broken vertical lines. The texts displayed on the vertical broken lines are the names of *a priori genes*.

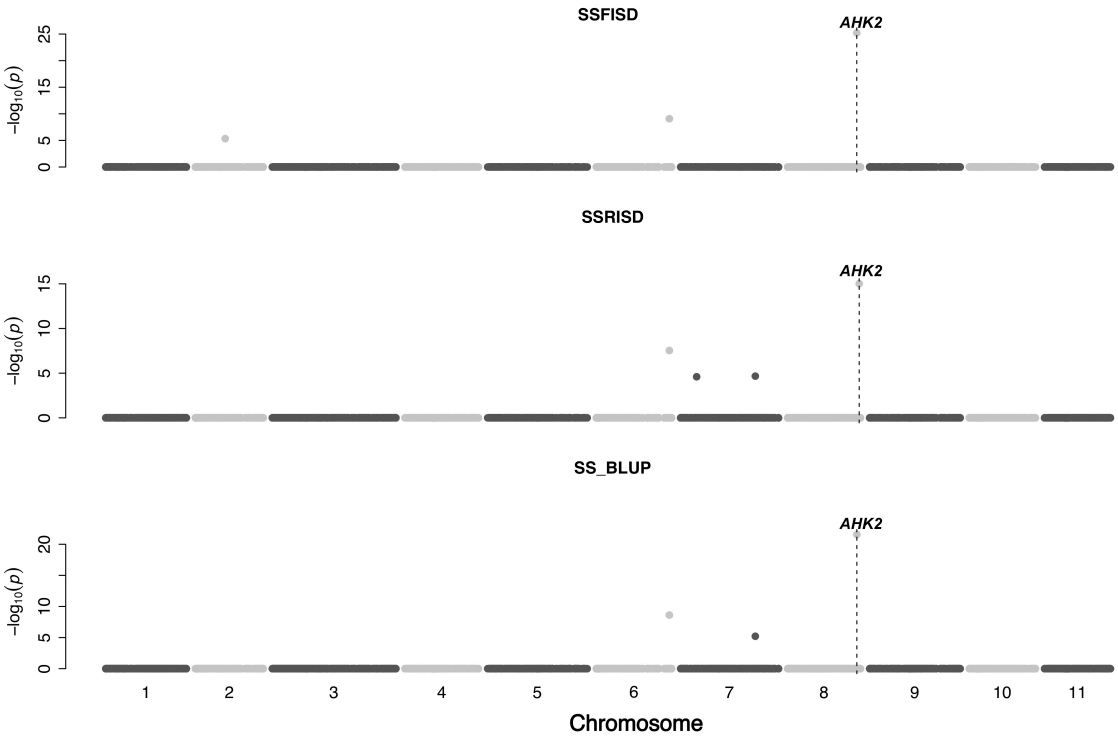


Figure S 3: QTL plot for seed size traits in the cowpea MAGIC population.QTL plots for maturity under full irrigation and short day (SSFISD), maturity under restricted irrigation and short day (SSRISD), and BLUPs of environments (SS_BLUP). The chromosome numbers are located on the x-axis and the

negative log of the *P*-values on the y-axis. The genetic position of the colocalization between QTL and *a priori* genes are indicated by broken vertical lines. The texts displayed on the vertical broken lines are the names of *a priori genes*.

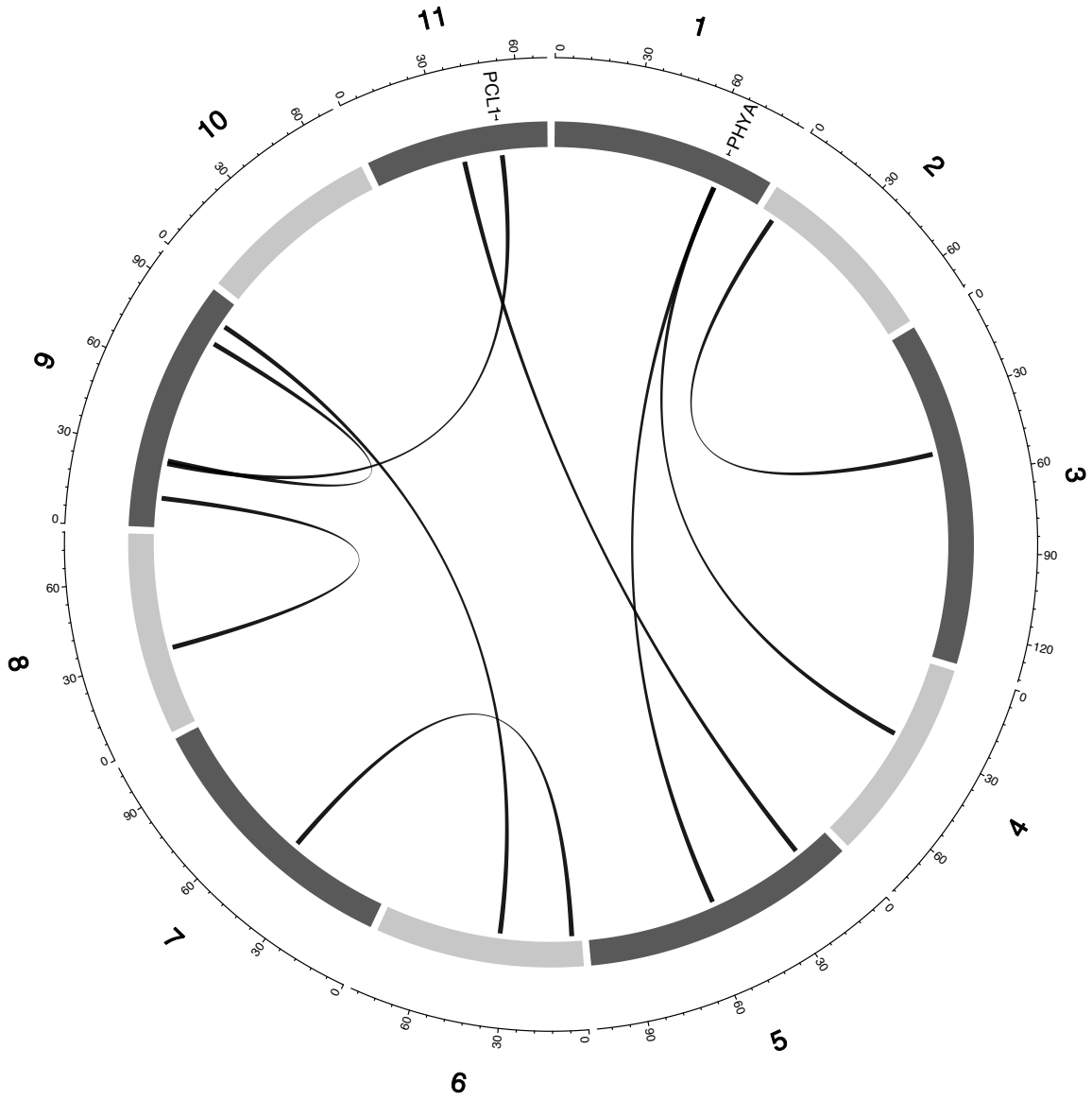


Figure S 4: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for FTFILD (Flowering time under full irrigation and long day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized *a priori* genes are texts between chromosomes and genetic map.

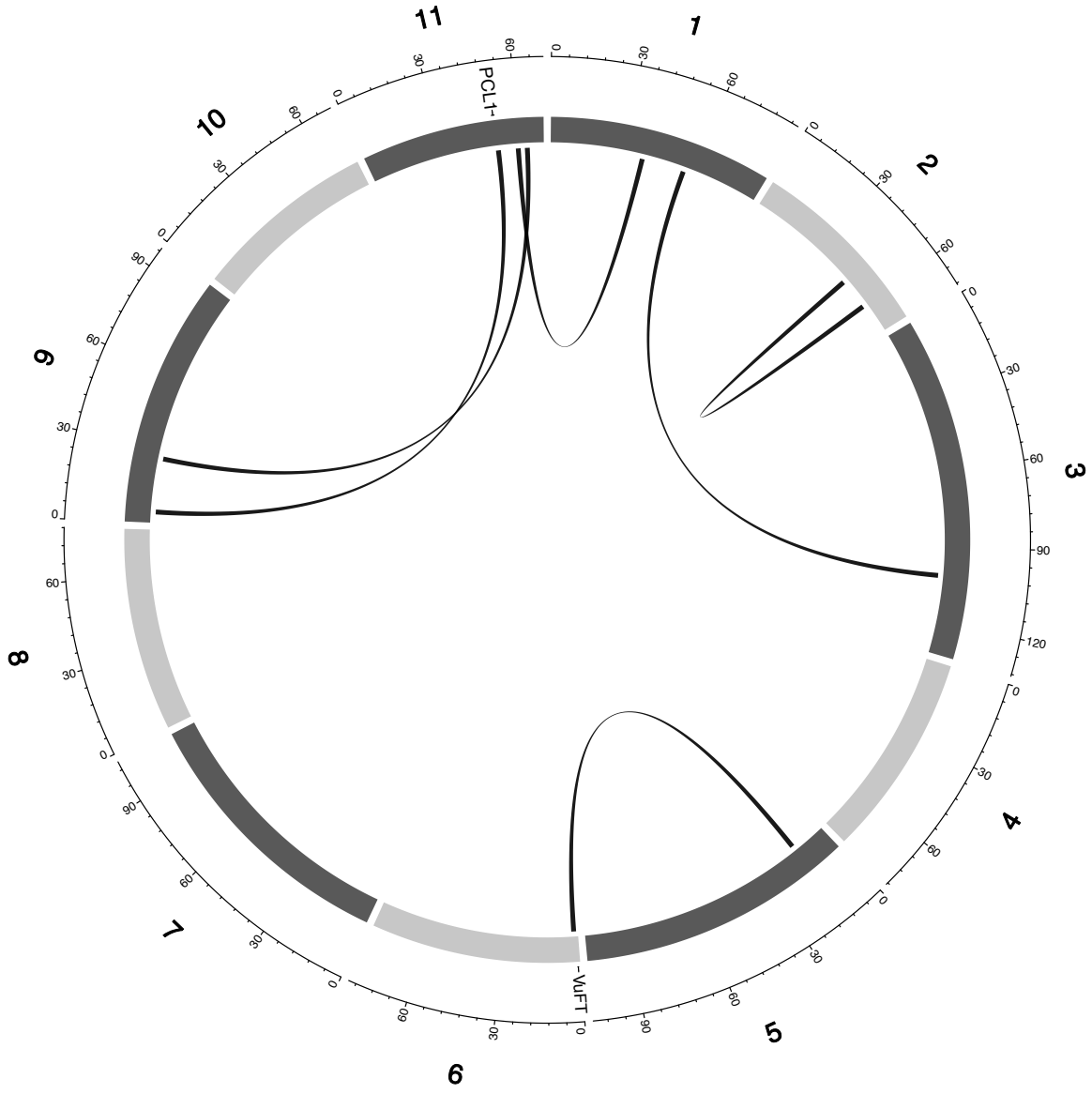


Figure S 5: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for FTRILD (Flowering time under restricted irrigation and long day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

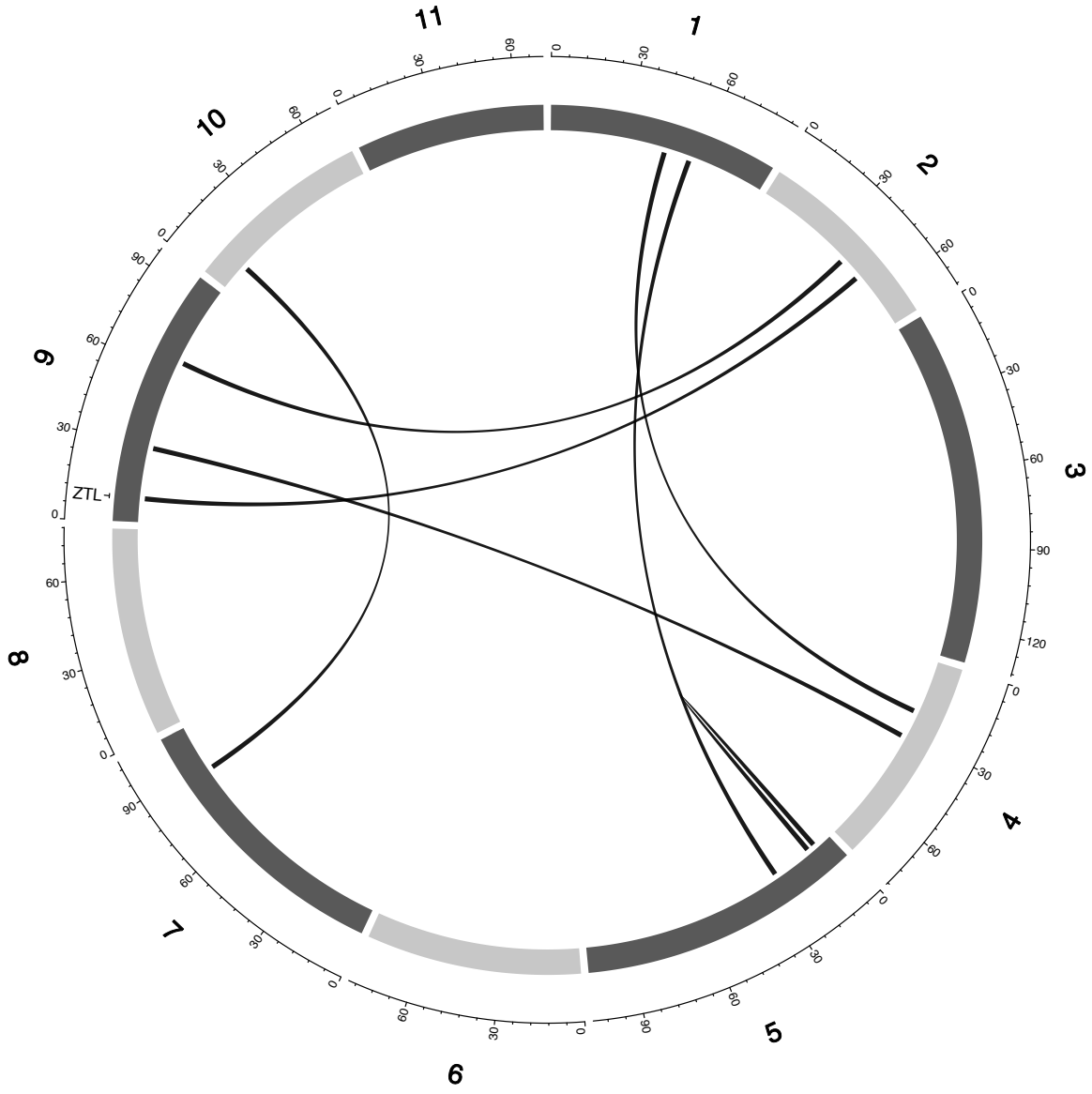


Figure S 6: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for FTFISD (Flowering time under full irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

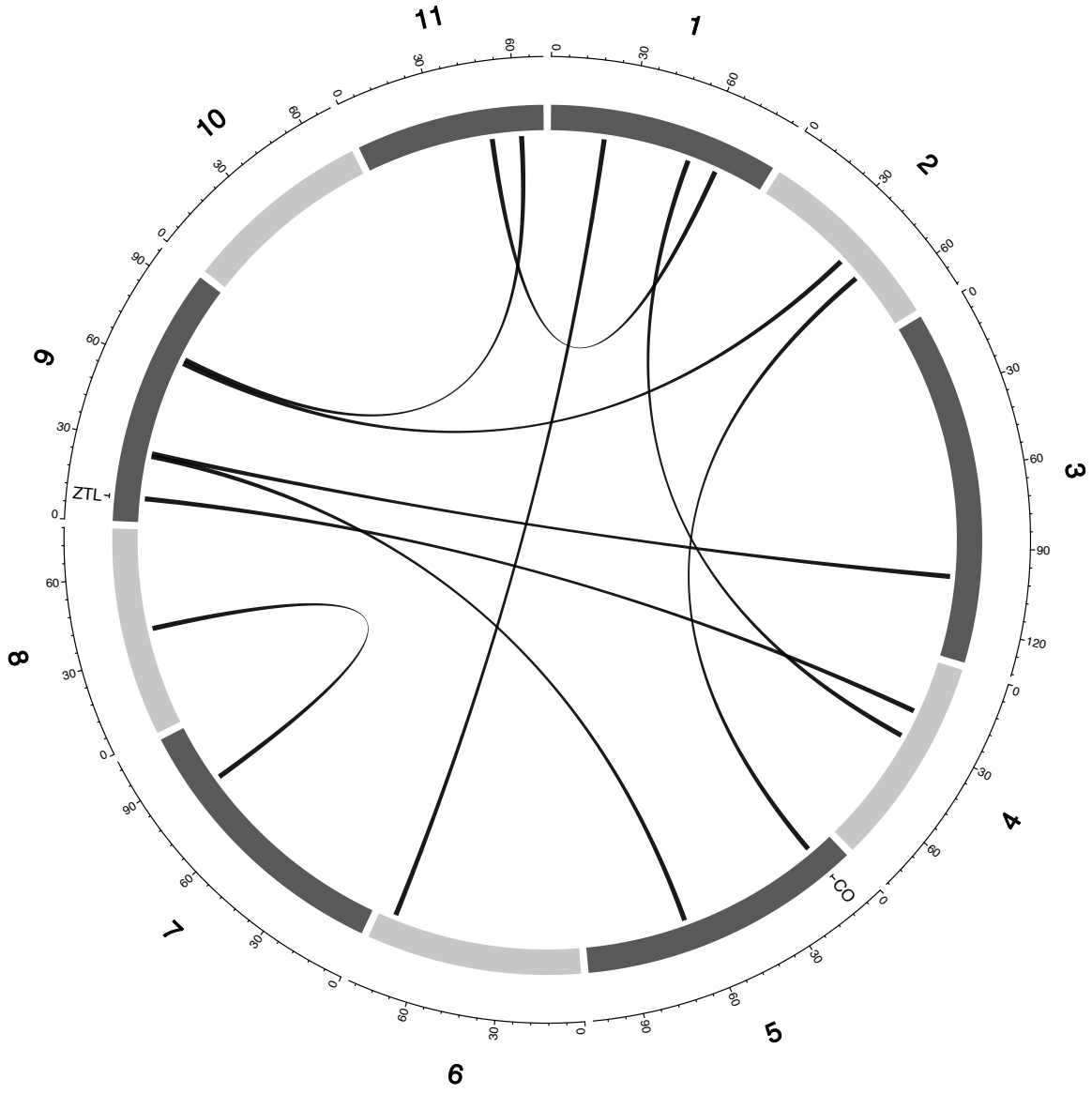


Figure S 7: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for FTRISD (Flowering time under restricted irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

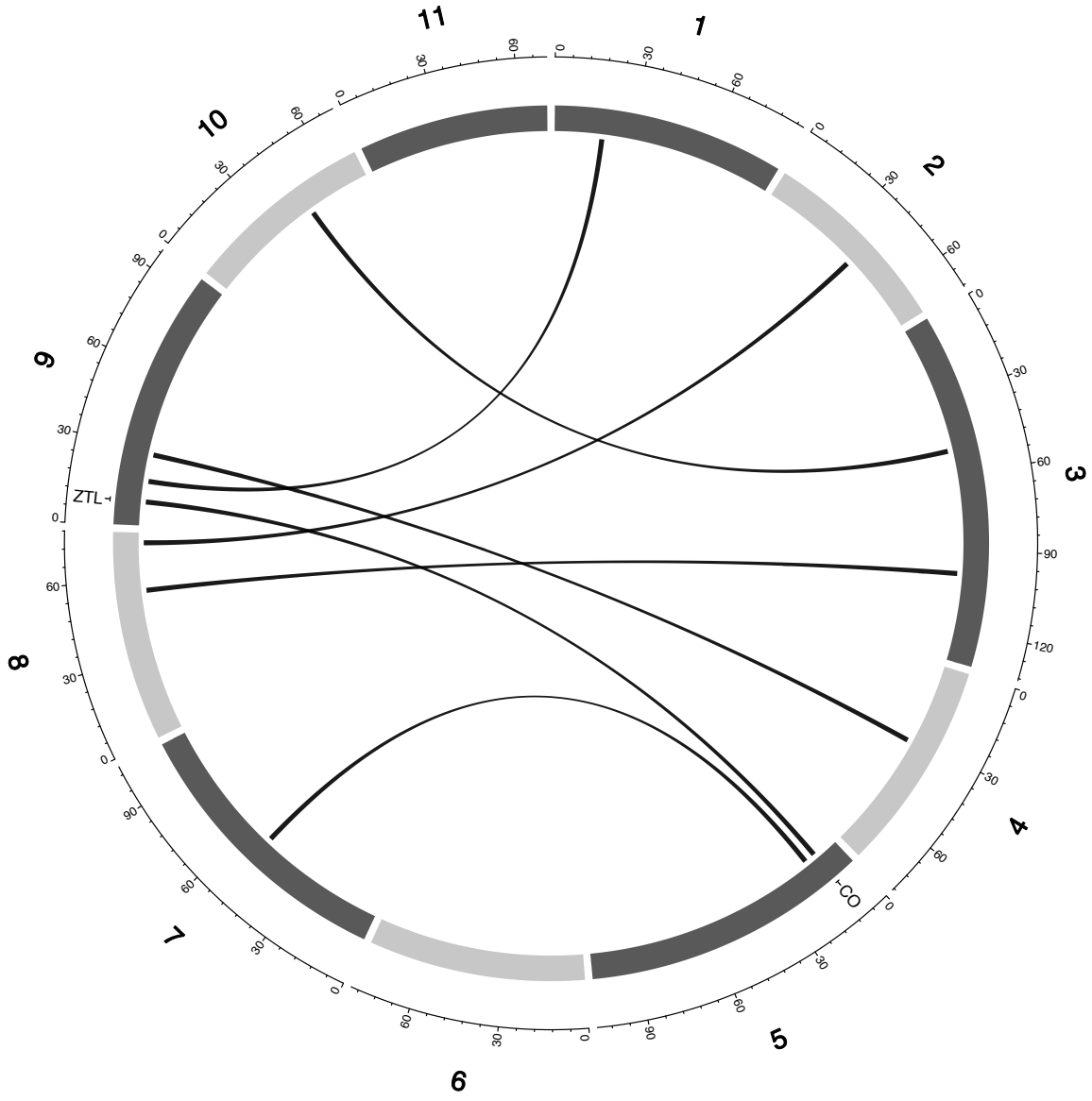


Figure S 8: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for MFISD (Maturity under full irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

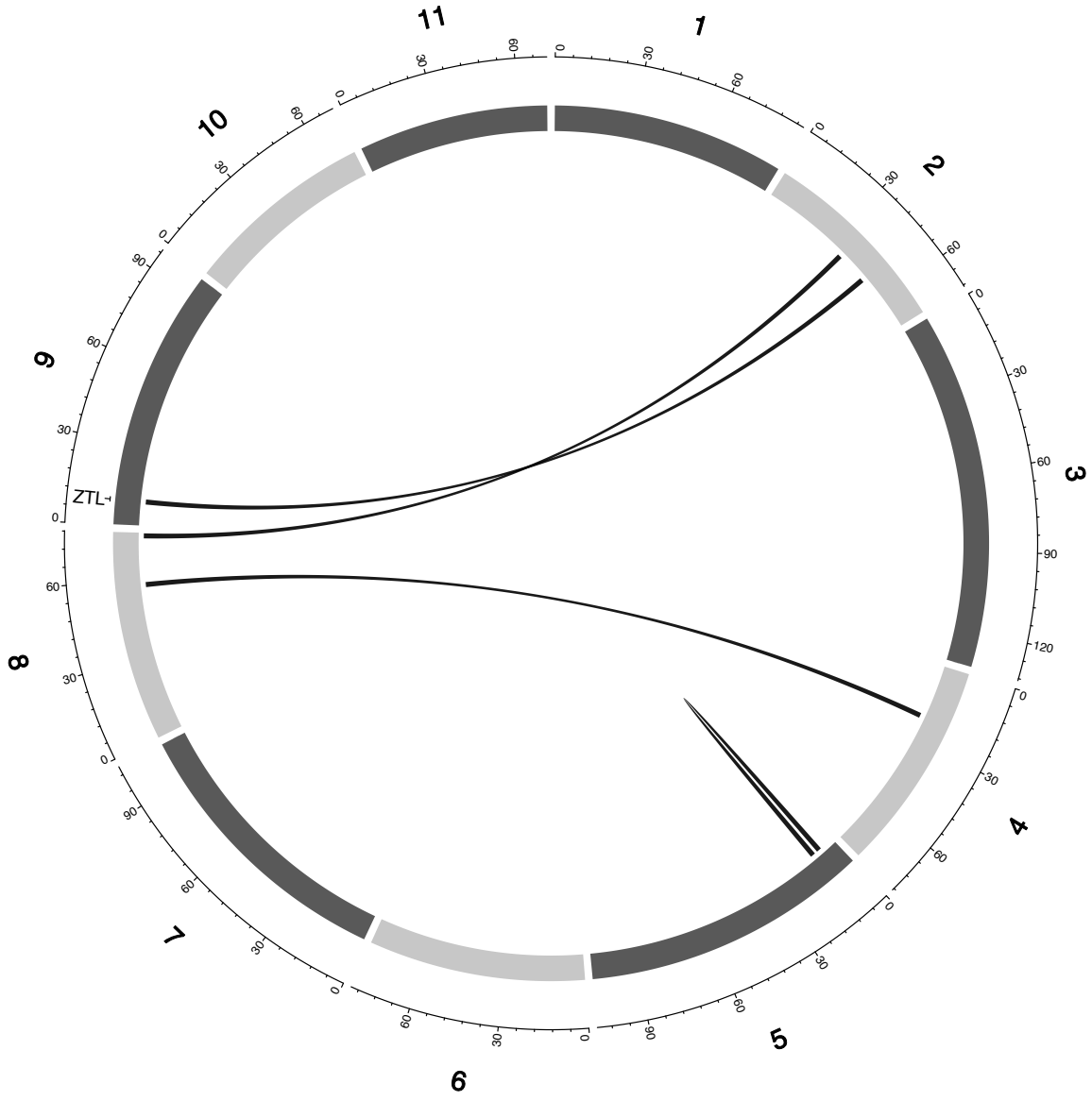


Figure S 9: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for MRISD (Maturity under restricted irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

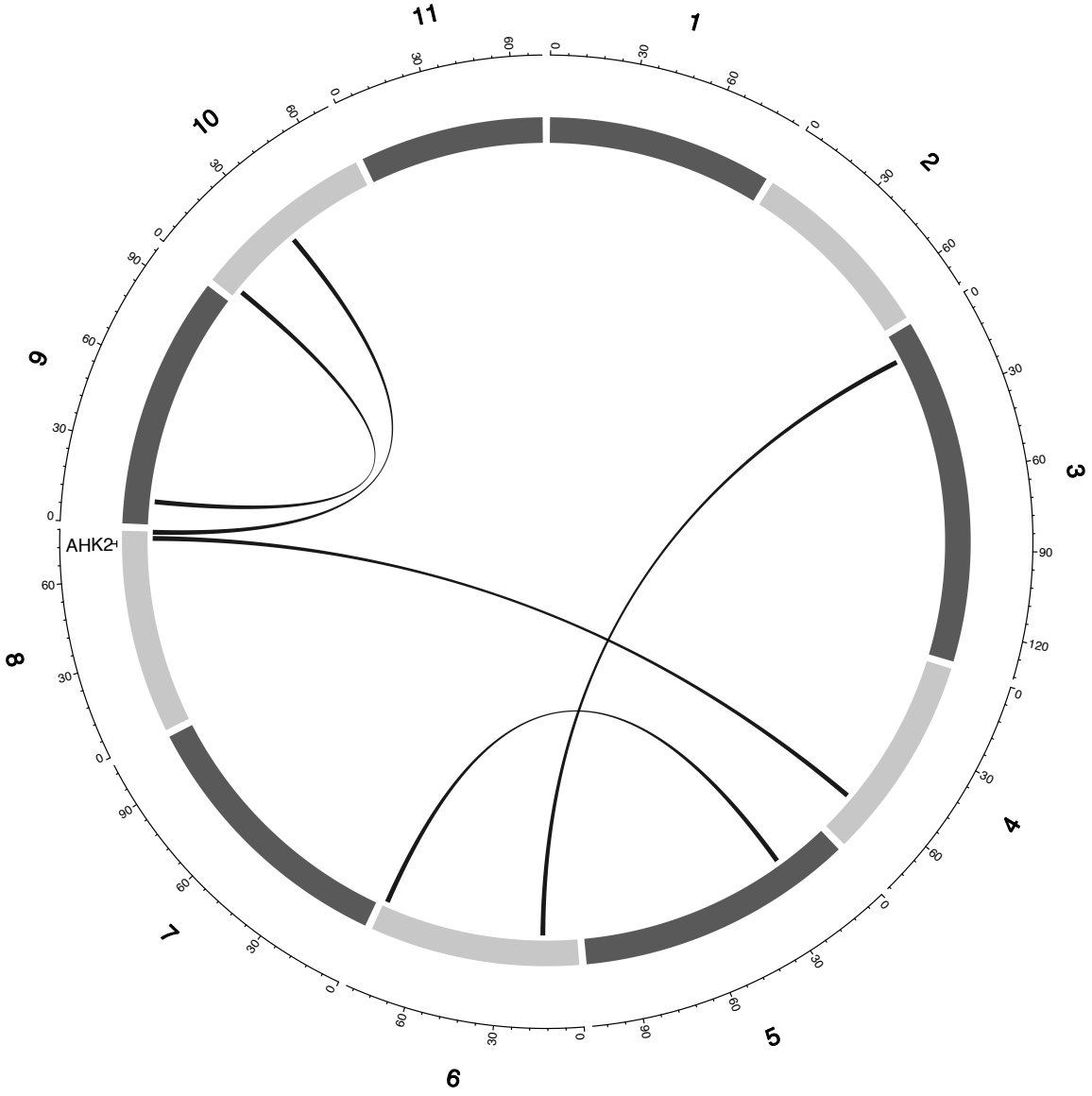


Figure S 10: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for SSFISD (Seed Size under full irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

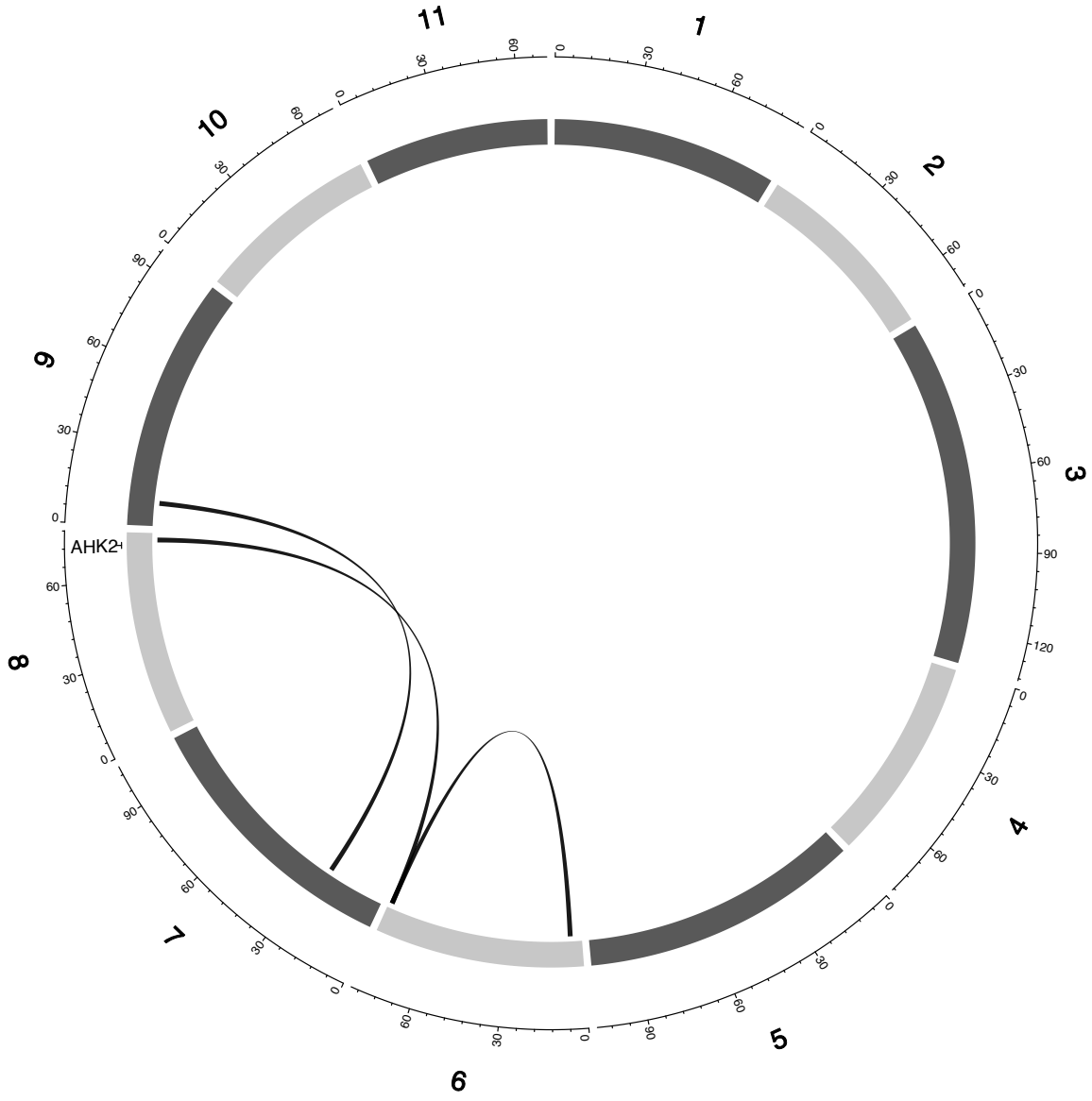


Figure S 11: Genetic map of the cowpea multiparent advanced generation inter-cross population (MAGIC) with pairwise interactions between epistatic QTL for SSRISD (Seed size under restricted irrigation and short day). Chromosomes are shown in shades of gray, two-way interacting loci are connected with black solid lines, and colocalized a priori genes are texts between chromosomes and genetic map.

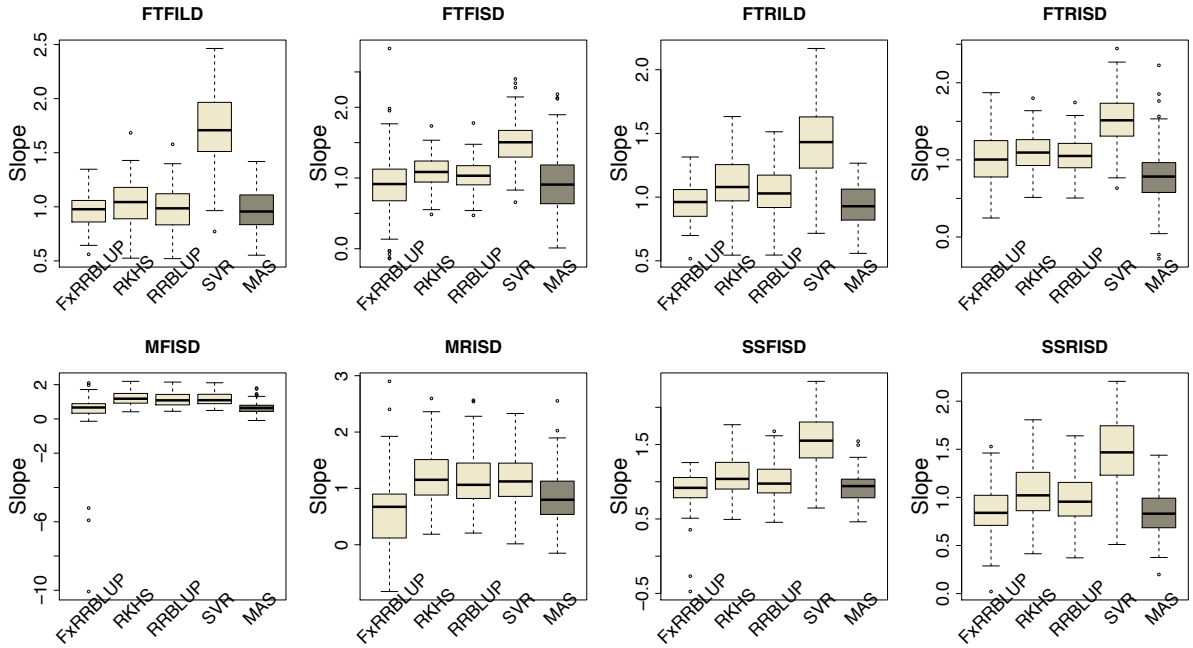


Figure S 12: Comparison of Slope of regression between observed and predicted trait values across GS and MAS models. Boxplots in each panel showed the distribution of slope values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short day (FTRISD), maturity under full irrigation and short day (MFISD), maturity under restricted irrigation and short day, seed size under full irrigation and short day (SSFISD), and seed size under restricted irrigation and short day (SSRISD).

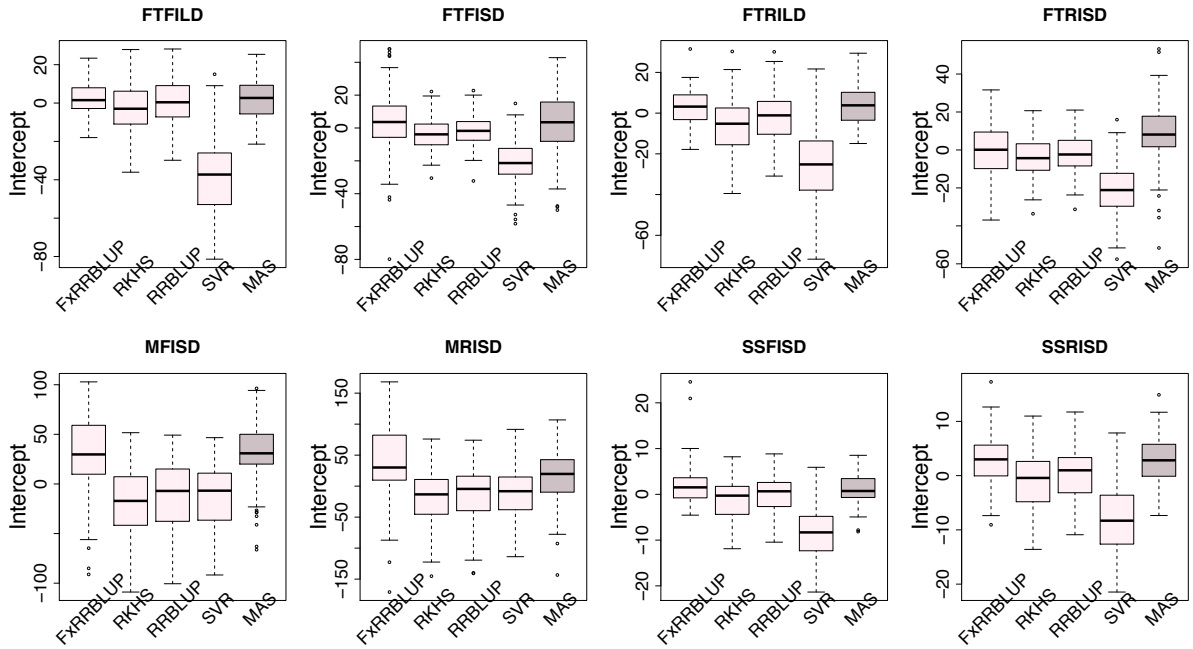


Figure S 13: Comparison of intercept of regression between observed and predicted trait values across GS and MAS models. Boxplots in each panel showed the distribution of intercept values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short

day (FTRISD), maturity under full irrigation and short day (MFISD), maturity under restricted irrigation and short day, seed size under full irrigation and short day (SSFISD), and seed size under restricted irrigation and short day (SSRISD).

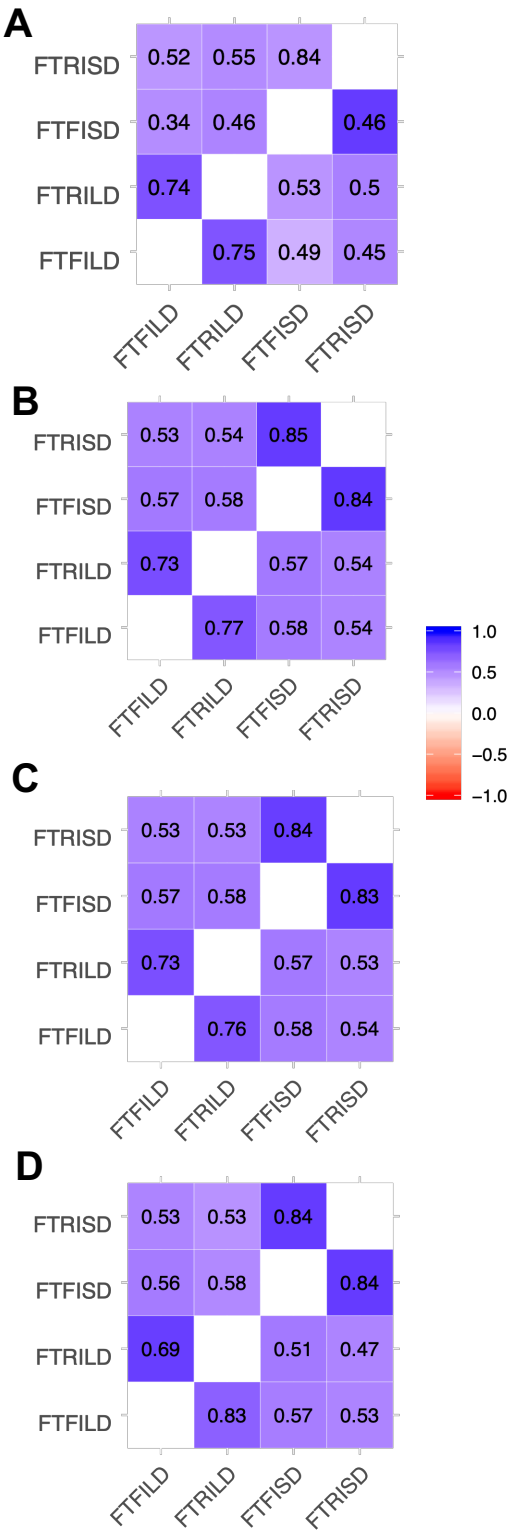


Figure S 14: Environment by environment prediction values across GS models. Boxplots in each panel showed the distribution of intercept values across 100 cycles for (A) FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), (B) RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), (C) RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), and (D) SVR (Support Vector Regression: Non-Parametric model) for flowering time under full irrigation and long day (FTFILD), flowering time under restricted irrigation and long day (FTRILD), flowering time under full irrigation and short day (FTFISD), flowering time under restricted irrigation and short day (FTRISD)

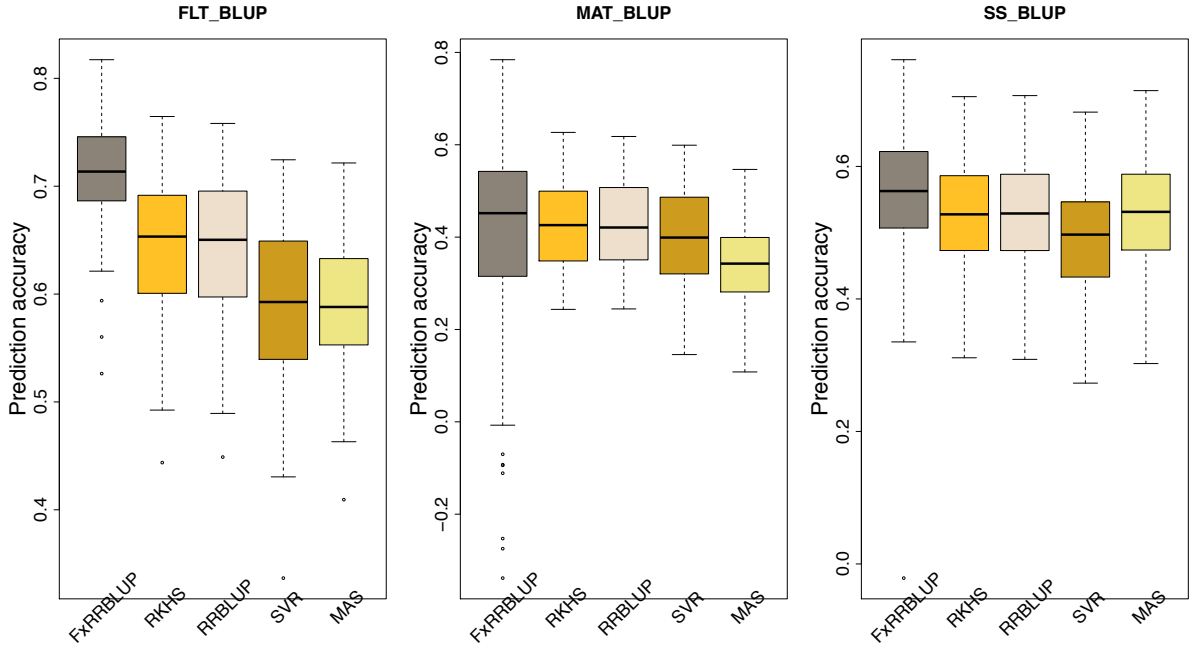


Figure S 15: Comparison of prediction accuracy across GS and MAS models. Boxplots in each panel showed the distribution of prediction accuracy values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time BLUP (FLT_BLUP), maturity BLUP (MAT_BLUP), seed size BLUP (SS_BLUP).

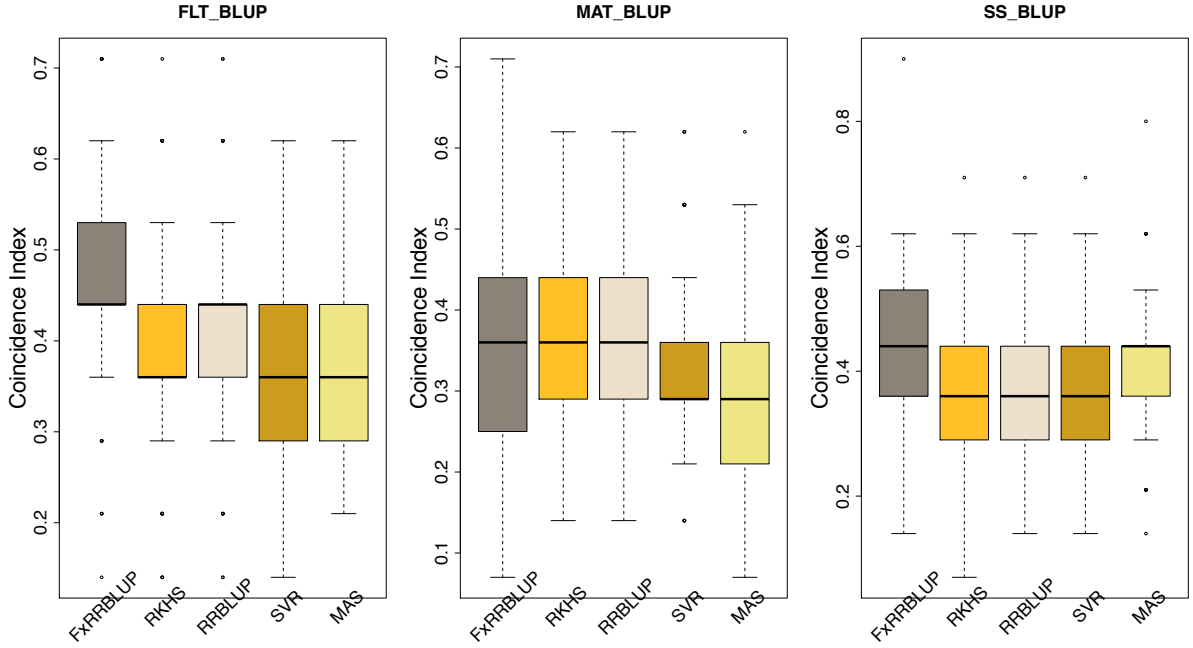


Figure S 16: Comparison of coincidence index across GS and MAS models. Boxplots in each panel showed the distribution of coincidence index values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time BLUP (FLT_BLUP), maturity BLUP (MAT_BLUP), seed size BLUP (SS_BLUP).

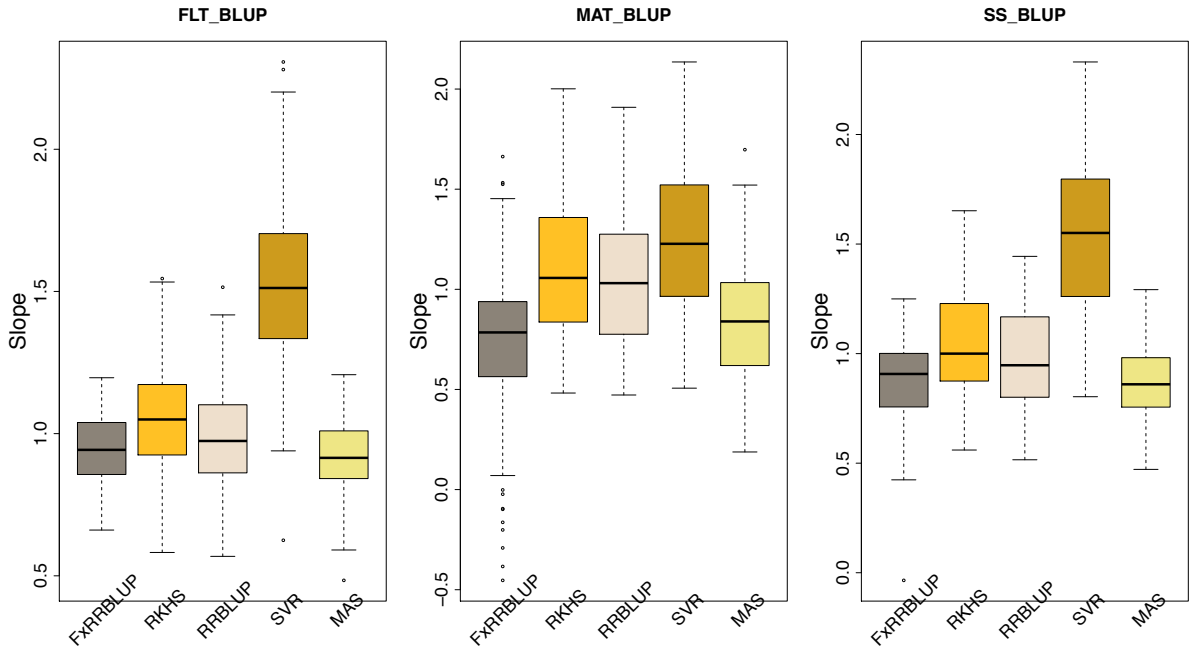


Figure S 17: Comparison of slope values across GS and MAS models. Boxplots in each panel showed the distribution of slope values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time BLUP (FLT_BLUP), maturity BLUP (MAT_BLUP), seed size BLUP (SS_BLUP).

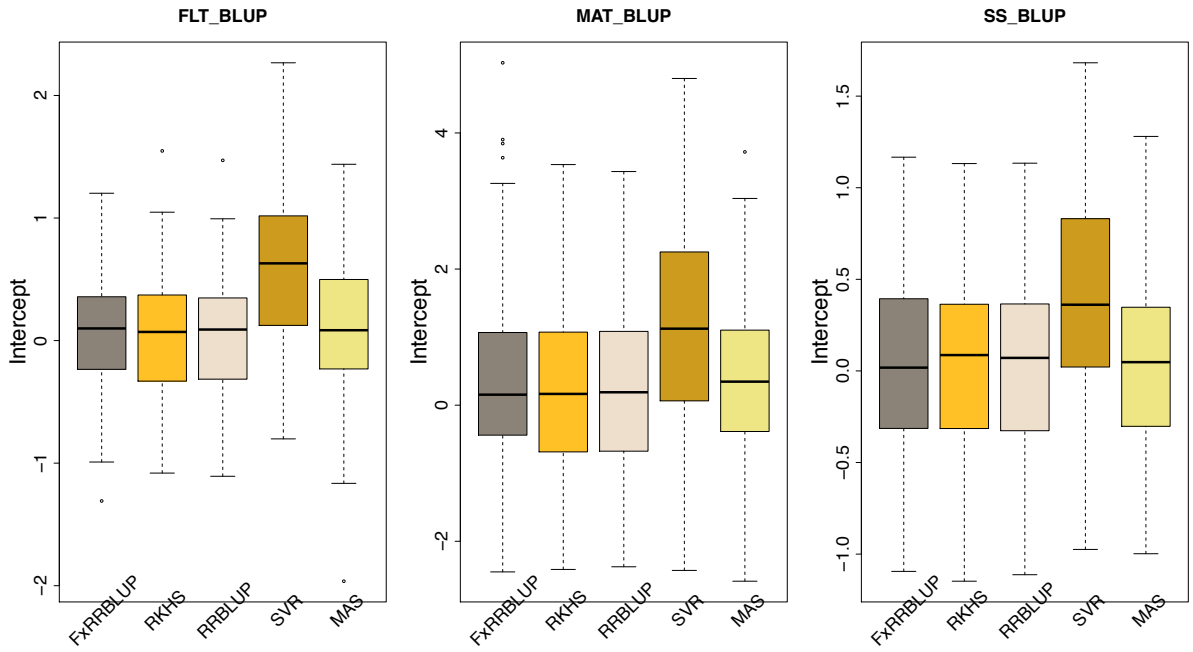


Figure S 18: Comparison of intercept values across GS and MAS models. Boxplots in each panel showed the distribution of intercept values across 100 cycles for FxRRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with fixed effects), RKHS (Reproducing Kernel Hilbert Space; Semi-Parametric model), RRBLUP (Ridge Regression Best Linear Unbiased Prediction: Parametric model with no fixed effects), SVR (Support Vector Regression: Non-Parametric model), and MAS (Marker Assisted Selection) for flowering time BLUP (FLT_BLUP), maturity BLUP (MAT_BLUP), seed size BLUP (SS_BLUP).

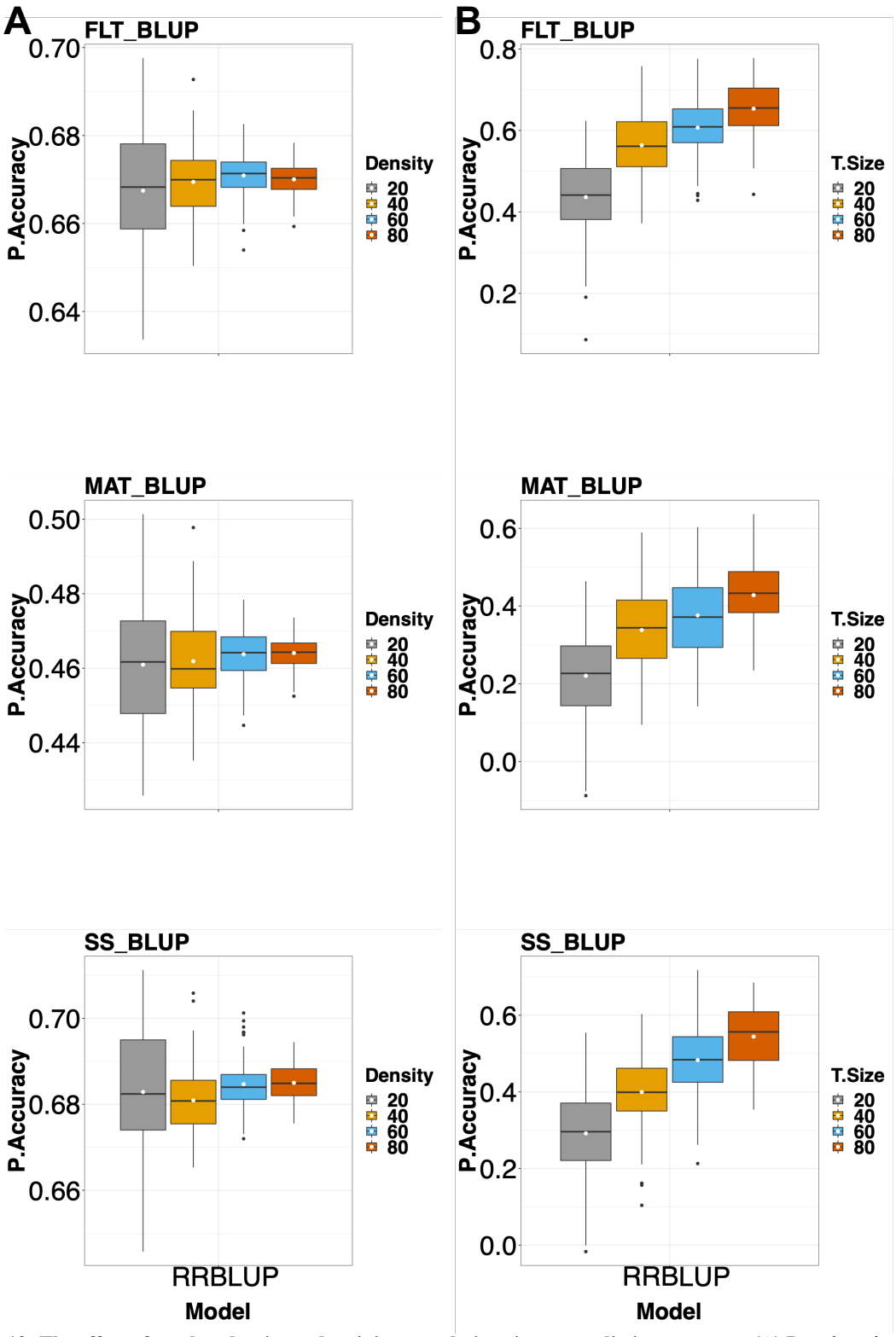


Figure S 19: The effect of marker density and training population size on prediction accuracy. (A) Boxplots showing comparison among different marker densities (20%, 40%, 60%, and 80%). (B) Boxplots showing comparison among different training population sizes (20%, 40%, 60%, and 80%).

Supplementary Tables

Table S 1: Environment by Environment correlation for flowering time.

Traits	FTFILD	FTRILD	FTFISD	FTRISD
FTFILD	-	0.76	0.50	0.44
FTRILD		-	0.57	0.52
FTFISD			-	0.86
FTRISD				-
Heritability (h^2)	0.41	0.42	0.48	0.46

Table S 2: Environment by environment correlation for maturity and seed size.

Traits	MFISD	MRISD	SSFISD	SSRISD
Heritability (h^2)	0.30	0.21	0.47	0.39
MFISD	-	0.64	-	-
SSFISD			-	0.80

Table S 3: QTL identified by stepwise regression explaining at least 6% of phenotypic variation.

Trait	QTL	Chr.	Pos.	PVE	ADE	MAF
FLT_BLUP	qVu9:24.77	9	24.7747	24.2	3.1	0.28

FLT_BLUP	qVu9:5.85	9	5.8471	13.0	2.1	0.4
FLT_BLUP	qVu11:49.06	11	49.0616	8.9	1.7	0.36
FLT_BLUP	qVu5:8.5	5	8.5038	7.8	1.6	0.48
FLT_BLUP	qVu4:29	4	29.0011	7.1	1.5	0.49
FTFILD	qVu9:24.77	9	24.7747	25.0	7	0.28
FTFILD	qVu9:22.65	9	22.648	19.5	5.6	0.49
FTFILD	qVu11:49.06	11	49.0616	10.9	4.4	0.36
FTFISD	qVu9:8.37	9	8.367	10.5	1.5	0.35
FTFISD	qVu4:31.3	4	31.2954	8.4	1.3	0.47
FTFISD	qVu5:8.5	5	8.5038	7.9	1.9	0.13
FTFISD	qVu4:20.34	4	20.3441	7.5	1.7	0.14
FTFISD	qVu2:40.37	2	40.3707	7.1	1.2	0.37
FTRILD	qVu9:23.36	9	23.3558	27.5	6.9	0.29
FTRILD	qVu5:8.5	5	8.5038	7.3	3.3	0.48
FTRILD	qVu11:49.06	11	49.0616	7.0	3.3	0.36
FTRISD	qVu4:19.99	4	19.9946	10.6	2.2	0.15
FTRISD	qVu1:55.11	1	55.111	8.2	1.4	0.47
FTRISD	qVu9:8.37	9	8.367	8.0	1.4	0.35
FTRISD	qVu4:31.3	4	31.2954	6.8	1.3	0.47
FTRISD	qVu5:3.9	5	3.8966	6.8	1.3	0.46
MAT_BLUP	qVu2:45.2	2	45.203	9.5	3.3	0.31
MAT_BLUP	qVu9:5.85	9	5.8471	8.6	3	0.38
MAT_BLUP	qVu4:19.99	4	19.9946	6.0	3.5	0.15
MFISD	qVu5:13.76	5	13.7582	8.2	3.5	0.35
MFISD	qVu9:5.85	9	5.8471	6.6	3	0.49
MRISD	qVu2:45.2	2	45.203	9.6	6.1	0.31
MRISD	qVu9:8.37	9	8.367	8.6	5.5	0.39
SS_BLUP	qVu8:74.29	8	74.2918	25.2	2.1	0.25
SS_BLUP	qVu6:78.35	6	78.3467	10.7	1.2	0.49
SSFISD	qVu8:74.21	8	74.2124	29.3	2.5	0.24
SSFISD	qVu6:78.35	6	78.3467	9.2	1.3	0.49
SSRISD	qVu8:76.81	8	76.8132	19.7	2.4	0.23
SSRISD	qVu6:78.35	6	78.3467	10.0	1.5	0.49

Quantitative trait loci (QTL), Chromosome (Chr.), Position (Pos. in centimorgan), Additive effect (ADE), Phenotypic variation explained (PVE), and Minor allele frequency (MAF).

1173 **Table S 4: Epistatic QTL identified by SPAEML and their effect sizes.**

Trait	QTL1	ADE1	MAF1	QTL2	ADE2	MAF2
FLT_BLUP	qVu9:25.39	3.1	0.28	qVu11:62.84	1.9	0.14
FLT_BLUP	qVu9:5.86	2.1	0.39	qVu11:42.83	1.8	0.14
FLT_BLUP	qVu5:12.79	1.7	0.34	qVu6:78.36	0.5	0.47
FLT_BLUP	qVu4:31.3	1.3	0.48	qVu6:1.47	1.6	0.15
FLT_BLUP	qVu1:66.57	1.6	0.12	qVu9:26.8	1.5	0.26
FLT_BLUP	qVu4:30.21	0.2	0.44	qVu7:45.81	0.4	0.43
FTFILD	qVu9:25.39	7	0.28	qVu11:50.94	4.6	0.26
FTFILD	qVu5:12.79	2.9	0.34	qVu11:35.28	1.6	0.40
FTFILD	qVu1:66.38	3.3	0.22	qVu4:31.03	2.3	0.47
FTFILD	qVu1:66.57	4.3	0.12	qVu5:52.97	0.9	0.37
FTFILD	qVu6:32.5	1.2	0.12	qVu9:86.49	1.1	0.34
FTFISD	qVu4:31.3	1.2	0.48	qVu9:28.65	0.8	0.41
FTFISD	qVu1:55.11	1.2	0.37	qVu5:25.01	0.4	0.10
FTFISD	qVu2:48.05	0.9	0.30	qVu9:8.37	1.4	0.39
FTFISD	qVu7:84.88	0.6	0.38	qVu10:10.07	0.8	0.36
FTFISD	qVu5:5.81	0.6	0.16	qVu5:8.91	1	0.49
FTRILD	qVu9:25.39	6.6	0.28	qVu11:62.84	3.8	0.14
FTRILD	qVu5:12.79	3.7	0.34	qVu6:0.68	2.4	0.39
FTRISD	qVu4:20.34	2.1	0.14	qVu9:8.37	1.3	0.39
FTRISD	qVu1:54.81	1.3	0.37	qVu4:31.3	1.2	0.48
FTRISD	qVu1:66.38	1.2	0.22	qVu11:49.06	1.1	0.39
FTRISD	qVu7:80.11	0.3	0.47	qVu8:37.41	0.6	0.48
MAT_BLUP	qVu2:48.05	2.8	0.30	qVu9:8.37	3	0.39
MFISD	qVu5:12.79	3.4	0.34	qVu7:50.14	2	0.27
MFISD	qVu2:39.29	2.4	0.36	qVu8:73.15	2.4	0.34
MFISD	qVu5:8.5	2.8	0.19	qVu9:8.37	2.8	0.39
MRISD	qVu2:48.05	5.3	0.30	qVu9:8.37	5.5	0.39
MRISD	qVu2:35.19	5.4	0.21	qVu8:75.88	4.8	0.24
SS_BLUP	qVu6:78.36	1.1	0.47	qVu7:18.1	0.4	0.11
SS_BLUP	qVu6:78.35	0.8	0.13	qVu8:74.29	2.1	0.25
SSFISD	qVu4:62.75	0.8	0.10	qVu8:74.29	2.5	0.25
SSFISD	qVu5:20.54	0.6	0.11	qVu6:78.36	1.2	0.47
SSRISD	qVu6:78.35	1	0.13	qVu8:74.29	2.2	0.25
SSRISD	qVu6:3.67	0.8	0.12	qVu6:78.36	1.4	0.47

1174 Quantitative trait loci (QTL), Linkage group (LG), Position (Pos. in centimorgan), Additive
1175 effect (ADE), , and Minor allele frequency (MAF).
1176

1177 **Table S 5: Mean and standard deviation of prediction accuracy across GS and MAS models.**

Trait	FxRRBLUP	RKHS	RRBLUP	SVR	MAS
FT_BLUP	0.71±0.05	0.65±0.07	0.65±0.07	0.59±0.08	0.59±0.06
FTFILD	0.68±0.05	0.55±0.08	0.55±0.08	0.54±0.08	0.64±0.06
FTFISD	0.56±0.20	0.59±0.07	0.59±0.07	0.55±0.08	0.33±0.11
FTRILD	0.68±0.07	0.58±0.08	0.58±0.07	0.50±0.08	0.61±0.08
FTRISD	0.58±0.17	0.58±0.07	0.58±0.07	0.53±0.08	0.25±0.10
MT_BLUP	0.40±0.23	0.42±0.09	0.42±0.09	0.40±0.10	0.33±0.09
MFISD	0.30±0.20	0.39±0.09	0.39±0.09	0.36±0.09	0.20±0.10
MRISD	0.25±0.29	0.37±0.11	0.37±0.11	0.34±0.12	0.30±0.10
SS_BLUP	0.56±0.10	0.52±0.08	0.53±0.08	0.49±0.09	0.53±0.09
SSFISD	0.58±0.14	0.54±0.08	0.54±0.08	0.51±0.09	0.57±0.08
SSRISD	0.50±0.11	0.45±0.08	0.45±0.08	0.43±0.09	0.47±0.10

1178

1179 **Table S 6: Mean and standard deviation of coincidence index of GS and MAS models.**

Trait	FxRRBLUP	RKHS	RRBLUP	SVR	MAS
FLT_BLUP	0.47±0.11	0.40±0.11	0.42±0.11	0.37±0.10	0.37±0.09
FTFILD	0.49±0.09	0.37±0.10	0.37±0.10	0.35±0.11	0.45±0.10
FTFISD	0.43±0.16	0.44±0.10	0.44±0.10	0.42±0.09	0.30±0.11
FTRILD	0.40±0.09	0.36±0.10	0.35±0.10	0.28±0.10	0.37±0.10
FTRISD	0.45±0.15	0.43±0.09	0.42±0.10	0.42±0.10	0.26±0.09
MT_BLUP	0.34±0.15	0.35±0.09	0.36±0.10	0.33±0.11	0.27±0.10
MFISD	0.32±0.12	0.30±0.09	0.30±0.10	0.31±0.09	0.26±0.09
MRISD	0.31±0.18	0.34±0.10	0.35±0.10	0.33±0.10	0.33±0.10
SS_BLUP	0.44±0.11	0.37±0.11	0.36±0.10	0.36±0.10	0.42±0.11
SSFISD	0.48±0.11	0.44±0.09	0.44±0.09	0.43±0.09	0.46±0.10
SSRISD	0.42±0.11	0.36±0.11	0.37±0.11	0.36±0.10	0.40±0.10

1180
1181