AGRONOMIC, GENETIC AND GENOMIC APPROACHES FOR PREDICTING HETEROSIS IN SORGHUM [*Sorghum bicolor* (L.) Moench]

by

FRANK MAULANA

B.S., Bunda College of Agriculture, University of Malawi, 2004
M.S., Kansas State University, 2011

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

The approach used to identify inbred lines that can produce superior hybrids is costly and time-consuming. It requires creation of all possible crosses and evaluation of the crosses to estimate combining abilities for the desired traits. Predicting heterosis or hybrid performance in any way possible may help to reduce the number of crosses to be made and evaluated. In this study, four sets of experiments were conducted to determine whether heterosis can be predicted based on inbred line performance, genetic distance between parents and genomic prediction model.

The first experiment was aimed at assessing the levels of genetic diversity, population structure and linkage disequilibrium (LD) in 279 public sorghum inbred lines, based on 66,265 SNPs generated using the genotyping-by-sequencing (GBS) platform. The inbred lines were developed at different times over the last two decades and harbor robust diversity in pedigree and agronomic characteristics. Some of the inbreds are resistant to Acetolactate synthase (ALS) and Acetyl co-enzyme-A carboxylase (ACC) inhibitor herbicides. The mean polymorphic information content (PIC) and gene diversity across the entire inbreds were 0.35 and 0.46, respectively with non-herbicide resistant inbreds harboring more diversity than the herbicide resistant ones. The population structure analysis clustered the inbred lines into three major subgroups according to pedigree and fertility-reaction with the maintainer lines (B-lines) distinctly forming a separate cluster. Analysis of molecular variance (AMOVA) revealed more variation within subgroups than among subgroups. Substantial linkage disequilibrium (LD) was detected between the markers in the population with marked variation between chromosomes. This information may facilitate the use of the

inbreds in sorghum breeding programs and provide perspectives for optimizing marker density for gene mapping and marker-assisted breeding.

The second experiment, based on 102 F1 hybrids developed by intercrossing closely and distantly related inbreds, was conducted to investigate the relationship of genetic distance between parents with hybrid vigor or heterosis. The F1 hybrids alongside their parents were evaluated at two environments in a randomized complete block design with three replications. The results show that correlations of genetic distance between parents with hybrid performance and heterosis were variable and dependent on the trait. Though most were statistically non-significant and not strong to be used as predictor for heterosis, the results tend to show that certain level of genetic distance between parents is needed to capture maximum heterosis and hybrid performance.

The objective of the third research study was to determine whether traits measured on parents can be used to predict hybrid performance in sorghum and to assess the combining ability of selected inbreds. Forty-six parental inbred lines and 75 F_1 hybrids generated from intercrossing the inbreds were evaluated in four environments in a randomized complete block design with three replications. The average performance of the parents (mid-parent) was significantly correlated with hybrid performance for thousand kernel weight, days to flowering and plant height. Significant general (GCA) and specific (SCA) combining abilities were observed for most traits, with highly significant GCA effects observed for most traits as compared to SCA indicating that additive genetic effects are more important in affecting the inheritance of the traits measured. Results show that studying parental inbred line performance could generate important information for predicting hybrid performance in sorghum.

The fourth experiment was aimed at assessing the efficacy of genomic prediction of hybrid performance in sorghum. Genomic prediction was performed with five-fold cross-validation procedure on 204 F1 hybrids developed using 102 inbred lines. A total of 66,265 SNP markers generated using genotyping-by-sequencing were used in this study. Results showed that increasing training population size increased prediction accuracies for all traits with the effect being different for different traits. Also, considering additive effects alone versus additive and dominance effects in the model showed similar trend of prediction accuracy but the full model (considering both additive and dominance effects of the markers) provided better prediction at least for some of the traits. The results suggest that genomic prediction could become an effective tool for predicting the performance of untested sorghum hybrids thus adding efficiency to hybrid selection.

AGRONOMIC, GENETIC AND GENOMIC APPROACHES FOR PREDICTING HETEROSIS IN SORGHUM [*Sorghum bicolor* (L.) Moench]

by

FRANK MAULANA

B.S., Bunda College of Agriculture, University of Malawi, 2004
M.S., Kansas State University, 2011

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agriculture
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Dr. Tesfaye Tesso

Copyright

Abstract

The approach used to identify inbred lines that can produce superior hybrids is costly and time-consuming. It requires creation of all possible crosses and evaluation of the crosses to estimate combining abilities for the desired traits. Predicting heterosis or hybrid performance in any way possible may help to reduce the number of crosses to be made and evaluated. In this study, four sets of experiments were conducted to determine whether heterosis can be predicted based on inbred line performance, genetic distance between parents and genomic prediction model.

The first experiment was aimed at assessing the levels of genetic diversity, population structure and linkage disequilibrium (LD) in 279 public sorghum inbred lines, based on 66,265 SNPs generated using the genotyping-by-sequencing (GBS) platform. The inbred lines were developed at different times over the last two decades and harbor robust diversity in pedigree and agronomic characteristics. Some of the inbreds are resistant to Acetolactate synthase (ALS) and Acetyl co-enzyme-A carboxylase (ACC) inhibitor herbicides. The mean polymorphic information content (PIC) and gene diversity across the entire inbreds were 0.35 and 0.46, respectively with non-herbicide resistant inbreds harboring more diversity than the herbicide resistant ones. The population structure analysis clustered the inbred lines into three major subgroups according to pedigree and fertility-reaction with the maintainer lines (B-lines) distinctly forming a separate cluster. Analysis of molecular variance (AMOVA) revealed more variation within subgroups than among subgroups. Substantial linkage disequilibrium (LD) was detected between the markers in the population with marked variation between chromosomes. This information may facilitate the use of the

inbreds in sorghum breeding programs and provide perspectives for optimizing marker density for gene mapping and marker-assisted breeding.

The second experiment, based on 102 F1 hybrids developed by intercrossing closely and distantly related inbreds, was conducted to investigate the relationship of genetic distance between parents with hybrid vigor or heterosis. The F1 hybrids alongside their parents were evaluated at two environments in a randomized complete block design with three replications. The results show that correlations of genetic distance between parents with hybrid performance and heterosis were variable and dependent on the trait. Though most were statistically nonsignificant and not strong to be used as predictor for heterosis, the results tend to show that certain level of genetic distance between parents is needed to capture maximum heterosis and hybrid performance.

The objective of the third experiment was to determine whether traits measured on parents can be used to predict hybrid performance in sorghum and to assess the combining ability of selected inbreds. Forty-six parental inbred lines and 75 F_1 hybrids generated from intercrossing the inbreds were evaluated in four environments in a randomized complete block design with three replications. The average performance of the parents (mid-parent) was significantly correlated with hybrid performance for thousand kernel weight, days to flowering and plant height. Both general (GCA) and specific (SCA) combining abilities were significant for most traits, with highly significant GCA effects observed for most traits as compared to SCA indicating that additive genetic effects are more important in affecting the inheritance of the traits measured. Results show that studying parental inbred line performance could generate important information for predicting hybrid performance in sorghum.

The fourth experiment was aimed at assessing the efficacy of genomic prediction of hybrid performance in sorghum. Genomic prediction was performed with five-fold cross-validation procedure on 204 F1 hybrids developed using 102 inbred lines. A total of 66,265 SNP markers generated using genotyping-by-sequencing were used in this study. Results showed that increasing training population size increased prediction accuracies for all traits with the effect being different for different traits. Also, considering additive effects alone versus additive and dominance effects in the model showed similar trend of prediction accuracy but the full model (considering both additive and dominance effects of the markers) provided better prediction at least for some of the traits. The results suggest that genomic prediction could become an effective tool for predicting the performance of untested sorghum hybrids thus adding efficiency to hybrid selection.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to thank God Almighty for His mercy and blessing on me and for giving me strength to carry out this research work. I would like also to express my sincere appreciation to my major advisor, Dr. Tesfaye Tesso, for his constructive guidance towards my study from the beginning to the end. My sincere thank goes to my committee members, Dr. Mary Beth Kirkham, Dr. Geoffrey Morris, Dr. Ramasamy Perumal, Dr. Jesse Poland and committee chair, Dr. Zhilong Yang for their advice and valuable suggestions towards my Ph.D research study that ultimately helped to improve this work. I would like also to thank my fellow labmates in the sorghum breeding team: Jebril Jebril, Lauren Lang, Dilooshi Weerasooriya, Dr. Dereje Dugassa and Daniel Hopper for their unwavering support during field operations and data collection.

I am also thankful to all my family members who have been with me all the way in both good and bad times. Special appreciation goes to my mother and father, Abbey Kapangaza, and Reginald Maulana, respectively, who tried their best to send me to school so that I can grow up to become a productive person in the society. My sincere gratitude and appreciation go to my lovely wife, Enellece Maulana and my children: Mayamiko and Mirriam Maulana for their patience, continued support and encouragement during my study.

Lastly, I would like to thank the Kansas Grain Sorghum commission for the financial support towards my study, and the Kansas State University, Department of Agronomy for provision of required facilities used to conduct my research.

General Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is one of the most important grain crops worldwide. It is used as staple food in the developing countries in Africa and Asia, and mainly as feed source in the developed world. In the United States, sorghum ranks third after maize and wheat in total production (FAO, 2004). Unlike in the developing countries where mainly open-pollinated varieties are grown, sorghum in the US and other industrialized countries are exclusively grown as hybrids to exploit the superior performance that results from heterosis (Duvick, 1999; Birchler et al., 2003). Hybrid vigor or heterosis is a situation where the performance of a hybrid is better than its inbred parents and it is the most important component of crop productivity. The concept of heterosis was first utilized in maize (*Zea mays* (L) breeding, which significantly increased maize yields by 15% compared to the superior open-pollinated varieties. Furthermore, hybrid maize accounted for 65% of total maize cultivation contributing to a tremendous increase of annual maize production by the late twentieth century (Duvick, 1999). In rice, the hybrid yields were increased by 20-30% which ultimately increased average production by 44.1% (Cheng et al., 2007). In sorghum, at the end of the twentieth century, almost half of the world sorghum production was hybrids, contributing about 35-40% yield gains in the United States (Duvick, 1999)

Although hybrid breeding programs have been a great success in increasing yields in several cereal crops including sorghum, the process of identifying parental inbred lines that can produce superior hybrids still remains the most time-consuming and expensive activity. In sorghum, for example, the current approach is to select parents and cross them to tester parents and evaluate the crosses to estimate hybrid performance and combining ability of the parents for the desired traits in multiple locations and years. Superior parents (males and females) are then intercrossed in all possible combination and tested to identify the most promising hybrid for commercial consideration. This approach takes time and also requires a

lot of resources. Therefore, predicting hybrid performance in any way possible may help to reduce the number of crosses to be made and evaluated, thus may ultimately reduce phenotyping costs.

A number of previous studies have reported inconsistent results about the relationship between the performance of inbred parents and their hybrids. Some of the studies reported low correlation between inbred line and hybrid performance especially for complex traits such as grain yield (Hallauer and Miranda, 1988), implying that inbred line performance for complex traits such as yield may not be a good predictor of hybrid performance. This is due to the genotype-by-environment interaction effect leading to high phenotypic plasticity as reflected in their low heritability value (Sadras and Slafer, 2012). However, in other studies positive correlations between the inbred line (mid-parent) and hybrid performance in maize have been reported (Flint Garcia et al., 2009; Prado et al., 2013; Ertiro et al., 2013).

Furthermore, quantitative genetic theory states that there is a positive correlation between genetic distance between parental inbred lines and heterosis. However, a strong linkage between a large proportion of markers with the QTLs controlling the trait of interest is needed. Again, heritability of the trait of interest should be high enough and with strong dominance effects. Failure to satisfy these important pre-requisites may lead to no correlation between genetic distance and heterosis (Fred et al., 1986; Zhang et al., 1994, 1995; Godshalk et al., 1990; Melchinger et al., 1990a, b). In many cereal crops including rice, maize and wheat, genetic distance between parental inbred lines based on molecular or allozyme markers have been used to identify hybrids with superior performance (Zhang et al., 1994; 1995; Martin et al., 1995; Lanza et al., 1997).

Positive correlations between marker based genetic divergence of the parental lines and heterosis in their hybrids have been reported (Reif et al., 2003; Krystkowiak et al., 2009) and this has been recognized as potential tool for identifying parents that would produce

superior hybrids (Smith et al., 1990; Zhang et al., 1995; Lanza et al., 1997). This has been found to be working across species including rice (Zhang et al., 2010), sorghum (Jordan et al., 2003) and maize (Smith et al., 1990). In rice, significant and positive correlations between molecular marker genetic distance and heterosis have been reported (Zhao et al., 2008). Also, Zhang et al (2010) found positive correlations between the breadth of genetic distance among parental lines and heterosis for plant height and panicle length. Therefore, genetic diversity measures may serve as tool to predict heterosis and thus reduce costs associated with developing and phenotyping a large number of crosses to identify superior hybrids. There were also few reports that do not agree with these findings. In a study conducted by Zhang et al. (2006) and Liao et al. (1998) no correlation was observed between the genetic distance among the parents and heterosis implying that genetic distance may not be the best predictor of heterosis (Zhang et al., 1994; Zhang et al., 1995; Xiao et al., 1996).

Hybrid is the most preferred type of cultivar for various crops including maize, sorghum and rice, and has immensely contributed to increased yields globally. Nevertheless, selection of superior hybrids still remains a trial and error process, and therefore, the greatest challenge in hybrid breeding is how to predict the performance of crosses based on the existing data on the parents. Over the years, marker-assisted selection (MAS) has been successfully used in different crops to incorporate major genes and/or large-effect quantitative trait loci (QTLs) in cereal crops (Gregorio et al., 2013; Tuberosa et al., 2007; Araus et al., 2008). However, it is not a viable option for complex traits controlled by many genes of small effect and/or by a combination of major and minor genes. Genomic prediction is an alternative to MAS in that it utilizes genome-wide markers to build the prediction model used to estimate the breeding values of individuals that have been genotyped but not phenotyped (Meuwissen et al., 2001). Moreover, genomic prediction helps to improve gain from selection, reduce phenotyping costs and speed up the development of new cultivars by

reducing the breeding cycle (Heffner, 2010). Recently, genomic prediction models have successfully been used to predict the performance of untested hybrids in wheat (Zhao et al., 2013), maize (Technow et al., 2012), rice (Xu et al., 2014) and canola (Jan et al., 2016).

This research study has four components. The first part focuses on investigating the genetic diversity, population structure and linkage disequilibrium in sorghum public inbred lines using genotyping-by-sequencing (GBS) approach. The second part of the study focuses on investigating whether there is any association between genetic distance estimated using genome-wide SNP markers generated using GBS between parental inbred lines and heterosis or hybrid vigor. The third part of the study focuses on investigating whether hybrid performance or heterosis can be predicted based on inbred line performance. The last and fourth part focuses on determining whether genomic prediction models can be used in sorghum hybrid breeding program to predict the performance of untested sorghum hybrids.

Chapter 1 - Heterosis: Literature Review

Heterosis

Heterosis is when the performance of the hybrid is better than the performance of its inbred parents. It dates back to 1876 when Charles Darwin observed that the progenies of a cross between two maize inbred parents expressed 25% increase in plant height compared to progenies from inbred parents. Since its rediscovery (Shull, 1908; East, 1908), heterosis has resulted in dramatic increase in crop productivity in many crops including sorghum, maize and rice. In maize, for example, grain yield increased nearly by fivefold from the late 1930s to 2004 largely due to the success in exploiting heterosis (Duvick, 2005). In sorghum, hybrid breeding program was fully adopted after the discovery of cytoplasmic-genetic male sterility system which made crossing easier. Similar to maize, heterosis in sorghum hybrid is expressed in terms of grain yield, where one parent contributes a high grain number per panicle than the other (Miller and Kebede, 1984). Reports from earlier studies show that grain yields of sorghum hybrids increased by 58 and 22% compared to a better parent under dryland and irrigated conditions, respectively (Quinby et al., 1958). Similarly, Doggett (1969) conducted hybrid evaluation trials in four countries at 391 locations and found hybrid performance was better than that of the better parent, with the yield advantage more pronounced under dryland environments.

Genetic basis of heterosis

Despite its marked expression for several traits in various crops, there is no broad consensus regarding its genetic basis of heterosis. However, three hypotheses have been proposed as the mechanisms underlying its genetic basis and these include dominance (Jones, 1917; Collins, 1921), over-dominance (Shull, 1908; Shull,1946) and epistasis (Schnell and Cockerham, 1992). The dominance hypothesis states that heterosis is due to dominant alleles from either parent cancelling the genetic effects of deleterious recessive alleles contributed

by the other parent in the heterozygous hybrid (Davenport, 1908). The over-dominance hypothesis states that the combination of two alleles at a given locus produces a genetic effect that is superior to that of either of the homozygous combinations of those alleles at that locus, implying that the two alleles complement each other and there is over-expression of a particular set of genes in the heterozygote. Epistasis hypothesis states that heterosis is due to the interactions between genes present at two or more different loci that produce the phenotype.

Experimental studies to determine the genetic basis of heterosis

Experiments have been conducted to validate any of these hypotheses to the exclusion of the others mainly using maize as a model crop. Using generation means or diallele analyses in maize, over-dominance and epistasis have been reported to be the two mechansims underlying the genetic basis of heterosis, although this claim was later refuted by Hinze and Lamkey (2003) and Mihaljevic et al. (2005). However, despite its popularity, it was widely recognized as inadequate because it could not explain progressive heterosis or rapid rate of inbreeding depression in tetraploids as well as absence of a decline in degree of heterosis over 50 years of genetic improvements (Duvick, 2001; Birchler et al., 2003). In maize, for example, QTL mapping study for grain yield was conducted and 11 QTLs for grain yield which showed over-dominance or over-dominance (Stuber et al., 1992). Fine mapping of one major QTL on chromosome 5 dissected this region into two smaller QTLs in repulsion phase linkage, which also demonstrated dominance (Graham et al., 1997; Reif et al., 2005). Similarly in another Design III study in maize (Lu et al., 2003) proposed that QTL for grain yield in maize exhibits true over-dominance.

In rice, at least for one study (Yu et al., 1997), epistasis was found to be responsible for heterosis in yield and yield components compared to over-dominance and dominance. In addition, estimates of over-dominant gene action contributing to heterosis have been reported

in maize using North Carolina Design III mating design (Halluaer et al., 2010). Furthermore, the importance of dominance versus over-dominance was demonstrated by recurrent selection studies (Robin et al., 1956) whereby populations were tested in crosses with each other, or with an inbred female parent. After comparing the response to selection of different populations that were selected with regard to performance of a cross with an inbred tester compared with a population tester, dominance is reported to be the primary genetic basis of heterosis. But again if over-dominance is the primary genetic basis of heterosis, the inbred tester may improve the population more than the population tester because alleles are fixed in the inbred while in a population, they are intermediate in frequency. In this study the inbred and population tester both improved performance of the population and this was in agreement with the importance of dominance relative to over-dominance. The idea was that if over-dominance was the primary basis of heterosis, the populations would diverge because of selection and also increase the number of homozygous alternative alleles in the populations to maximize heterozygosity and the population cross performance. The result may enhance the performance of the population crosses and decrease performance of the populations *per se.*

If dominance and epistasis were the primary mechanisms of genetic basis of heterosis, the favorable allele frequency would increase in the population, and in the crosses of the population resulting in increased performance of the populations and their crosses. The finding from this study was that there was substantial increase in the performance of populations which supported the importance of dominance versus over-dominance. Results of recent QTL mapping studies in maize also showed dominance as the major contributor to heterosis for yield and yield components and growth parameters such as plant height (Garcia et al., 2008). Besides, studies conducted in rice using recombinant inbred line (RIL) population, from the cross between indica and japonica showed that dominance was the primary basis of heterosis based on evidence from QTL, the absence of significant epistatic

interactions, and the relatively weak relationship of marker heterozygosity with performance for most traits. However, two inbred lines from the population exceeded the hybrid performance, suggesting that under the dominance hypothesis, it is possible to produce a homozygous individual with all the favorable alleles that produced the observed hybrid performance.

Although QTLs associated with over-dominance in populations derived from heterotic maize hybrids for traits such as yield and plant height have been reported (Edwards et al., 1987; Stuber et al., 1992), subsequent genetic dissection of a QTL with estimated over-dominance gene action suggested that the QTL detected originally might be separated into two, QTLs linked in repulsion phase with genetic effects due to dominance (Graham et al., 1997; Lariepe et al., 2012). Moreover, earlier study done in tomato reported over-dominance as the possible genetic mechanism driving heterosis for two major yield components (number of flowers per plant and weight of the fruit) contributing to overall yield in hybrid tomatoes (Krieger et al., 2010).

Likewise, there is also substantial evidence on the role epistasis may have in affecting heterosis. Wolf and Hallauer (1997) used various empirical and statistical approaches to support the role of epistasis in heterosis. In their study, they used triple testcross analysis by comparing the relative performance of segregating progeny when testcrossed to both parents and to the F_1 hybrid. And deviation in performance of the F_1 testcross from the average of the parental testcrosses is in agreement with gene action due to epistasis. Following this approach, epistasis for multiple traits including yield, yield components, and development time stages among progeny of the heterotic hybrid B73 × Mo17 was detected. Furthermore, a number of QTL mapping studies in rice have been reported to underline the role of epistasis (Yu et al., 1997; Li et al., 2001; Hua et al., 2003). Depending on the type of experimental materials and approaches used in each experiment, this epistasis effect can take different

forms additive × additive epistasis or dominance × dominance epistasis (Yu et al., 1997; Li et al., 2001; Hua et al., 2003). In general, depending on the type of parents, species and the nature of the population from which parents are derived or the type of progeny tested, the relative contribution of dominance, over-dominance and epistasis seem to vary greatly with dominance swimmingly is the major contributor.

Traditionally, marker-assisted selection (MAS) has been used in crop improvement to indirectly select desirable traits for improvement. Over the years, several types of molecular markers including restriction fragment length polymorphism (RFLP), simple sequence repeats (SSRs), random amplification of polymorphic DNA (RAPD), single nucleotide polymorphism (SNPs) and amplified fragment length polymorphisms (AFLPs) have been developed and used effectively in crop improvement or genetic analysis of various traits. However, the use of single nucleotide polymorphisms (SNPs) as DNA markers for plant genotyping has led to significant improvement in capability to score variation in specific DNA targets. SNPs are most abundant in a genome and suitable for whole-genome analysis (Zhu et al., 2003). More importantly, the cost associated with SNP discoveries have tremendously decreased because SNP-based marker techniques have been improved in marker density compared with the earlier genotyping approaches.

The development of genotyping-by-sequencing (GBS) approach has provided a great opportunity for SNP discovery and genotyping at an affordable price (Elshire et al., 2011). Depending on the objective of the study, the extracted genomic DNA is digested with a particular restriction enzyme and these restriction enzymes include *ApeKI*. Once the digestion of the DNA is done, the barcode adapter is ligated to the sticky ends of the digested DNA. Then to increase the number of DNA copies, the samples are subjected to PCR amplification. Subsequently, the amplified DNA samples are pooled together for sequencing. GBS has accorded the research community an opportunity to identify common genetic

variants associated with traits of interest in genome-wide association studies (GWAS) (Morris et al., 2013), genomic diversity studies (Fu et al., 2014) and genomic prediction studies (Zhao et al., 2013; Windhaussen et al., 2012). With GBS, there is no need to understand the nature of the genome of the species before analysis and also new SNPs are discovered while genotyping (Poland and Rife, 2012; Narum et al., 2013).

Prediction of hybrid vigor or heterosis

The major challenge in hybrid breeding programs is lack of efficient method to select parents that can produce superior hybrids, and also select superior hybrids out of many potential single-cross hybrids without extensive field evaluation. In sorghum hybrid breeding programs, the current approach is to select potential parents and cross them in all possible combinations and evaluate the crosses in multiple locations and years to estimate combining ability for the desired traits. However, this approach is time-consuming and very expensive because it requires a lot of resources such as land which is becoming a limiting factor and worse still after field evaluation most of the hybrids end up being discarded because of low general performance. To address this challenge, two approaches have been proposed which include the use of agro-morphological and high-density molecular marker data to predict hybrid performance.

Prediction of hybrid performance based on agronomic traits

Different approaches have been proposed to predict heterosis or hybrid performance based on agronomic data. In sorghum, for example, the performance of parents coupled with their combining ability estimates are important criteria used to select parents for developing superior hybrids. The combining ability analysis is a powerful and traditional approach to test the value of parents to produce superior hybrids. Selecting parents based on the combining ability increases the chance of developing high performing hybrids. In fact, parents with

higher general combining ability (GCA) estimates tend to produce superior hybrids compared to those with lower GCA (Sandeep et al., 2010).

Various research studies have been performed to investigate whether inbred line performance can be used to predict the performance of derived hybrids (Prado et al., 2013; Flint-Garcia et al., 2009; Zaidi et al., 2007). Although, in other related studies have shown that line *per se* performance is not a good predictor of hybrid performance (Hallaluer and Miranda, 1988), Zaidi et al. (2007) found positive correlation between hybrid performance with inbred performance under moisture stress conditions indicating possibility of predicting hybrid performance using inbred line performance in maize. Similarly, incorporating the pedigree and *per se* performance data in a best linear unbiased prediction of general and specific combining ability enhanced the prediction efficiency of hybrid performance for grain yield and grain dry matter content (Schrag et al., 2010).

Genetic distance based prediction of hybrid performance

Previous studies have shown that there is a positive correlation between genetic distance and performance of hybrids (Bernardo, 1995; Marsan et al., 1998). Positive correlation was reported between genetic distance among maize parents and the degree of heterosis obtained from intercrossing the parents in maize (Smith et al., 1990), sorghum (Jordan et al., 2003; Saghai-Maroof et al., 1997) and rice (Zhang et al., 1994). Likewise, positive correlations for grain yield between marker polymorphism among parents with mid-parent (r = 0.48) and better-parent heterosis (r = 0.65) have been reported in sorghum (Rajendrakumar et al., 2013). Nevertheless, most of the reported correlations have not been strong enough to predict heterosis (Rao et al., 2004; Rani and Rao, 2009) perhaps due to lack of strong association between marker-based heterozygosity estimate and heterozygosity at quantitative trait loci (QTLs) affecting trait of interest (Charcosset et al., 1991) and epistasis (Moll et al., 1965). In sorghum, however, several studies have focused on assessing the

genetic diversity of sorghum germplasm (Mutegi et al., 2011) and association studies for complex traits (Srinivas et al., 2009; Nagaraja Reddy et al., 2013) but no much effort focusing on prediction of heterosis based on genetic distance between elite parental inbred lines.

Genomic prediction of hybrid performance

Success stories about the use of genomic prediction models in animal breeding (Goddard et al., 2010; Habier, 2010) stimulated the plant breeding community to start using genomic prediction models to predict hybrid performance (Piepho, 2009; Lorenza and Bernado, 2009; Zhao et al., 2013). Additionally, recent advances in next-generation technologies such as genotyping-by-sequencing (GBS) that can generate huge amount of genotypic data covering the whole genome at a lower price has further excited the science community to undertake genomic prediction studies to improve selection efficiency. In marker-assisted selection (MAS), only markers with significant and large effects are used to build a prediction model resulting in only a small portion of genetic variance being explained. In contrast, in genomic prediction (GP), all genome-wide markers are used potentially resulting in all the genetic variance being explained by the markers; and also the markers are assumed to be in linkage disequilibrium (LD) with the quantitative trait loci (QTL) so that the number of effects per QTL to be estimated is small. Using the prediction model, the genomic estimated breeding values (GEBVs) are estimated based on the sum of all marker effects (Meuwissen et al., 2001). One of the advantages of genomic prediction is that it can speed up the breeding cycle (Meuwissen et al., 2001).

Furthermore, earlier studies have shown the possibility of predicting hybrid performance using whole-genome prediction (Zhao et al., 2013; Xu et al., 2014). In genomic prediction/ selection for hybrid performance, a training population consisting of genotypes that have been both genotyped (parental genotypes) and phenotyped (hybrid phenotypes) is

used to develop a genomic prediction model that incorporates genotypic data from a candidate population of untested hybrids and produce genomic estimated breeding values (GEBV). To estimate the accuracy of prediction model, true breeding value is correlated with the genomic estimated breeding values (GEBVs) using training and validation sets.

A number of genomic prediction models have been used in predicting hybrid performance in various crops (Technow et al., 2012; Zhao et al., 2013), which include ridge regression best linear unbiased prediction model (RR-BLUP) (Endelman, 2011). Genomic prediction of hybrid performance using RR-BLUP is modeled as: $Y = 1_n\mu + K_A a + K_D d + e$, where 1_n = a vector of ones, and n and μ represent the number of hybrids and the across environment mean, in that order. K_A is the design matrix (n x m) for the additive marker effects, in which m represents the number of markers. The K_A is additively coded as -1, 0 and 1, where "-1" and "1" are representing homozygous genotypic classes A_2A_2 and A_1A_1 and "0" represents heterozygous genotypic class A_1A_2 for each SNP locus. K_D is the design matrix for the dominance marker effects coded as 0, 1, 0 with score "0" standing for homozygous genotypic classes A_2A_2 and A_1A_1 and "1" for the heterozygous genotypic class A_1A_2.The additive and dominance effects of the i^{th} marker are represented as a and d, respectively, in the prediction model while e represents the j^{th}hybrid residual effect.

The additive and dominance marker effects are assumed to be normally distributed, N $(0, \sigma_a^2)$ and N $(0, \sigma_d^2)$ with constant variance of additive effects (σ_a^2) and dominance effects (σ_d^2), respectively. This model predicts the genotypic value of untested hybrids using effects estimated for each marker from the hybrid phenotype (Whittaker et al., 2000; Piepho, 2009). In maize, simulation studies have shown that grain or biomass can be predicted with high accuracy (Albrecht et al., 2011; Gonzalez-Camacho et al., 2012; Riedelsheimer et al., 2012; Zhao et al., 2012) thereby increasing the rate of genetic gain because prediction accuracy of GEBVs is linearly related to the response to selection.

However, a number of factors have shown to affect prediction accuracies including population size of the training set, trait heritability, pedigree information of the genotypes and linkage disequilibrium (LD) between markers with quantitative trait loci (QTLs) associated with the trait, relatedness of genotypes in the training and validation sets and the prediction model used. The higher the training population size, the better the prediction accuracy becomes. Additionally, genotypes from different populations tend to have lower prediction accuracies than those from the same population. For example, a decrease in prediction accuracy as high as 93% has been reported for hybrids in training and validation populations with no common parent shared among them (Gowda et al., 2013a). Again prediction accuracies tend to decrease for genotypes in the training population that are not related with those in the test population (Albrecht et al., 2011). Moreover, when heritability of the trait is low, prediction accuracies tend to be low and vice-versa. Furthermore, different prediction models show variable prediction accuracies for complex traits. Nevertheless, among the models so far tested, RR-BLUP has shown to provide stable and high prediction accuracies for quantitative traits compared to other models (Heslot et al., 2012; Piepho et al., 2012b; Iwata and Jannink, 2011). Additionally, it is very robust and has low computational load, hence it is well suited for genomic prediction of hybrid performance for complex traits (Piepho et al., 2009).

References

Albrecht, T., V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, M. Ouzu- nova, H. Simianer, and C.C. Schon. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet.123:339-350.

Araus, J. L., G. A. Slafer, C. Royo, and M. D. Serret. 2008. Breeding for yield potential and stress adaptation in cereals. Crit. Rev. Plant Sci. 27:377-412.

Bernardo, R. 1995. Relationship between single-cross performance and molecular marker

heterozygosity. Theor. Appl. Genet. 83(5): 628-634.

Birchler, J.A., D. L. Auger, and N.C. Riddle. 2003. In search of the molecular basis of heterosis. Plant Cell. 15:2236-2239.

Charcosset, A., M. Lefort-Buson, and A. Gallais.1991. Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. Theor. Appl. Genet. 81(5): 571-575.

Cheng, S.H., J.Y.Zhuang, Y.Y. Fan, J.H. Du, and L.Y. Cao. 2007. Progress in research and development on hybrid rice: a super-domesticate in China. Ann. Bot. 100: 959-966.

Collins, G. 1921. TEOSINTE IN MEXICO The closest wild relative of Maize is Teosinte-a forage plant hitherto known only as an annual. A perennial form discovered in Southern Mexico should prove of value to the breeder. Heredity. 12:339-350.

Davenport, C.B.1908. Degeneration, albinism and inbreeding. Sci. 28(718): 454-455.

Doggett, H.D. 1969. Yields of hybrid sorghum. Exp. Agric. 5:1.

Duvick, D.N. 1999. Heterosis: feeding people and protecting natural resources. In: Coors JG Pandey S (ed) The genetics and exploitation of heterosis in crops. Crop Sci. pp: 19-29.

Duvick, D. N. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). Maydica 50:193-202.

Duvick, D.N. 2001. Biotechnology in the 1930s: The development of hybrid maize. Nat. Rev. Genet. 2: 69-74.

East, E.M. 1908. Inbreeding in corn. Conn Agric Exp Stn Rep 1907:419-428.

Edwards, M. D., C. W. Stuber, and J. F. Wendel. 1987. Molecular-marker-facilitated investigation of quantitative-trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genet. 116:113-125.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5): p.e19379.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4(3):250-255.

Ertiro, B.T., H, Zeleke, D. Friesen, M. Blummel, and S.T. Afriyie. 2013. Relationship between the performance of parental inbred lines and hybrids for food-feed traits in maize (*Zea mays* L.) in Ethiopia. Field Crops Res. 153:86-93.

Flint-Garcia, S.A., E.S.Buckler, P.Tiffin, E. Ersoz and N. M. Springer. 2009. Heterosis is Prevalent for Multiple Traits in Diverse Maize Germplasm. PLoS ONE 4(10):e7433.

Frei, O.M., C.W. Stuber , M.M. Goodman. 1986. Use of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci. 26:37-42.

Fu, Y.B., B. Cheng, G.W. Peterson. 2014. Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping-by-sequencing. Genet. Resour. Crop Evol. 61:579-594.

Garcia, A.A.F., S. Wang, A.E. Melchinger, and Z.B. Zeng. 2008. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genet.180(3):1707-1724.

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. Cairns et al. 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. Theor. Appl. Genet. 125:759-771.

Goddard, M. E., B. J. Hayes, and T. H. Meuwissen. 2010. Genomic selection in livestock populations. Genet. Res. 92:413-421.

Godshalk, E.B., M. Lee, and K.R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor. Appl. Genet. 80:273-280.

Gowda, M., Y. Zhao, H. P. Maurer, E. A. Weissmann, T. Wurschum, and J. C. Reif. 2013a. Best linear unbiased prediction of triticale hybrid performance. Euphytica 191: 223-230.

Graham, G. I., D. W. Wolff, and C. W. Stuber. 1997. Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci. 37:1601-1610.

Gregorio, G.B., M.R. Islam, G.V. Vergara, and S. Thirumeni. 2013. Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. SABRAO J. Breed. Genet. 45(1): 31-41.

Habier, D. 2010. More than a third of the WCGALP presentations on genomic selection. J. Anim. Breed. Genet.127:336-337.

Hallauer, A.R., M.J. Carena, and J.B. Miranda Filho. 2010. Quantitative Genetics in Maize Breeding, 3rd ed. Handbook of Plant Breeding Volume 6. Springer, New York. 663 pages.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative Genetics in Maize Breeding, 2nd ed. Iowa State University Press, Ames, IA. SB191.M2 H29.

Heffner, E.L., A.J. Lorenz, J. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50:1681-1690.

Heslot, N., H.P. Yang, M. E. Sorrells, and J.L. Jannink. 2012.Genomic selection in plant breeding: a comparison of models. Crop Sci. 52:146-160.

Hinze, L. L., and K.R. Lamkey. 2003. Absence of epistasis for grain yield in elite maize hybrids. Crop Sci. 43(1): 46-56.

Hua, J., Y. Xing, W. Wu, C. Xu, and X. Sun et al. 2003. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA 100: 2574-2579.

Iwata, H., and J.L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 51:1915-1927.

Jan, H.U., A. Abbadi, S. Lücke, R.A. Nichols, and R.J.Snowdon. 2016. Genomic prediction of testcross performance in canola (*Brassica napus*). PLoS One 11(1): e0147769.

Jones, D. F. 1917. Dominance of linked factors as a means of accounting for heterosis. Genet 2: 466-479.

Jordan, D.R., Y. Tao , I.D. Godwin , R.G. Henzell, M. Cooper , and C.L.McIntyre. 2003. Prediction of hybrid performance in grain sorghum using RFLP markers. Theor. Appl. Genet. 106:559-567.

Krieger, U., Z.B. Lippman, and D. Zamir, 2010. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat. Genet. 42(5): 459-463.

Krystkowiak, K., T.Adamski, M. Surma, and Z. Kaczmarek. 2009. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (*Triticum aestivum* L.). Euphytica. 165:419-434.

Lanza, L.L.B., C.L. de Souza Jr, L.M.M. Ottoboni, M.L.C. Vieira, and A.P. De Souza. 1997.Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor. Appl. Genet. 94(8):1023-1030.

Larièpe, A., B. Mangin, S. Jasson, V. Combes, F. Dumas, P. Jamin, C. Lariagon, D. Jolivot, D. Madur, J. Fievet, and A. Gallais. 2012. The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent

overdominance among traits of agronomical interest in maize (*Zea mays* L.). Genet.190(2):795-811.

Li, Z.K., L.J. Luo, and H.W. Mei et al. 2001. Over-dominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genet.158:1737-1753.

Liao, F.M., K.L. Zhou, H.H. Yang, and Q.S. Xu. 1998. Genetic difference of parents and its relation to heterosis in hybrid rice. Chin. J. Rice Sci.12(4):193-199.

Lorenzana, R. E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in bi-parental plant populations. Theor. Appl. Genet. 120:151-161.

Lu, H., J. Romero-Severson, and R. Bernardo. 2003. Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor. Appl. Genet. 107:494-502.

Maroof, M.A., G.P. Yang, Q. Zhang, and K.A. Gravois.1997. Correlation between molecular marker distance and hybrid performance in US southern long grain rice. Crop Sci. 37(1): 145-150.

Marsan, P. A., P. Castiglioni, F. Fusari, M. Kuiper, and M. Motto. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96(2): 219-227.

Martin, J.M., T.E. Talbert, S.P. Lanning, and N.K. Blake. 1995. Hybrid performance in wheat as related to parental diversity. Crop Sci. 35:104-108.

Melchinger, A.E., K.R. Lamkey, and W.L. Woodman. 1990a. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci. 30:1033-1040.

Melchinger, A.E., M. Lee, K.R. Lamkey, A.R. Hallauer , and W.L. Woodman. 1990b. Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor. Appl. Genet. 80:488-496.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genet.157:1819-1829.

Mihaljevic, R., C. C. Schön, H. F. Utz, and A. E. Melchinger. 2005. Correlations and QTL correspondence between line per se and testcross performance for agronomic traits in four populations of European maize. Crop Sci. 45:114-122.

Miller, F.R., and Y. Kebede. 1984. Genetic contributions to yield gains in sorghum, 1950 to 1980. In WR Fehr, ed, Genetic Contributions to Yield in Five Major Crop Plants. Crop Science Society of America Special Publication 7. CSSA, Madison, WI, pp 1-14.

Moll, R.H., J.H. Longquist, J.V. Fortuna, and E.C.Johnson.1965. The relation of heterosis and genetic divergence in maize. Genet.52:139-144.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, S. Kresovich. 2013. Population genomic and genome-wide association studies of agronomic traits in sorghum. PNAS 110(2): 453-458.

Mutegi, E., F. Sagnard, K. Semagn, M. Deu, M. Muraya, B. Kanyenji, S. de Villiers, D. Kiambi, L. Herselman, and M. Labuschagne. 2011. Genetic structure and relationships within and between cultivated and wild sorghum [*Sorghum bicolor* (L.) Moench] in Kenya as revealed by microsatellite markers. Theor. Appl. Genet. 122:989-1004.

Nagaraja Reddy, R., R. Madhusudhana , S. Murali Mohan, D.V. Chakravarthi, S.P. Mehtre, N. Seetharama, and J.V. Patil. 2013. Mapping QTL for grain yield and other

agronomic traits in post-rainy sorghum [*Sorghum bicolor* (L.) Moench]. Theor.Appl. Genet.126:1921-1939.

Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller , and P.A. Hohenlohe . 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22: 2841-2847.

Piepho, H. P. 2009. Ridge regression and extensions for genome-wide selection in maize. Crop Sci. 49:1165-1176.

Piepho, H., J. Ogutu, T. Schulz-Streeck, B. Estaghvirou, A. Gordillo, and F. Technow.2012. Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. Crop Sci.52:1093-1104.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5(3): 92-102.

Prado, S.A., L.G. Brenda, A. D. Novoa, D. Foster, M. L. Senior, C. Zinselmeier, M.E. Otegui, and L. Borrás. 2013. Correlations Between Parental Inbred Lines and Derived Hybrid Performance for Grain Filling Traits in Maize. Crop Sci. 53:1636-1645.

Quinby, J.R., N.W. Kramer, J.C. Stephens, K.A. Lahr, and R.E. Karper. 1958. Grain Sorghum Production in Texas. Tex. Agric. Exp. Stn. Bull. 912.

Rajendrakumar,P., K. Hariprasanna, I. Jaikishan, R. Madhusudhana, and J.V. Patil. 2013. Potential of microsatellite marker polymorphism in the prediction of grain yield heterosis in sorghum [*Sorghum bicolor* (L.) Moench]. In: Rakshit S, Das IK, Shyamprasad G, N. Seetharama, and J.V.Patil.2012. Morphological and molecular diversity reveal wide variability among sorghum Maldandi landraces from India. J. Plant Biochem. Biotechnol. 21:145-156.

Rani, K.J., and S.S. Rao. 2009. Relationship between heterosis and genetic divergence in rabi sorghum [*Sorghum bicolor* (L.) Moench]. Res.Crops 10:319-322.

Rao, M.G., I. Reddy, R.S. Kulkarni, S. Ramesh, and S.L. Reddy. 2004. Prediction of heterosis based on genetic divergence of parents through regression analysis in sunflower (*Helianthus annuus* L.). Helia 27(41): 51-58.

Reif, J. C., A. R. Hallauer, and A. E. Melchinger. 2005. Heterosis and heterotic pattern in maize. Maydica 50:215-223.

Reif, J.C., A.E. Melchinger, X.C. Xia, M.L. Warburton, D.A. Hoisington, S.K. Vasal et al. 2003. Use of SSRs for establishing heterotic groups in subtropical maize. Theor. Appl. Genet.107: 947-957.

Riedelsheimer, C., F. Technow, and A. E. Melchinger. 2012. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics 13(1): 452.

Robinson, H.F., R.E. Comstock, A. Khalil, and P.H. Harvey. 1956. Dominance versus over-dominance in heterosis: evidence from crosses between open-pollinated varieties of maize. American Naturalist, pp.127-131.

Sadras, V.O., and G.A. Slafer. 2012. Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. Field Crops Res. 127:215-224.

Sandeep, R.G., M.R. Gururaja Rao, S. Ramesh, H. Chikkalingaiah, and Shivanna. 2010. Parental combining ability as a good predictor of productive crosses in sweet sorghum [*Sorghum bicolor* (L.) Moench]. J.Appl. Nat. Sci. 2:245-250.

Schnell, F., and C. Cockerham. 1992. Multiplicative vs. arbitrary gene action in heterosis. Genet. 131:461-469.

Schrag, T.A. J. Mohring, A.E. Melchinger, B. Kusterer, B.S. Dhillon, H.P. Piepho, M. Frisch. 2010. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor. Appl. Genet.120: 451-461.

Shull, G.H. 1908. The composition of a field of maize. Am Breeders Assoc. Rep. 4:296-301.

Shull, G.H. 1946. Hybrid seed corn. Sci. 103:547-550.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J.Wall.1990. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield heterosis and RFLPs. Theor. Appl. Genet. 80:833-840.

Srinivas, G., K. Satish, R. Madhusudhana, R. Nagaraja Reddy, S. Murali Mohan, and N.Seetharama.2009. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor. Appl. Genet. 118:1439-1454.

Stuber, C. W., S. Lincoln, D.W. Wolff, T. Helentjaris, and E.S. Lander. 1992. Genet. 132: 823-839.

Technow, F., C. Riedelsheimer, T. A. Schrag, and A. E. Melchinger.2012.Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. 125:1181-1194.

Tuberosa, R., S. Salvi, S. Giuliani, M. C. Sanguineti, and M. Bellotti et al. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47: 120-141.

Whittaker, J. C., R. Thompson, and M. C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res.75:249-252.

Windhausen, V. S., G. N. Atlin, J. Crossa, J. M. Hickey, and P. Grudloyma et al. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda) 2:1427-1436.

Wolf, D.P., and R. Hallauer. 1997. Triple testcross analysis to detect epistasis in maize. Crop Sci. 37(3): 763-770.

Xiao, J., J. Li, L. Yuan, S.R. McCouch, and S.D. Tanksley. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor. Appl. Genet. 92:637-643.

Xu, S., D. Zhu, and Q. Zhang. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. 111(34):12456-12461.

Yu, S.B., J.X. Li, C.G. Xu, Y.F. Tan, Y.J.Gao, X.H. Li, Q. Zhang, and M.S. Maroof, 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proceed. Natl. Acad. Sci. 94(17): 9226-9231.

Zaidi, P.H., P. Mani Selvan, R. Sultana, A. Srivastava, A.K. Singh, G. Srinivasan G, R.P. Singh, and P.P.Singh. 2007. Association between line per se and hybrid performance under excessive soil moisture stress in tropical maize (*Zea mays* L.). Field Crops Res. 101:117-126.

Zhang, Q.F,Y.J. Gao, S.H. Yang, R.A. Ragab, M.A. Saghai Maroof, and Z.B. Li. 1994. A half-diallel analysis of heterosis in elite hybrid rice based on RFLP and microsatellites. Theor. Appl. Genet. 89:185-192.

Zhang, T., L. Han, J.D. Xu, K.F. Jiang, X.J. Wu, X.D. Wang, and J.K. Zheng. 2006. Correlation between genetic distance and yield heterosis of hybrid aromatic rice. Sci. Agric. Sin 39(4): 831-835.

Zhang, Q.F., Y.J. Gao, M.A.S. Maroof, S.H. Yang and J.X. Li. 1995. Molecular divergence and hybrid performance in rice. Mol. Breed.1:133-142.

Zhang, T., X.L. Ni, K.F. Jiang, H.F. Deng, H.E. Qing, Q.H. Yang, Y.A.N.G. Li, W.A.N. Xian-Qi, Y.J. Cao, and J.K. Zheng. 2010. Relationship between heterosis and parental genetic distance based on molecular markers for functional genes related to yield traits in rice. Rice Sci. 17(4): 288-295.

Zhao, Q.Y., Z. Zhu, Y.D. Zhang, L. Zhao, T. Chen, O.F. Zhang, and W.L Wang. 2008. Correlation analysis between genetic distance of SSR markers and heterosis japonica. The Fifth National Congress of Plant Molecular Breeding-cum-academic exchanges Proceedings.

Zhao, Y., J. Zeng, R.Fernando, and J.C. Reif. 2013. Genomic prediction of hybrid wheat performance. Crop Sci. 53(3): 802-810.

Zhu, Y.L, Q.J. Song, D.L. Hyten , C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, et al. 2003. Single- Nucleotide Polymorphisms in Soybean. Genet.163:1123-1134.

Chapter 2 - Genome-Wide Analysis of Genetic Diversity, Population Structure and Linkage Disequilibrium in Public Sorghum [*Sorghum bicolor* (L.) Moench] Inbred Lines.

Abstract

The next-generation sequencing technologies have provided excellent opportunity for in depth investigation of the genome structure and genetic diversity, as well as for mapping of complex traits in crop species at low cost. These are essential ingredients for designing and implementing effective crop improvement program. The objective of this study was to assess the levels of genetic diversity and characterize population structure and linkage disequilibrium in public sorghum inbred lines. A total of 279 public sorghum inbred lines developed at Kansas State University comprising 228 R-lines and 51 B-lines were included in this study. Having developed at different times over the last two decades, the materials harbor robust diversity in pedigree and agronomic characteristics with some materials having resistance to Acetyl co-enzyme-A carboxylase (*ACCase*) and Acetolactate synthase (*ALS*) inhibitor herbicides. The inbred lines were genotyped using genotyping-by-sequencing (GBS) platform and 282,536 SNPs with ≥ 1% minor allele frequency (MAF) and < 20% missing data were generated. After filtering for MAF, a total of 66,265 SNPs with ≥ 5% MAF and < 20% missing data were returned and used for analysis. The mean polymorphic information content (PIC) and gene diversity across the entire inbreds were 0.35 and 0.46, respectively. The non-herbicide resistant inbreds harbored more diversity than those resistant to herbicides as revealed by PIC and gene diversity values. The neighbor-joining tree, principal component and STRUCTURE analyses clustered the inbred lines into three subgroups according to pedigree and fertility-reaction. Thus inbred lines derived from closely related pedigrees tended to cluster together; also the maintainer lines (B-lines) were distinctly

grouped to form a separate cluster. Analysis of molecular variance revealed that variation within subgroups was much higher than that among subgroups. Substantial linkage disequilibrium was detected between the markers in the population with marked variation between chromosomes. This information may facilitate the use of the inbreds in sorghum breeding programs and provide perspectives for optimizing marker density for gene mapping and marker-assisted breeding.

Key words: SNP, Single nucleotide polymorphisms, PIC, polymorphism information content; Neighbor-joining tree; Linkage disequilibrium; MAF, Minor allele frequency; Analysis of molecular variance; GBS, Genotyping-by-sequencing.

Introduction

The availability of low cost next-generation sequencing technologies such as genotyping-by-sequencing (GBS) has provided a great opportunity for deeper investigation of genome structure and genetic diversity of crop species. These in turn will generate valuable information for understanding of the genetics of key plant traits and facilitate their improvement through breeding. Molecular markers have been widely utilized in plant breeding and genetics for gene mapping, genetic analysis and other biological applications (Semagn et al., 2012; Satish et al., 2012). Knowledge of genetic diversity/similarity and structure of working germplasm is very important for selecting potential parental sources for initiating breeding populations (Benchimol et al., 2000; Lu et al., 2009; Reif et al., 2003).

Over years, a number of molecular markers which include amplified fragment length polymorphisms (AFLP), restriction fragment length polymorphism (RFLP), single nucleotide polymorphisms (SNPs), random amplified polymorphic DNA (RAPD) and simple sequence repeats (SSR) have been developed to serve these purposes. Each time a new tool came up it added value or brought more power either through adding accuracy of marker genotyping and scoring, reducing cost and time needed to run the assays or through increasing robustness such as increasing the number of alleles that can be evaluated at a time. In this regard single nucleotide polymorphic markers are the latest genotyping tools with much improved power due to their abundance in the genome and suitability for analysis on a wide range of genomic scales (Rafalski, 2002; Zhu et al., 2003). Additionally, SNPs have become very popular because of their low cost per data point, locus specificity, codominance and potential for high throughput analysis (Schlotterer, 2004; Chagne et al., 2007; Rafalski, 2002). In addition, they are applied in a broad range of genetic studies including genetic diversity and population structure, mapping of quantitative trait loci (QTL), genome-wide association mapping (GWAS) studies and marker-assisted breeding and genomic selection.

Genetic diversity is an important ingredient of a successful plant breeding program. Genetic gains in recurrent selection are highly dependent, among others, on the extent of genetic variability not only in the base population but also of the population at each selection cycle. In crops where hybrid technology is commercially exploited, the performance of a hybrid cultivar is highly dependent on the degree of genetic variability between its parental inbreds (Ganapathy et al., 2012). Therefore, knowledge of the extent of genetic variability in the breeding materials and integration of the information in the development of the cultivars are very vital (Chao et al., 2007). Furthermore, success in genetic studies such as QTL mapping largely depends on understanding of the genetic diversity and genetic structure of the population under study.

Sorghum is one of the most diverse crop species expressing an array of morphological and agronomic features as well as broad genomic variability (Zheng et al., 2011; Mace et al., 2013; Morris et al., 2013). The USDA National Plant Genetic Resource Center alone is housing over 40,000 sorghum germplasm accessions (GRIN). Several international and national institutions in the developed and developing world maintain large number of sorghum germplasm collections. However, the photoperiod sensitivity of the crop arising from its tropical adaptation highly undermined the use of the germplasm that despite its abundance, only limited sources are being used in temperate sorghum breeding programs. An effort to circumvent this through converting selected materials to photoperiod insensitivity was a wise approach but after nearly 20 years of efforts only several hundred materials were converted for use in temperate breeding programs (Klein et al., 2016). These and photoperiod insensitive collections and early cultivars and germplasm developed over the last few decade collectively form the basis of sorghum germplasm currently utilized in temperate breeding programs. While efforts are underway to bring more germplasm into use such as through the reintroduction of the sorghum conversion program (USDA, 2009) and increased partnership,

it is also wise to study and document the genetic information of the existing public inbred lines currently utilized as key source of germplasm by both public and private breeding programs. The newly developed genomic platforms and bioinformatics tools present excellent opportunity to elucidate these in a way never possible before and help chart ways for the best use of these resources.

The objective of this study was to assess the genetic diversity, population structure and linkage disequilibrium (LD) in public sorghum inbred lines genotyped using genome-wide SNP markers generated using genotyping-by-sequencing (GBS) platform.

Materials and methods

Genetic materials

A total of 279 public sorghum inbred lines were used in this study. The collection included 228 pollinator lines and 51 seed parent lines developed at KSU and Texas A&M sorghum breeding programs. The lines represent diverse pedigrees and have diverse morphological and agronomic characteristics with some 52 of them having resistance to Acetyl co-enzyme-A carboxylase (*ACCase*) inhibitor herbicides and 89 of them to Acetolactate Synthase (*ALS*) inhibitor herbicides. The remaining 138 lines have been released over several years and may carry traits for drought tolerance, disease and insect resistance.

Genomic DNA extraction and genotyping-by-sequencing (GBS)

Sorghum seeds of the parental inbred lines were planted in the greenhouse at Kansas State University (KSU) using 96-cell flat trays filled with Metro-mix 360 (Sun Gro) growing medium. Ten to fourteen days after planting, young and fresh leaf tissues were harvested from a single seedling of each inbred line for genomic DNA extraction using the cetyltrimethylammonium bromide (CTAB) method (Doyle, 1987). The Quant-iT PicoGreen *ds*DNA Assay Kit (Invitrogen) was used to quantify the concentrations of the DNA samples. SNP genotyping was performed using genotyping-by-sequencing (GBS) at the Institute of

Genomic Diversity, Cornell University. GBS procedure identified 282,536 SNPs with minor allele frequency (MAF) ≥ 0.01 and < 20% missing data. A set of 66,265 SNPs was used for the analysis after filtering SNPs with MAF < 5% using PLINK v1.07 (Purcell et al., 2007)

Statistical analysis

Genetic diversity and familial relatedness

The mean gene diversity and polymorphic information content (PIC) (Botstein, 1980) were calculated using PowerMarker v 3.25 (Liu and Muse, 2005). The Nei's genetic distance (Nei, 1972) between inbred lines was calculated in R program (R Development Core Team) of *Adegenet* package. Relative kinship values for all pairwise comparisons among the lines were calculated based on 66,265 SNPs with MAF ≥ 0.05 and < 20 % missing data using TASSEL 5.2.14 (Bradbury et al., 2007) and the "scaled Identity by state (IBS)" (Endelman and Jannink, 2012). The genotypes were coded as 0, 1 and 2, equal to the count of one of the alleles at that locus and the missing genotype values are replaced with the average genotypic score at that locus before a relationship matrix is estimated. All negative relative kinship values between two individuals indicating less relationship between two random individuals were changed to zero.

Population structure, neighbor-joining tree and principal component analyses

Population structure, neighbor-joining (NJ) tree and principal component analyses (PCA) were performed to determine the genetic grouping of this collection and also to assess whether there is clear genetic clustering between B-lines (male-sterile maintainer lines) and R-lines (fertility-restorer lines). Population structure analysis was performed using the STRUCTURE program v 2.3.4 with the admixture model (Falush et al., 2012). The burn-in period of 30,000 iterations and a run of 30,000 replications of Markov Chain Monte Carlo after burn in were used. For each run, 5 independent runs of the STRUCTURE were performed with an assumed number of subpopulations (*k*) varying from 1 to 10, leading to 50

structure outputs. The inbred lines with membership coefficients of greater than 80% were assigned to a subgroup and those with less than this threshold were assigned to a mixed subgroup. The genetic distance matrix was generated in TASSEL 5.2.14 (Bradbury et al., 2007) using neighbor-joining clustering method and the NJ tree was visualized using Molecular Evolutionary Genetic Analysis (MEGA 5.10) (Tamura et al., 2011)

PCA was performed using package *prcomp* (Becker et al., 1988) in R program. The number of principal components to use in clustering the inbred lines was determined by generating a scree plot in which the proportion (eigenvalues) of each principal component (PC)'s contribution to total variation was plotted against the number of PCs. An "*elbow*" point on the scree plot denotes the appropriate number of PCs to use (Figure A.1). The 3D scatter plot was produced using *rgl* package (Murdoch et al., 2013) in R program.

Analysis of molecular variance (AMOVA) and population differentiation

Analysis of molecular variance (AMOVA) (Weir, 1996) and population differentiation (F_{st}) were calculated using GenAlex 6.502 (Excoffier et al., 2005) to partition variation among and within subgroups. For AMOVA and population differentiation (F_{st}), the inbred lines were assigned to different subgroups based on the results from the NJ, population structure, PCA and a priori (pre-determined) subgroups based on pedigree information (i.e. *ACCase*, *ALS* and regular inbred lines).

Linkage disequilibrium (LD) analysis

Linkage disequilibrium (LD), a squared allele frequency correlation coefficient (r^2) between loci was estimated in TASSEL 5.2.14 (http://www.maizegenetics.net) using a sliding window of 50. LD was estimated between all SNPs on the whole genome and each chromosome separately. LD estimates were also calculated for different subgroups identified by the STRUCTURE program and pedigree-based subgroups. One set of 50 inbred lines was randomly selected from each subgroup to avoid bias that differences in sample sizes of the

different subgroups may cause. All SNP pairs with *p*-values less than 0.05 were considered to be in significant LD. LD decay distance curves were estimated on the whole genome basis and each chromosome separately using curvilinear regression in SPSS software version 22.0 (IBM Corp. 2013).

Results

Genotyping-by-sequencing (GBS)

Genotyping-by-sequencing generated 282, 536 SNPs distributed across the genome with MAF ≥ 0.01 and < 20% missing data. And more than half of all SNPs scored were rare (MAF < 0.05) (Table 2.1). The number of SNPs per chromosome ranged from 15,867 on chromosome 8 to 46,338 on chromosome 1. After SNPs with minor allele frequency of (MAF) < 5% were filtered, the total number of SNPs were reduced to 66, 265 with number of SNPs per chromosome subsequently reducing ranging from 3950 on chromosome 7 to 10,189 on chromosome 1. A further filtering of SNPs with less than 10% MAF reduced the total number of SNPs to 42, 661, with the number of SNPs reducing correspondingly with lowest number (2379) of SNPs again occurring on chromosome 7 and the highest (6235) on chromosome 1 (Table 2.1). Thus the rare alleles appear to have occurred on all chromosomes and filtration of rare alleles at different MAF thresholds removed SNPs from all chromosomes such that the ranking of the number of SNPs per chromosome after filtrations rarely changed (Table 2.1). Nevertheless, the analysis was performed on 66, 265 SNPs after 5% filtration.

Table 2.1 Summary of single nucleotide polymorphisms (SNPs) identified by genotyping-by-sequencing (GBS) presented by chromosomes at three different minor allele frequency (MAF) thresholds.

Chromosome	Number of SNPs after filtration at MAF			Rank of # of alleles per locus after filtration at MAF		
	≥ 0.01	≥ 0.05	≥ 0.1	MAF ≥ 0.01	MAF ≥ 0.05	MAF ≥ 0.1
1	46338	10189	6235	1	1	1
2	36420	8946	5924	3	2	2
3	39736	8798	5447	2	3	3
4	31678	7162	4636	4	4	4
5	18086	5454	3682	9	7	7
6	25786	6724	4441	5	5	5
7	19675	3950	2379	8	10	10
8	15867	4388	3093	10	9	8
9	22800	4965	3045	7	8	9
10	24150	5689	3779	6	6	6
Total	282536	66265	42661			

Genetic diversity, familial relatedness and genetic distance analyses

The genetic diversity and population structure analysis based on 66, 265 SNPs grouped the inbreds into three subgroups, G1, G2 and G3. The gene diversity, PIC and mean relative kinship values across and within subgroups of inbreds are presented in Table 2.2. The mean gene diversity and PIC values across the entire set of inbreds were 0.46 and 0.35, respectively. Among the model-based subgroups, mean gene diversity was different with G2 having the highest diversity (0.42) and G3 the lowest (0.37). G1 had mean gene diversity of 0.40. G1 had the lowest PIC value (0.33) followed by G2 (0.35) and G3 (0.38). For the pedigree-based subgroups, the *ACCase* herbicide resistant subgroup had the lowest mean gene diversity (0.39) and PIC value (0.31) followed by the *ALS* herbicide resistant subgroup with gene diversity of 0.41 and PIC of 0.32. The regular lines subgroup had the highest mean gene diversity (0.44) and PIC value (0.34) among pedigree-based subgroups. These explain the degree of past research in each of these germplasm categories where breeding of regular parental lines brought more diversity to the subgroup compared to the *ALS* and *ACCase* herbicide resistance research that are relatively a recent initiative.

Relative kinship values between pairs of individuals provide valuable information for quantitative genetic studies and reflect the approximate similarity between two given individuals over the average probability of identity between two random individuals. Across the inbred lines, mean relative kinship value between inbred lines was 0.061, ranging from 0 to 1.46. Out of all inbred lines more than 55% of the pairwise comparisons had relative kinship values of less than 0.5, indicating that majority of the lines are unrelated (Table 2.2 and Figure 2.1). The mean relative kinship values between inbred lines for both model-based clustering and pedigree-based subgroups also varied (Table 2.2). The mean relative kinship value for G1 was

0.076 ranging from 0 to 1.79. The mean relative kinship values were 0.082 and 0.23 for G2 and G3, respectively. For pedigree-based subgroups, the regular inbred subgroup had the highest mean relative kinship value (0.10), followed by *ACCase* and the *ALS* herbicide resistant subgroups with 0.084 and 0.083, respectively. The pairwise relative kinship values between lines ranged from 0 to 1.71 for *ACCase* herbicide resistant subgroup, from 0 to 2.0 for *ALS* herbicide resistant subgroup and from 0 to 1.65 for regular inbred lines. In addition, approximately 43, 47 and 49% of the pairwise relative kinship values for *ACCase*, *ALS* herbicide resistant subgroups and regular lines, in that order, were less than 0.5 indicating no relationship between these lines (Table 2.2).

The genetic distance between pairwise comparisons of all 279 inbred lines ranged from 0.10 to 0.70 and overall average distance was 0.49; the vast majority (80%) of the pairs of lines ranged from 0.41 to 0.70 (Figure 2.2). Among the R-lines, about 61% of them had relative kinship values of less than 0.5, with the mean of 0.29. The relative kinship values ranged from 0 to 1.55. On average the genetic distance among the R-lines was 0.30, ranging from 0.05 to 0.68, with about 58% of them having their pairwise genetic distance of less or equal to 0.5. Among the B-lines, the genetic distance ranged from 0.06 to 0.67 with 54% of them having the genetic distance of less than 0.5. In addition, the average relative kinship value was 0.08, ranging from 0 to 1.45.

Table 2.2 Genetic diversity and mean relative kinship within subgroups identified based on STRUCTURE analysis and pedigree-classification.

Model-based subgroups					
Subgroup	Sample size	Gene diversity	PIC	Mean of relative kinship	Kinship values < 0.5 (%)
G1	76	0.40	0.33	0.076	45.8
G2	113	0.42	0.35	0.082	43.8
G3	90	0.37	0.38	0.23	51.3
Pedigree-based subgroups					
ACC R-Lines	52	0.39	0.31	0.084	43.4
ALS R-Lines	89	0.41	0.32	0.083	46.8
Regular lines	138	0.44	0.34	0.10	49.4
Total	279	0.46	0.35	0.061	55.0

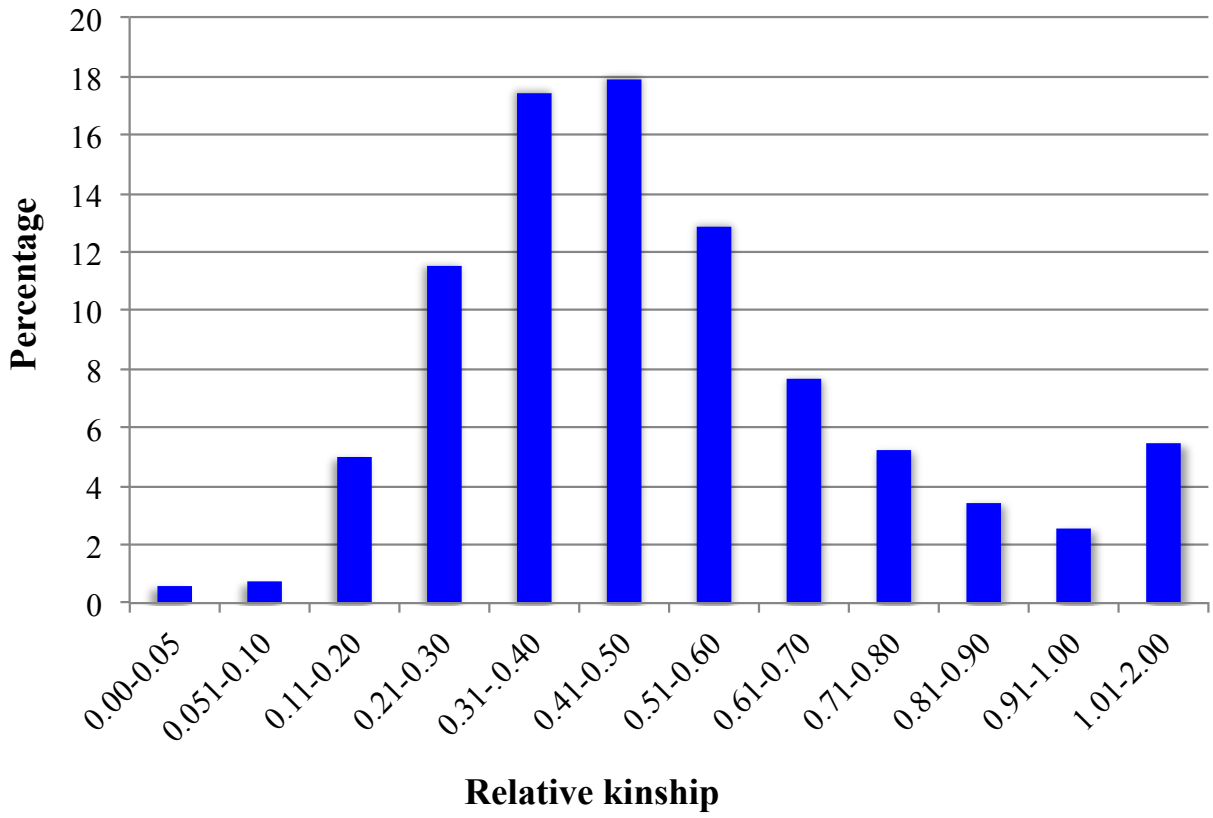


Figure 2.1 Distribution of pairwise relative kinship values in percentages for 279 sorghum public inbred lines genotyped using 66, 265 SNPs. Relative kinship values close to 0 indicate no relationship.

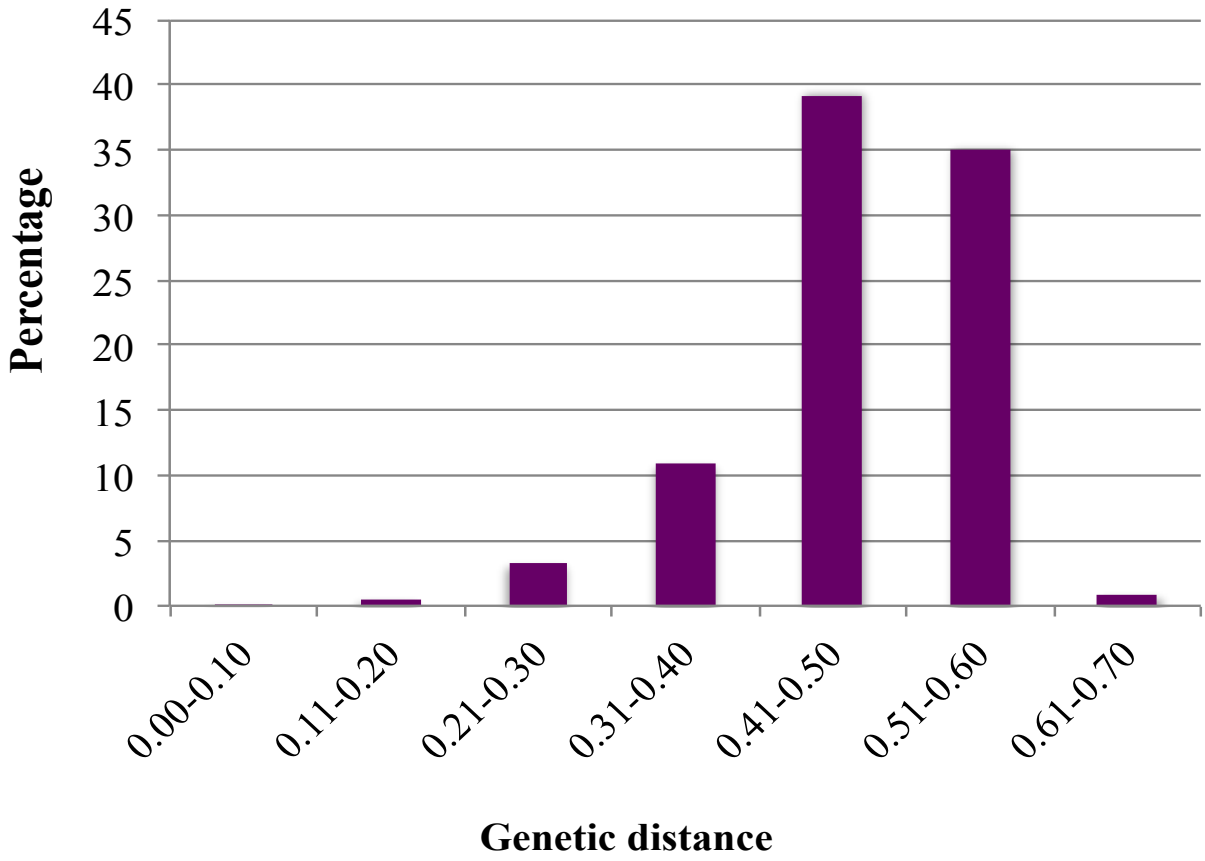


Figure 2.2 Distribution of pairwise genetic distance between 279 public sorghum inbred lines genotyped using 66,265 SNPs.

Neighbor-joining tree, population structure and principal component analyses

The neighbor-joining tree analysis grouped the 279 inbred lines into three major subgroups and with several minor clusters within each major subgroup (Figure 2.3). Of the total 65 inbreds in Subgroup 1, 47 were B-lines, 37 were *ACCase* herbicide resistant and 7 were *ALS* herbicide resistant and the remaining were regular pollinator lines. Subgroup 2 consisted of diverse inbred lines including 70 *ALS*, and 5 *ACCase* herbicide resistant lines, 15 regular R- and 4 B-lines. Subgroup 3 comprised 68 regular lines and 9 *ALS* herbicide resistant lines (Figure 2.3). Most of the inbred lines with the common parents in their pedigree tended to cluster in the same subgroup. Also, majority of the B-lines (maintainers of male sterility) were clustered in one minor subgroup and were clearly separated from the R-lines (fertility-restorer lines). This may be the result of many years of separate breeding activity for B-and R-lines.

Genome-wide SNP marker data were subjected to population structure analysis using the STRUCTURE program which also grouped the inbred lines into three major subgroups (Figure 2.4). Again, PCA was done on SNPs that were converted to numerical values. The scree plot was generated whereby the eigenvalues of an individual PC's contribution to the total variation was plotted against the number of PCs (Figure A.1). The scree plot showed that the "elbow" point occurred at 3 implying that three principal components are required to cluster the inbred lines while each subgroup at 0.8 membership coefficient cutoff level has left several lines as admixture. The first three principal components explained 11.8, 10.08 and 5.13% of the overall genetic variation explained by the marker data (Figure 2.5). The three

PCs clustered the parental inbred lines into subgroups generally based on pedigree information and fertility restoration capacity (Figure 2.5).

However, there were some subtle differences among the neighbor-joining tree, population structure and PCA analyses. The extent of genetic differentiation appears to be sharper for neighbor-joining tree than the population structure and principal component analyses. Some lines that were clustered tightly in the neighbor-joining tree appeared to be loosely clustered in the STRUCTURE analysis (Figures 2.3 and 2.4). But this has to do with the nature of the analyses. However, the subgrouping by the two methods were generally in agreement with each other in that the same group of lines were pooled into the same cluster in both methods with minor expected discrepancies. Similar to the NJ tree and STRUCTURE analyses, the principal component analysis also separated the inbred lines from one another with most of *ALS*, *ACCase* herbicide resistant and the regular B-lines separated from the rest of the inbreds.

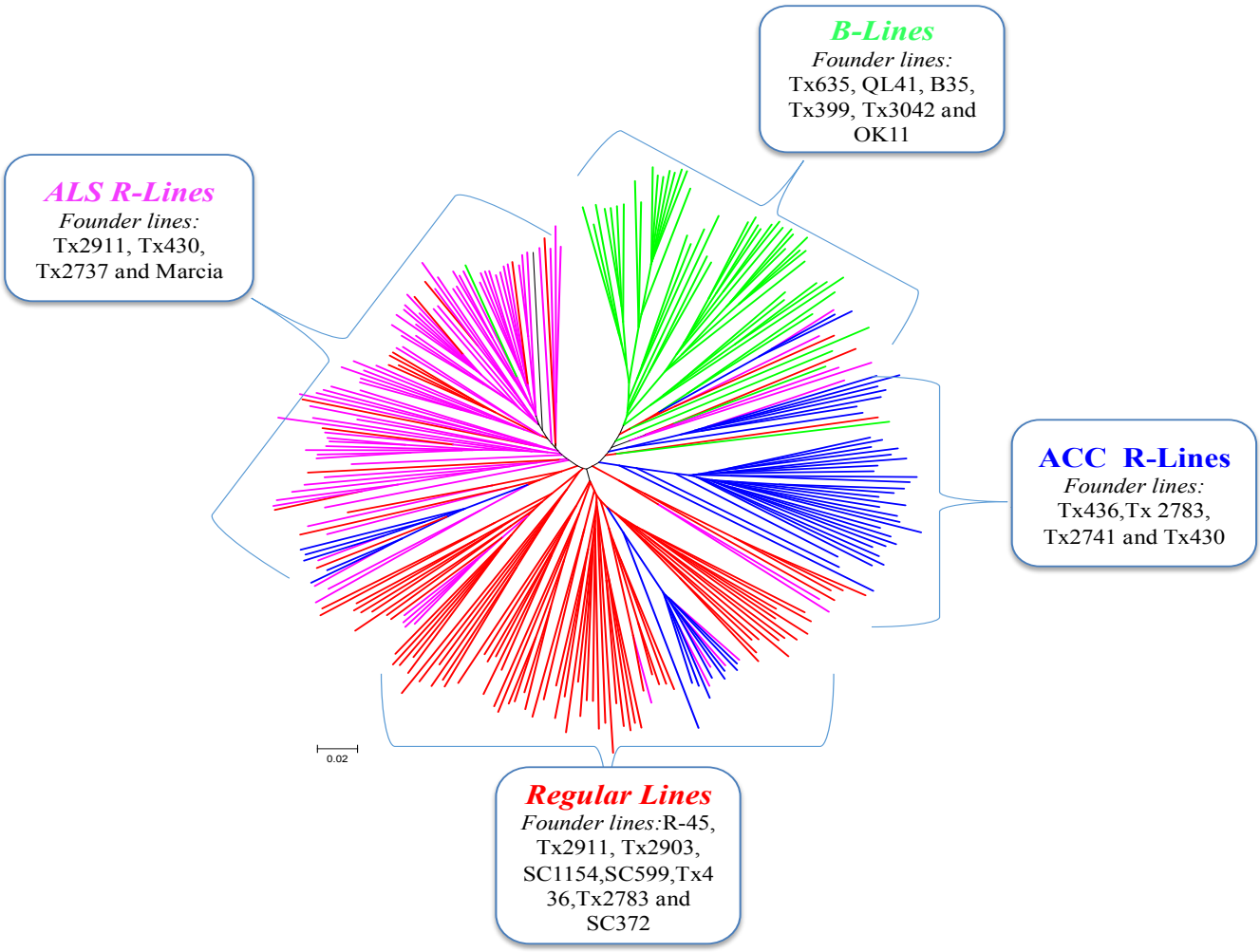


Figure 2.3 Genetic relationship among 279 public sorghum inbred lines assessed by the neighbor-joining tree method. The branches are color-coded based on pedigree information and fertility restoration capacity (B vs. R-Lines).

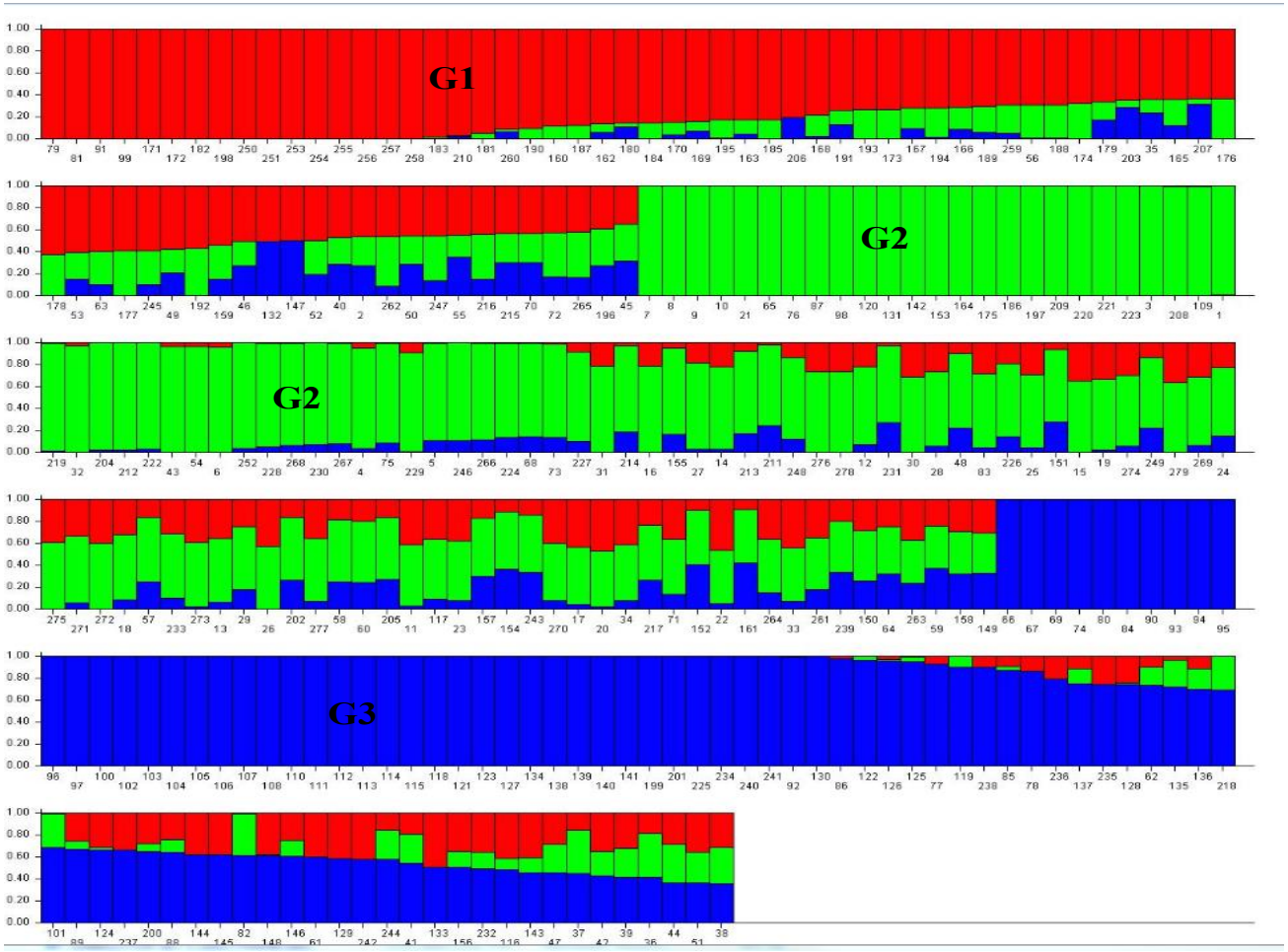


Figure 2.4 Population structure analysis results. Numbers on the *y*-axis show subgroup memberships. G1, G2 and G3 are subgroups identified by STRUCTURE program (G1= red, G2=green and G3=blue).

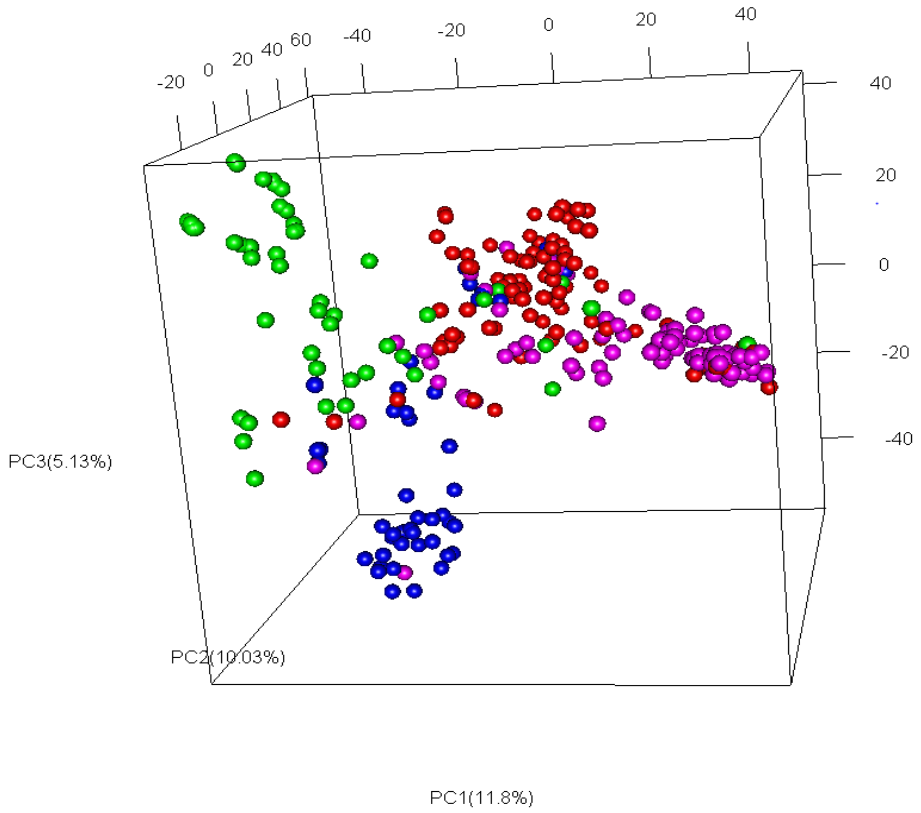


Figure 2.5 Principal component analysis (PCA) results of 279 public sorghum inbred lines based on 66,265 SNP data. The colors for the subgroups are similar to the neighbor-joining tree results presented in Figure 2.3. Red = Regular Lines; Blue= *ACCase* herbicide resistant lines; Green = B-Lines and Pink= *ALS* herbicide resistant lines.

Analysis of molecular variance (AMOVA) and population differentiation

The AMOVA revealed that for the model-based classification, 19% and 81% of the molecular variation were found among subgroups and within subgroups, respectively (Table 2.3). Similarly, for the pedigree-based classification, 27% of the total genetic variation was accounted for by among subgroups and 73% within subgroups (Table 2.3). Pairwise genetic distance between the subgroups was tested using pairwise *F*-statistics. For the model-based subgroups, the *F*-statistics value was the largest (*Fst* = 0.42) between G1 and G2 (Table 2.4), while it was the smallest (*Fst* = 0.31) between G1 and G3. The *F*-statistics value for the pedigree-based subgroups was the largest (*Fst* = 0.39) between subgroups 2 and 3, while it was the lowest (*Fst* = 0.23) between subgroups 1 and 2 (Table 2.4).

Table 2.3 Analysis of molecular variation of three subgroups classified by different methods

Model-based subgroups				
Source of variation	*df*	Sum of squares	Variance components	Percent variation
Among subgroups	2	147547.43	867.06	19
Within subgroups	276	860261.23	3116.89	81
Total	278	1007808.66	4451.49	100
Pedigree-based subgroups				
Among subgroups	2	76155.36	1039.33	27
Within subgroups	276	935791.19	2810.06	73
Total	278	1011946.56	3849.4	100

Table 2.4 Pairwise *F*-statistics value between each pair of the three subgroups identified based on STRUCTURE analysis (above diagonal) and pedigree information (below diagonal).

Subgroup	G1	G2	G3
G1	-	0.42	0.31
G2	0.23	-	0.32
G3	0.35	0.39	-

G1, G2 and G3 are subgroups.

Linkage disequilibrium (LD) analysis

The linkage disequilibrium (LD) results are presented in Table 2.5. On average the LD (r^2) in the entire set of inbred lines was 0.27 with 79.8% of SNP marker pairs in significant LD ($p < 0.05$). Among the model-based subgroups, the mean LD (r^2) ranged from 0.37 for G2 to 0.42 for G3 with 67.8 and 59.7% of marker pairs in significant LD ($p < 0.05$), respectively (Table 2.5). On the other hand, the average LD among pedigree-based subgroups ranged from, $r^2 = 0.31$ for regular inbred lines to $r^2 = 0.37$ for ACC R-inbred lines. Again ALS R-inbred lines had the highest percentage number of SNP marker pairs (43.4) with LD estimates (r^2) greater than 0.35, followed by ACC R-inbred lines (42.3). The regular inbred lines had the lowest (31.5%). Additionally, ACC R-lines had the highest percentage number of LD estimates for SNP marker pairs with $p< 0.0001$, while regular inbred lines had the lowest (Table 2.5). Linkage disequilibrium decayed with increasing physical distance within and across chromosomes and the LD pattern varied across the chromosomes (Figure 2.6 and Figures B.1-B.5)

Table 2.5 Linkage disequilibrium as measured by r^2 and its p-value for across and within subgroups of sorghum public inbred lines.

Model-based subgroups							
Subgroup	Sample size	No. of SNP pairs	$r^2 > 0.35$ (%)	p-values < 0.0001, %	Mean (r^2)	Average p	Significant SNP pairs (%)
G1	50	478786	44.5	38.2	0.41(±0.38)	0.19	72.3
G2	50	499685	41.3	36.9	0.37(±0.33)	0.21	67.8
G3	50	534983	42.7	36.5	0.42(±0.42)	0.17	59.7
Pedigree-based subgroups							
ACC R-Lines	50	495485	42.3	38.9	0.37(±0.33)	0.15	69.1
ALS R-Lines	50	533593	43.4	37.2	0.36(±0.32)	0.16	65.3
Regular Lines	50	530856	31.5	28.7	0.31(±0.31)	0.18	57.9
Total	279	639159	29	64.4	0.27(±0.25)	0.12	79.8

Numbers in parentheses are standard deviations

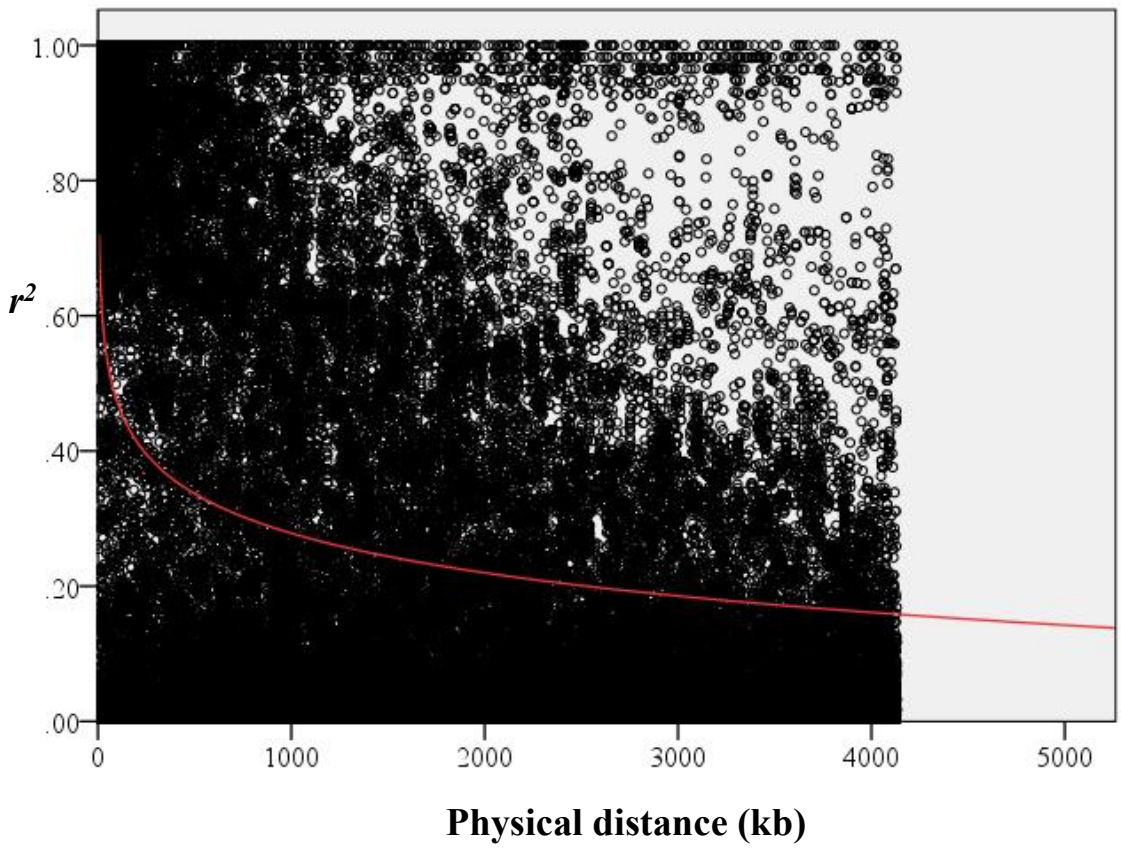


Figure 2.6 Scatter plot of genome-wide linkage disequilibrium (r^2) against physical distance (kb) and estimated genome-wide LD decay curve.

Discussion

The information on genetic diversity, population structure, and LD in this collection should provide important information for developing new breeding strategies for utilizing the germplasm for sorghum improvement. Compared with earlier reports, the genetic diversity discovered across the 279 public sorghum inbred lines was high as reflected in gene diversity value (0.49). The mean gene diversity and PIC detected in this study both within subgroups and across the entire set may not be considered large compared to what can be observable from landrace collections. But given that the materials are elite breeding lines subjected to intensive selection primarily for improved agronomic traits, the level of diversity observed was significant. This may be the result of the long history of the breeding program introducing new germplasm materials into breeding populations which is a key for maintaining robust genetic variation as observed in this study. Similar results have been reported in maize (Wu et al., 2016) where genetic variability based on GBS generated SNPs among inbred lines from CIMMYT and other regions of the world were found to be robust with gene diversity and PIC values of 0.31 and 0.25, respectively.

Moreover, the STRUCTURE program, PCA and NJ tree analyses clustered the populations according to pedigree information where lines sharing large proportion of common pedigree were apparently forming strong clusters. Accordingly, the markers distinctly separated the B-lines from the R-lines. This is most likely due to the difference in the pedigree of the materials and does not necessarily indicate that B- and R- germplasm groups are distinctly “heterotic” given that the B/R designation is the result of one or few nuclear genes (Cui et al., 1996; Klein et al., 2001). Seed parent (B) breeding materials in most hybrid sorghum

breeding programs come from the kafir sorghum working group (Harvey, 1977) while the pollinator parents tend to be durra, caudatum and other races. Such result was not unexpected because over time, sorghum seed and pollinator parents have been developed using these distinct gene pools resulting in most of the B-lines sharing the same genetic background and thus genetically departing from the R-lines. Because the B- and R- designations are based on few nuclear genes, this research finding could be different if the study was performed on original germplasm population not subjected to genetic improvement. Perhaps for the same reason, previous similar studies have failed to sort B- and R- lines (Menz et al., 2004).

The relatively close association within the *ALS* and *ACCase* herbicide resistant pollinators can be explained in the same way as above. Resistance breeding to both herbicide chemistries was a relatively recent endeavor and was conducted by single institution. Hence, the amount of diversity that could be captured in this short period in both *ALS* and *ACCase* herbicide resistant materials is expected to be low. Moreover, since the resistance donor is just a single source, hence all *ALS* or *ACCase* herbicide resistant lines have high chance of sharing common allele that may contribute some degree of similarity regardless of genes in other regions of the genome that a relatively low diversity is expected among these classes of inbreds.

Genome-wide LD in the current study was 0.27 with 79.8% of SNP marker pairs in significant LD ($p < 0.05$). Both the magnitude and distribution of LD across genomes are important factors affecting the precision of association studies, and the marker-assisted breeding effectiveness (Sorrells and Yu, 2009). Because of the rapid rate of inbreeding that quickly brings alleles to fixation, usually a higher level of LD is expected in sorghum than in maize (Tian et al., 2009; Gore et al., 2009). LD among the pedigree based sub-groups was higher than the overall LD perhaps due to smaller

population size. The *ACCase* and *ALS* herbicide resistant subgroups had higher LD than the regular subgroups which may be either because they have smaller size compared to the regular lines or because they consist of more minor frequency alleles that may have dragged along from the wild herbicide resistance gene donors. The LD across the genome decayed gradually as the physical distance increased (Figure 2.6) and we observed variable levels of LD decay distances in all of the 10 chromosomes (Figures B.1-B.5). In almost all the chromosomes, we observed a decrease in LD with increase in physical distance though with few exceptions. These results corroborate the previous findings (Wang et al., 2013; Bouchet et al., 2012; Wen et al., 2011). However, in other chromosomes it was observed that LD decreased as physical distance increased up to a certain point and then started increasing again with increased physical distance (data not shown). Similar results have been reported by Wang et al. (2013) where they observed LD value of 0.08 at the physical distance of > 10Mb on SBI-06 almost similar to the average r^2 value (0.084) at 30-50 kb across all the 10 chromosomes. Contrary to previous studies that have reported shorter LD of between 10-15kb (Hamblin et al., 2005), in the present study, LD decayed by > 100kb both across the genome as well as in all 10 chromosomes (Figure 2.6 and Figures B.1-B.5). Longer LD decay distance in inbred lines is expected due to the fact that these materials have undergone selfing and selection for several generations, hence creating large LD blocks (Takano-Kai et al., 2009; Huang et al., 2012).

Again, LD decay detected in this study was higher than what has been reported in earlier studies (Yan et al., 2009; Wang et al., 2013; Hamblin et al., 2005). This was expected because of the difference in the types of materials used in those studies (inbred lines vs. landraces). LD decays relatively faster with increasing genetic map distance in wild relatives and landraces than in inbred lines (Tenaillon et

al., 2001; Ching et al., 2002). This is perhaps due to the fact that landraces or wild weedy species have not been subjected to strong directional forces frequently used in plant breeding to create inbred lines (Cardwell et al., 2006; Hamblin et al., 2010; Morell et al., 2005).

Conclusion

In this study the GBS approach was used for SNP discovery and genotyping public sorghum inbred lines. The study demonstrated that public sorghum inbred lines, although they have gone through rigorous selection pressure under strict selfing conditions, still maintained reasonable genetic diversity. The relatively narrow genetic diversity estimated in parental lines resistant to ALS and the *ACCase* inhibitor herbicides results from the relatively short history of herbicide resistance breeding. Continued effort to bring fresh germplasm into the breeding program will help maintain robust diversity among the inbreds. The ever expanding genomic and bioinformatic platforms are providing more power to breeders and geneticists to dissect this diversity and utilize the germplasm resources in a way it can make more impact.

References

Becker, R. A., J. M. Chambers, and A. R. Wilks. 1988. Wadsworth & Brooks/Cole, Pacific Grove, PA. The New S Language.

Benchimol, L.L., C.L. Souza Junior, A.A.F. Garcia, P.M.S. Kono, C.A. Mangolin, A.M.M. Barbosa et al. 2000. Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. Plant Breed. 119:491-496.

Botstein, D. 1980. A theory of modular evolution for bacteriophages. Ann. N.Y. Acad. Sci. 354:484-490.

Bouchet, S., D. Pot, M. Deu, J.F. Rami, and C. Billot et al. 2012. Genetic structure, linkage disequilibrium and signature of selection in sorghum: lessons from physically anchored DArT markers. PLoS One 7: e33470.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, and Y. Ramdoss et al. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Caldwell, K.S., J. Russell, P. Langridge, and W. Powell. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genet. 172:557-567.

Chagne, D., J. Batley, D. Edwards, J.W. Forster.2007. In single nucleotide polymorphisms genotyping in plants. Springer pp:77-94.

Chao, S., W. Zhang, J. Dubcovsky, and M. Sorrells. 2007. Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market class. Crop Sci. 47:1018-1030.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante, and A.J. Rafalski. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3:19.

Cui, X., R.P.Wise, and P.S. Schnable. 1996. The rf2 restorer gene of male-sterile t-cytoplasm maize. Sci. 272:1334-1336.

Doyle, J.J.1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Endelman, J.B. and J. L. Jannink. 2012. Shrinkage estimation of the realized matrix. G3 2:1405-1413.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin 3.01: An integrated software package for population genetics data analysis. Evol. Bioinformatics Online 1:47-50.

Falush, D., M. Stephens, and J.K. Pritchard. 2012. Inference of Population Structure Using Genotype Data: Linked Loci and Correlated Allele Frequencies. Genet. 164:1567-1587.

Ganapathy, K.N., S.S. Gomashe,S. Rakshit, B. Prabhakar, S.S. Ambekar, R.B. Ghorade, B.D. Biradar, U. Saxena, and J.V. and Patil. 2012. Genetic diversity revealed utility of SSR markers in classifying parental lines and elite genotypes of sorghum [*Sorghum bicolor* (L.) Moench]. Australian J. Crop Sci. 6(11): 1486.

Gore, M. A., J.M. Chia, R. J. Elshire, Q. Sun, and E. S. Ersoz et al. 2009. A first-generation haplotype map of maize. Sci. 326:1115-1117.

Hamblin, M. T., M. G. Salas Fernandez, A. M. Casa, S. E. Mitchell, and A. H. Paterson et al. 2005. Equilibrium processes cannot explain high levels of

short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genet. 171:1247-1256.

Hamblin, M.T., T.J. Close, P.R. Bhat, S. Chao, J.G. Kling, K.J. Abraham, T. Blake, W.S. Brooks, B. Cooper, C.A. Griffey, and P.M. Hayes. 2010. Population structure and linkage disequilibrium in US barley germplasm: implications for association mapping. Crop Sci.50 (2): 556-566.

Harvey, P.R. 1977. Sorghum germplasm base in the U.S. In proc. 32nd Com and Sorghum Res. Conf. Amer. Seed Trade Assoc. Chicago, IL. p. 186-198.

Hung, H.Y., L.M. Shannon, F. Tian, P.J. Bradbury, C. Chen, S.A. Flint-Garcia, M.D. McMullen, D. Ware, E.S. Buckler, J.F. Doebley, and J.B. Holland. 2012. ZmCCT and the genetic basis of day length adaptation underlying the post-domestication spread of maize. Proc. Natl. Acad. Sci. USA 109: E1913-E1921.

IBM Corp. Released 2013. IBM SPSS Statistics for windows, version 22.0 Armonk, N.Y: IBM Corp.

Klein, R.R., F.R. Miller, S. Bean, and P.E. Klein, 2016. Registration of 40 Converted Germplasm Sources from the Reinstated Sorghum Conversion Program. J. Plant Registrations 10(1): 57-61.

Klein, R.R, R. Rodriguez-Herrera, J.A. Schlueter, P.E. Klein, Z.H. Yu , and W.L. Rooney. 2001. Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor. Appl .Genet.102:307-319.

Liu, K., and S.V. Muse, 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Lu, Y., J.Yan, C. Guimarães, S. Taba, Z. Hao , S. Gao et al. 2009.Molecular

characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor. Appl. Genet.120:93-115.

Mace, E.S., S. Tai, E.K. Gilding, Y. Li, P.J. Prentis, L. Bian, B.C. Campbell, W.Hu, D.J. Innes, X. Han, and A. Cruickshank. 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nature communications, 4.

Menz, M. A., R. R. Klein, N. C. Unruh, W. L. Rooney, and P. E. Klein et al. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44: 1236-1244.

Morrell, P. L., D. M. Toleno, K. E. Lundy, and M. T. Clegg. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. spontaneum) despite high rates of self-fertilization. Proceed. Natl. Acad. Sci. USA 102:2442-2447.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, S. Kresovich. 2013. Population genomic and genome-wide association studies of agronomic traits in sorghum. PNAS 110(2): 453-458.

Murdoch, A.D. 2013. rgl: 3D visualization device system (OpenGL). R package version 0.93.945, URL http://CRAN.R-project.org/package=rgl.

Nei, M.1972. Genetic distance between populations. Am. Nat. 106: 283-292.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. De Bakker , M.J. Daly, and P.C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559-575.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 5:94-100.

Reif, J.C. A.E. Melchinger, X.C. Xia, M.L. Warburton, D.A. Hoisington, S.K. Vasal et al. 2003. Use of SSRs for establishing heterotic groups in subtropical maize. Theor. Appl. Genet. 107:947-957.

Satish, K., Z. Gutema, C. Grenier, P.J. Rich, and G. Ejeta. 2012. Molecular tagging and validation of microsatellite markers linked to the low germination stimulant gene (*lgs*) for Striga resistance in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 124(6): 989-1003.

Semagn, K., C. Magorokosho, B.S. Vivek, D. Makumbi, Y. Beyene, S. Mugo, B.M. Prasanna, and M.L.Warburton. 2012. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC genomics 13(1):113.

Schlotterer, C. 2004. The evolution of molecular markers-just a matter of fashion? Nat. Rev. Genet. 5:63-69.

Sorrells, M.E., and J. Yu. 2009. Linkage disequilibrium and association mapping in the Triticeae. In C. Feuillet and G.J. Muehlbauer (ed.) Genetics and genomics of the Triticeae 7:655-683. Springer.

Takano-Kai, N., H. Jiang, T. Kubo, M. Sweeney, and T. Matsumoto et al. 2009. Evolutionary history of GS3, a gene conferring grain length in rice. Genet. 182:1323-1334.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, and M. Nei et al. 2011. Mega5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731-2739.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). Proc. Natl. Acad. Sci. USA 98:9161-9166.

Tian, F., N.M. Stevens, and E.S. Buckler. 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proceed. Natl. Acad. Sci. 106(Supplement 1): 9979-9986.

Wang, Y.H., H.D. Upadhyaya, A.M. Burrell, S.M.E. Sahraeian, R.R. Klein, and P.E.Klein.2013. Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, Sorghum bicolor. G3: Genes| Genomes| Genet. 3(5):783-793.

Weir, B.S. 1996. Genetic data analysis II. Sinauer Associated, Inc., Sunderland, MA.

Wen, W., J.L. Araus, S. Trushar, J. Cairns, G. Mahuku, and M. Bänziger et al. 2011. Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci. 51:2569-2581.

Wu, Y., F. San Vicente, K. Huang, T. Dhliwayo, D.E. Costich, K.Semagn, N. Sudha, M. Olsen, B.M.Prasanna, X. Zhang, and R. Babu. 2016. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor. Appl. Genet. 129(4): 753-765.

Yan, J., T. Shah, W.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451.

Zheng, L. et al. 2011. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol. 12: R114.

Zhu,Y.L, Q.J. Song, D.L. Hyten , C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, et al. 2003. Single-Nucleotide Polymorphisms in Soybean. Genet.163:1123-1134.

Chapter 3 - Association of Genetic Distance Between Parental Inbred Lines with Hybrid Vigor in Sorghum [*Sorghum bicolor* (L.) Moench]

Abstract

Hybrid vigor or heterosis is an important component of hybrid performance. The current approach used for selecting potential parental lines that can produce superior sorghum hybrids is very expensive and time consuming. It involves testing large number of testcross hybrids in multi-environment trials. New tools are needed to increase efficiency of selecting the most promising parental lines. Previous studies indicate that heterosis is higher in crosses that involve genetically diverse parents. The objective of this study was to investigate whether the molecular-based genetic distance between parental lines is associated with hybrid vigor in sorghum. A total of 279 public sorghum inbred lines were genotyped using 66,265 SNPs generated using genotyping-by-sequencing (GBS) and genetic distance between parents of pairwise comparisons was estimated. A total of 60 parental lines were selected including 30 closely and another set of 30 distantly related lines and used to develop 102 F1 hybrids. The F1 hybrids including three commercial hybrid checks (Seneca, Dekalb54-00 and Pioneer84G62) and their parental lines were evaluated at Kansas State University Research farm near Manhattan and the North East experimental station at Ottawa, KS during 2015 summer season. The experiment was laid in a randomized complete block design with three replicates. Data were collected on days to flowering, plant height, grain yield and yield components including panicle length, panicle weight, panicle yield, number of kernels per panicle and thousand kernel weight. The genetic distance between closely related parental inbred lines ranged

from 0.1 to 0.39 and was 0.4 to 0.70 for distantly related parents. The results show that correlations of genetic distance between parental lines with hybrid performance and heterosis were variable and dependent on the trait. Though most were statistically non-significant and may be too weak to be used as predictor for hybrid performance and heterosis, the results tend to show that certain level of genetic distance between parents is needed to capture maximum heterosis and hybrid performance.

Key words: Sorghum; Heterosis; Correlation; Hybrid performance; GBS, Genotyping-by-sequencing; Genetic distance.

Introduction

Hybrid vigor or heterosis is the most important component of sorghum hybrid performance. It is a phenomenon whereby an F1 hybrid between two individuals show increased vigor or yield compared to their inbred parents. The concept of hybrid vigor traces back to early experiments on heterosis and its complement inbreeding conducted by Shull (1908, 1909) and East (1908). In these studies it was observed that when maize plants were selfed their vigor and grain yield declined rapidly. However, when two inbred lines were crossed both vigor and grain yield of the F1 hybrid often exceeded the mean of the two parents. It was this observation that was made over 90 years ago, and methodology explained by Shull (1909) that gave rise to the modern hybrid breeding technology (Crow, 1998).

The concept of heterosis was first utilized in maize [*Zea mays* (L.)] breeding, which significantly increased yields by 15% and 20-30% in rice compared to the superior open-pollinated varieties (Cheng et al., 2007). By the late twentieth century hybrid maize accounted for 65% of total maize cultivation contributing to a tremendous increase of annual maize production (Duvick, 1999). And by the end of

the twentieth century, almost half of the world sorghum production were hybrids, contributing about 35-40% yield gains in the United States (Duvick, 1999).

Although hybrid breeding technology has been a great success in increasing yields in many cereal crops including sorghum, the process of developing and evaluating the performance of hybrids remains the most expensive and time-consuming activity. In sorghum, for example, a lot of resources are spent in developing parental inbred lines and evaluating their potential for hybrid performance, thus very expensive and time-consuming. Recently, prediction of hybrid performance has been attracting much interest in many hybrid breeding programs including sorghum (Jordan et al., 2003), rice (Singh et al., 2011) and maize (Hallauer et al., 1988; Drinic et al., 2002; Xu et al., 2004). Prediction of heterosis or hybrid performance may not only help to reduce costs associated with developing and phenotyping a large number of crosses but also may increase the speed at which superior hybrids can be identified. Several plant breeders have been looking for the possibility of predicting heterosis based on the biochemical, physiological, pedigree, morphological and molecular marker data.

Earlier attempts to use isozyme variation to predict heterosis for grain yield (Schwartz, 1960; Stuber et al., 1980; Frei et al., 1986) was not successful because it did not provide accurate prediction perhaps due to the limited number of isozyme loci that may not be linked to loci associated with traits of interest (Hadjinov et al., 1980; Lamkey et al., 1987). Later on several molecular markers were used to study heterosis. Of these, SNP markers appear to have the greatest potential because of their abundance in the genome. So far a number of studies have been conducted to determine the association between molecular marker based diversity with hybrid performance and heterosis in different crop species including rice (Singh et al., 2011),

sorghum (Jordan et al., 2003) and maize (Smith et al., 1990; Xu et al., 2004; Drinic et al., 2001).

However, most of the earlier studies have reported inconsistent results about the relationship of marker-based genetic distance between parental inbred lines with heterosis. For example, significant and positive correlations between molecular marker based genetic distance and heterosis have been reported in rice (Cai et al., 2005; Zhao et al., 2008). Similarly, Zhang et al (2010) reported positive correlations between genetic distance between parental lines with heterosis for plant height and panicle length. In contrast, other studies did not find any correlations (Bernardo, 1992; Munhoz et al., 2009; Liao et al.,1998; Zhang et al., 2006) implying that genetic distance may not be the best predictor of heterosis (Zhang et al., 1994, 1995; Xiao et al., 1996). Some of the possible reasons mentioned for lack of strong correlation of genetic distance between parents and heterosis include lack of strong linkage between a large proportion of markers with the QTLs controlling the trait of interest and low heritability of the trait with low dominance effects.

To date, there is little information available about the relationship between molecular marker-based genetic distance between parental inbred lines and heterosis for yield and yield components in sorghum. This study is aimed at investigating the association between genome-wide marker-based genetic distance between parental inbred lines and heterosis for key agronomic traits of sorghum.

Materials and methods

Genetic materials

A total of 279 public sorghum inbred lines were used in this study. These included 228 pollinator lines and 51 seed parent lines developed at KSU and Texas A&M sorghum breeding programs. The lines represent diverse pedigrees and have

diverse morphological and agronomic characteristics with some 52 of them having resistance to Acetyl co-enzyme-A carboxylase (*ACCase)* inhibitor herbicides and 89 of them to Acetolactate Synthase (*ALS*) inhibitor herbicides. The remaining 138 lines have been released over several years and may carry traits for drought tolerance, disease and insect resistance. Of these, sixty parental inbred lines were selected based on their genetic distance with 30 of them closely related with genetic distance ranging from 0.1 to 0.39 and another 30 distantly related lines with genetic distance ranging from 0.4 to 0.7. Each set was comprised of seed and pollinator parents. Crosses were made between the pollinators and seed parents in each set such that a total of 50 F1 hybrids among closely related parents and 52 hybrids among distantly related parents were created.

Experimental design and field management

The experiment was laid in a randomized complete block design with three replications. All of the hybrids (102) and the checks were evaluated for hybrid performance at Kansas State University (KSU) Research farm near Manhattan and North East Experimental station in Ottawa, KS during 2015 summer season. Manhattan location had silt loam: fine silty, mixed, superactive, mesic cumulic hapludolls soils. On average Manhattan location receives annual precipitation of about 907mm (35.7inches) with the average minimum and maximum temperatures of -18°C and 32°C, respectively. On average the annual precipitation for KSU Agronomy Research Farm Ashland Bottoms near Manhattan, KS was 338, 539 and 576mm for 2012, 2013 and 2014, respectively. Ottawa location had woodson silt loam soil and annual precipitation can reach as far as 1000mm (40 inches) on average. The mean annual minimum temperature for this location is -7°C and the annual average maximum temperature is about 32°C. The gross plot size was 2 rows,

5m long each spaced at 0.75m apart. Planting date for KSU Research Farm near Manhattan, KS was June 23, 2015 while it was June 24, 2015 at North East Experimental station in Ottawa, KS. Weeds were controlled with pre-plant herbicides including 1.2 Atrazine, 1 2/3 pints of Dual II Mg and 5.7 oz of Calisto. All weeds that emerged after planting were controlled with hand weeding throughout the experimental period.

Data collection

Data were collected on days to flowering, plant height, grain yield and yield components including panicle length, panicle weight, panicle yield number of kernels per panicle, and thousand kernel weight.

- Days to flowering (DF) was recorded as the number of days from planting to when 50% of the plants in each plot reached half bloom.
- Plant height (PH) was measured as the distance from the soil surface to the tip of the panicle at physiological maturity expressed in centimeters.
- Grain yield was recorded as the weight of the kernels harvested at maturity from each plot expressed in kilograms per hectare.

After physiological maturity, three panicles from main plants were randomly sampled from each plot for measuring yield components:

- Panicle length (PL) was determined as the mean length of the panicles measured from the base to the tip of the panicle.
- Panicle weight (PW) was recorded as the weight of panicle from individual plant.
- Panicle yield (PL) was measured as the weight of grains threshed from individual panicle.

- Number of kernels per panicle (KN) was determined by counting the sorghum kernels threshed from the panicle using a laboratory seed counter (Seed Counter Model 850-3, International Marketing and Design Corp., Lookout Road, San Antonio, TX, 78233, USA).
- Thousand kernel weight (TKW) was determined by measuring the weight of 250 kernels from each panicle and multiplying by four.

The yield components per plot basis was determined by taking the mean of the three panicles for all of the yield components. Data for panicle yield, thousand kernel weight and grain yield were adjusted to 12.5% moisture content before statistical analysis.

Statistical analysis

Two hundred and seventy-nine parental inbred lines were SNP genotyped using genotyping-by-sequencing (GBS) and generated 282,536 SNPs with minor allele frequency (MAF) ≥ 0.01 and < 20% missing data. The genotypic data was filtered for minor allele frequency using PLINK v1.07 (Purcell et al., 2007) resulting in 66,265 SNPs with MAF ≥ 0.05 for use in the analysis. Genome-wide Nei's (1972) genetic distance between parental lines was calculated using *adegenet* package in R program (R Development Core Team, 2010). This genetic distance estimate assumes that genetic differences between individuals are due to mutation and genetic drift.

Analysis of variance (ANOVA) was performed for each trait using SAS 9.4 PROC GLM (SAS Institute, Cary NC). Pearson correlation analyses were performed using the procedure CORR in SAS. Broad-sense heritability (H) for each trait was estimated across environments and replicates (Hallauer et al., 2010).

$$H = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/e + \sigma_e^2/er}$$

where σ^2_g, σ^2_{ge} and σ^2_e are the genetic, genotype-by-environment interaction and residue variance components, respectively. *H* was estimated for means across environments, r and e are number of replications and environments, respectively. Better-parent heterosis and mid-parent heterosis were calculated for all traits measured using Fehr (1987) method as follows:

$$\text{Better Parent Heterosis (BPH)} = \frac{(F_1 - BP)}{BP} \times 100$$

$$\text{Mid Parent Heterosis (MPH)} = \frac{(F_1 - MP)}{MP} \times 100$$

where F_1 is the performance of the F_1 hybrid, BP is the performance of the better parent involved in a cross and MP is the average performance of the parents involved in the cross. Significance of heterosis estimate was tested using a *t* test at α=0.05: $t = (X_{DHM}-X_{MP})/SE(X_{DHM}-X_{MP})$, where X_{DHM} is the average of all derived hybrid means, X_{MP} is the average of all mean parental values and SE is the standard error of the difference between averages (Prado et al., 2013).

Results

Analysis of variance and hybrid performance

The combined analysis of variance for days to flowering, plant height, grain yield and yield components is presented in Tables 3.1 and 3.2. The genotype and genotype-by-environment interaction effects were significant for all traits in both the parents and the hybrids (Tables 3.1 and 3.2).

Across environment performance results of the parents and the hybrids for eight agronomic traits are presented in Table 3.3. There were significant differences among the parents and the hybrids for all traits. Generally, the hybrids outperformed their parental inbred lines for all traits measured. On average, the hybrids were 9.4,

29.2, 39.4 and 31.4% higher for PL, PW, PY and KN, respectively, than the parents. Again, the hybrids had higher TKW than that of the parents (Table 3.3). Moreover, the hybrids started flowering 8 d earlier than the parents and were 21.7cm taller (Table 3.3).

The range for PL was 26.8 to 34.4cm and 26.2 to 35.8cm among parents and hybrids, respectively. The range for PW was 72.6 to 127.4g among the parents and 96.5 to 176g among the hybrids. Similarly, PY and KN among the parents ranged from 36.7 to 73.9g and 1431 to 2409, respectively. PY among the hybrids ranged from 55.5 to 114.8g while the range for KN was 1722 to 3564. The range for TKW among the parents and hybrids was 23.8 to 35.3g and 29.2 to 38.4g, respectively. DF among the parents ranged from 60 to 80 d, while the range was 53 to 67 d in the hybrids. Compared to the hybrids, the parents were shorter ranging from 56.3 to 132.5cm than 119.6 to 173.4cm for the hybrids. The mean PH for the parents was 120.2cm and 141.9cm for hybrids. The grain yield among parents ranged from 5256 to 12268.5 kg ha^{-1} compared to 8773.0 to 16944.1 kg ha^{-1} among hybrids.

Table 3.1 Mean squares from the combined analysis of variance for eight agronomic traits of sorghum inbred lines evaluated at Manhattan and Ottawa, KS during 2015 summer season.

Parental inbred lines				
Trait	Environment (E)	Genotype (G)	G x E	Error
Panicle length (cm)	111.2	17.2***	13.9***	5.7
Panicle weight (g)	49280.7	585.7*	732.8**	366.6
Panicle yield (g)	13116.8	328.8*	397.9**	197.2
Number of kernels panicle^{-1}	23334631.6	274537.5*	321210.9*	209749.0
Thousand kernel weight (g)	437.7	28.6***	19.6**	13.7
Days to flowering	804.3	145.3***	147.8***	26.7
Plant height (cm)	7904.9	309.2***	562.3***	86.2
Grain yield (kg ha^{-1})	2037969906	80143322*	19184053*	11503052

*,** and *** Significant at $p \leq 0.05$, 0.01 and 0.001, respectively.

Table 3.2 Mean squares from the combined analysis of variance of eight agronomic traits of sorghum hybrids evaluated at Manhattan and Ottawa, KS during 2015 summer season.

	F1 hybrids			
Trait	Environment (E)	Genotype (G)	G x E	Error
Panicle length (cm)	352.3	23.7***	5.3**	3.2
Panicle weight (g)	91043.4	1063.3***	617.1***	389.3
Panicle yield (g)	34889.5	550***	375.3***	166.4
Number of kernels panicle^{-1}	52348935.5	498548.0***	272843.3***	158105.2
Thousand kernel weight (g)	297.0	24.3***	12.5***	6.3
Days to flowering	2404.9	59.0***	14.4*	10.7
Plant height (cm)	5885	686.3***	149.1***	46.5
Grain yield (kg ha^{-1})	4816726847	13442369*	19647919***	10926986

*,** and *** Significant at $p \leq 0.05$, 0.01 and 0.001, respectively.

Table 3.3 Results for across environment performance of parental inbred lines and the hybrids evaluated at Manhattan and Ottawa, KS during 2015 summer season.

	Parental inbred lines				F1 Hybrids			
Trait	Mean	SD	Range	H	Mean	SD	Range	H
Panicle length (cm)	29.6	1.8	26.8-34.4	0.23	32.4	2.0	26.2-35.8	0.80
Panicle weight (g)	97.3	12.0	72.6-127.4	0.19	125.7	14.5	96.5-176.0	0.45
Panicle yield (g)	55.2	8.2	36.7-73.9	0.17	77.0	10.3	55.5-114.8	0.36
Number of kernels panicle $^{-1}$	1832	240	1431-2409	0.14	2407	325	1722-3564	0.49
Thousand kernel weight (g)	30.8	2.2	23.8-35.3	0.37	32.2	1.9	29.2-38.4	0.64
Days to flowering	69	4.9	60-80	0.32	61	3.2	53-67	0.77
Plant height (cm)	120.2	11.5	56.3-132.5	0.39	141.9	10.9	119.6-173.4	0.82
Grain yield (kg ha^{-1})	9328	1588.5	5256-12268	0.16	9825	1537.8	8773-16944	0.27

SD = Standard deviation of the mean; H = Broad-sense heritability

Heterosis estimates for eight agronomic traits

Data for across environment analysis for mid-parent heterosis (MPH) and better-parent heterosis (BPH) are presented in Table 3.4. Both MPH and BPH varied considerably for the different traits with the lowest mean MPH of -11.3 recorded for DF and the highest (36.3%) for grain yield. The range for these traits was -16.7 to 8.1 and -6.1 to 125.8%, respectively. MPH for PY ranged from -6.3 to 112.1 and -11.9 to 89.1 for KN with a mean of 35.5 and 27.25, respectively. The range of MPH for TKW was narrower, only -7.9 to 24.5% with a mean of 4.9%. In general all traits exhibited average positive MPH except DF where almost all hybrids were earlier than their respective inbred parents resulting in an overall average negative heterosis.

The result for BPH had similar trend with MPH but with mean values marked smaller than that of the MPH except DF. Like the MPH, mean BPH for all traits except DF was positive indicating that, on average, the hybrids outperformed the better parent though this may be different for individual hybrids. For all yield and yield components studied, the smallest positive BPH recorded was 2.4 % for TKW and the highest was 26.1% for PY. From the range of both MPH and BPH results, it is apparent that few hybrids performed less than their corresponding inbred parents for almost all of the yield components. Crosses with negative heterosis were not necessarily those derived from parents with close genetic relationships. In general the mean performance of the inbred parents compared to the mean performance of their hybrids and that of the highest inbred values compared to the highest hybrid values was less than 100% for all traits except DF.

Disaggregation of the data into hybrids developed from closely and distantly related parental lines further revealed differences in the levels of heterosis for all traits (Tables 3.5 and 3.6). Among the hybrids from closely related parents, mid-parent

heterosis (MPH) ranged from -10.3% for DF to 35.1% for grain yield. For better-parent heterosis (BPH), the range was -6.2% for DF to 26.5% for PY (Table 3.5). On the other hand, MPH ranged from -12.2% for DF to 37.5% for grain yield, while the values for BPH ranged from -7.2% for DF to 26.7% for grain yield (Table 3.6). Furthermore, a larger percentage number of the hybrids from distantly related parents had positive levels of heterosis compared to the hybrids from closely related parents (Table 3.6). Hybrids among distantly related parents had larger percentage number of hybrids with positive MPH for all traits except PL and grain yield compared to hybrids from closely related parents. Likewise, a larger percentage of hybrids among distantly related parents had larger positive BPH than those hybrids from closely related parents for almost all traits except KN, TKW and DF (Table 3.6).

Correlations among agronomic traits of F1 hybrids

The across environment pearson correlation coefficients between yield and yield components for the hybrids are presented in Table 3.7. PL was significantly and positively correlated with PW (r = 0.39), PY (r = 0.39) and KN (r = 0.31), but not correlated with grain yield and other agronomic traits. Similarly, PW was significantly correlated with PY (r = 0.94) and KN (r = 0.88) but not correlated with grain yield, TKW, DF and PH. On the other hand, PY was significantly and positively correlated with KN (r = 0.90) and the correlations with other traits were not significant. Moreover, TKW was significantly and positively correlated with PL (r = 0.24), PW (r = 0.20) and PY (r = 0.30) and its correlation with KN was negative and not significant, indicating that as the size of the kernel increases, the KN decreases. Perhaps due to higher photosynthetic area (more leaves), PH was significantly and positively correlated with PL (r = 0.43), PW (r = 0.28), PY (r = 0.38) and TKW (r = 0.49) but not with KN and DF. Again DF was significantly and positively correlated

with PW (r = 0.26), PY (r = 0.33) and KN (r = 0.25), but not correlated with the other agronomic traits (Table 3.7)

Table 3.4 Across environment results for mid-parent heterosis (MPH) and better-parent heterosis (BPH) for eight agronomic traits of sorghum [*Sorghum bicolor* (L.) Moench].

	Heterosis (%)							
	Mid-parent heterosis (MPH)				Better-parent heterosis (BPH)			
Trait	Mean	SD	Range	£ Hybrids with "+" heterosis	Mean	SD	Range	£ Hybrids with "+" heterosis
Panicle length (cm)	9.4	7.6	-16.0-25.7	91.8	6.5	7.9	-16.4-22.4	83.6
Panicle weight (g)	25.8	17.6	-6.0-71.6	94.5	20.0	19.9	-15.6-85.5	86.3
Panicle yield (g)	35.5	23.8	-6.3-112.1	97.3	26.1	21.6	-14.9-88.2	93.2
Number of kernels panicle^{-1}	27.2	20.8	-11.9-89.1	95.9	18.9	20.9	-20.3-77.8	83.6
Thousand kernel weight (g)	4.9	6.4	-7.9-24.5	76.7	2.4	7.8	-11.8-37.4	63.0
Days to flowering	-11.3	5.1	-16.7 to -8.1	0	-6.7	6.3	-17.9-8.1	19.2
Plant height (cm)	21.7	19.1	-2.0-89.7	94.5	13.8	10.3	-6.0-41.0	94.5
Grain yield (kg ha^{-1})	36.3	23.4	-6.1-125.8	95.8	24.9	22.7	-19.8-88.4	90.4

SD = Standard deviations; £ = proportion of hybrids exhibiting positive heterosis for the trait.

Table 3.5 Across environment results for mid-parent heterosis (MPH) and better-parent heterosis (BPH) for eight agronomic traits in sorghum hybrids derived from closely related parental lines.

	Heterosis (%)							
	Mid-parent heterosis (MPH)				Better-parent heterosis (BPH)			
Trait	Mean	SD	Range	£ Hybrids with "+" heterosis	Mean	SD	Range	£Hybrids with "+" heterosis
Panicle length (cm)	10.3	7.0	-8.2-25.7	97.2	7.2	7.4	-11.0-22.4	83.3
Panicle weight (g)	25.5	17.5	-6.0-62.8	91.7	20.4	20.4	-15.6-85.5	86.1
Panicle yield (g)	34.4	23	-2.7-88.2	97.2	26.5	20.6	-10.9-67.1	91.7
Number of kernels panicle^{-1}	27.9	20.7	-6.3-89.1	94.4	20.9	20.7	-15.6-77.8	88.9
Thousand kernel weight (g)	4.3	7.0	-7.9-24.5	75	2.4	8.4	-11.8-37.4	63.9
Days to flowering	-10.3	4.9	-17.8 -3.0	2	-6.2	5.6	-17.2-8.1	13.9
Plant height (cm)	22.8	19.5	-2.0-81.55	97.2	13.8	9.5	-6.0-35.8	94.4
Grain yield (kg ha^{-1})	35.1	27.6	-5.0-125.8	100	23.0	25.0	-15.9-88.4	88.9

SD = Standard deviations; £ = proportion of hybrids exhibiting positive heterosis for the trait.

Table 3.6 Across environment results for mid-parent heterosis (MPH) and better-parent heterosis (BPH) for eight agronomic traits of sorghum hybrids derived from distantly related parental lines.

	Heterosis (%)							
	Mid-parent heterosis (MPH)				Better-parent heterosis (BPH)			
Trait	Mean	SD	Range	£ Hybrids with "+" heterosis	Mean	SD	Range	£ Hybrids with "+" heterosis
Panicle length (cm)	8.5	8.2	-16.0-23.1	86.5	5.9	8.5	-16.4-20.2	83.8
Panicle weight (g)	26.1	17.9	-5.2-71.6	97.3	19.6	19.8	-8.54-69.1	89.2
Panicle yield (g)	36.6	24.6	-6.3-112.1	97.3	25.8	22.9	-14.9-88.2	94.6
Number of kernels panicle^{-1}	26.5	21.2	-11.9-72.4	100	17	21.3	-20.3-60.1	78.4
Thousand kernel weight (g)	5.4	5.8	-2.8-20.6	78.9	2.5	7.3	-10.7-27.9	63.2
Days to flowering	-12.2	5.2	-22.2 to -2.32	0	-7.2	7.0	-18.0-5.9	15.8
Plant height (cm)	20.6	19.0	-0.85-89.7	97.4	13.7	11.2	-2.8-41.0	94.7
Grain yield (kg ha^{-1})	37.5	18.8	-6.1-75.8	94.7	26.7	20.4	-19.8-69.5	92.1

SD = Standard deviations; £ = proportion of hybrids exhibiting positive heterosis for the trait.

Table 3.7 Pearson correlation coefficients (r) among eight agronomic traits of sorghum hybrids evaluated at Manhattan and Ottawa, KS during 2015 summer season.

Trait	Panicle length (cm)	Panicle weight (g)	Panicle yield (g)	Number of kernels panicle^{-1}	Thousand kernel weight (g)	Days to flowering	Plant height (cm)	Grain yield (kg ha^{-1})
Panicle length (cm)	-							
Panicle weight (g)	0.39**	-						
Panicle yield (g)	0.39**	0.94**	-					
Number of kernels panicle^{-1}	0.31**	0.88**	0.90**	-				
Thousand kernel weight (g)	0.24*	0.20*	0.30**	-0.17	-			
Days to flowering	0.01	0.26**	0.33**	0.25**	0.17	-		
Plant height (cm)	0.43**	0.28**	0.38**	0.17	0.49**	0.19	-	
Grain yield (kg ha^{-1})	0.09	0.14	0.21	0.16	0.31	-0.08	-0.06	-

*, ** Significant at $p \leq 0.05$ and 0.001

Correlation of genetic distance (GD) between parental inbred lines with F1 hybrid performance and heterosis

The across environment pearson correlation analysis results of genetic distance between parents with hybrid performance *per se,* MPH and BPH for sorghum hybrids are presented in Table 3.8. For PL, the genetic distance between parents was negatively correlated with F1 hybrid performance (r = -0.02), MPH (r = -0.10) and BPH (r = -0.07). For PW, the genetic distance between parents and F1 hybrid performance were negatively correlated with each other. However, the genetic distance between parents for PW had positive and non-significant correlations with MPH of (r = 0.11) and BPH (r = 0.04). Similar to PW, PY the genetic distance was negatively correlated with F1 hybrid performance (r = -0.09) for PY, MPH (r = 0.11) and BPH (r = 0.04). For KN, the genetic distance had non-significant and negative correlations with F1 hybrid performance (r = -0.04) and BPH (r = -0.13), but its correlation with MPH was positive and non-significant. As for TKW, the genetic distance was positively correlated with F1 hybrid performance (r = 0.09), MPH (r = 0.09) and BPH (r = 0.04). Moreover, the genetic distance between parents for DF was negatively correlated with F1 hybrid performance (r = -0.17), MPH (r = -0.28) and BPH (r = -0.16). Similar to DF, the genetic distance between parents for PH had negative correlations with F1 hybrid performance, MPH and BPH with correlation coefficients (r) of -0.07, -0.14 and -0.08, respectively. For grain yield, the genetic distance between parents was also negatively correlated with F1 hybrid performance (r = -0.02), MPH (r = -0.10) and BPH (r = -0.10).

The across environment results of pearson correlations among F1 hybrid performance, MPH and BPH for all traits studied are presented in (Appendices C-F). For PL, mid-parent performance was positively correlated with F1 hybrid

performance. Similarly PL, MPH and BPH were highly correlated with each other. Again mid-parent value had negative correlations with both MPH and BPH (Appendix C). PW was positively correlated with mid-parent value . Similar to PL, PW had significant and positive correlations with MPH (r = 0.62) and BPH (r = 0.52), while MPH and BPH were strongly correlated with each other (r = 0.90) (Appendix C). For PY, F1 hybrid performance was positively and significantly correlated with mid-parent value (r = 0.64), MPH (r = 0.56) and BPH (r = 0.94) (Appendix D). MPH and BPH were also highly correlated with each other (r = 0.94). Again, F1 hybrid performance for number of kernels per panicle (KN) had significant and positive correlations with mid-parent value (r = 0.46), MPH (r = 0.70) and BPH (r = 0.62) (Appendix D). Similarly, TKW and DF for F1 hybrid were significantly and positively correlated with mid-parent value, MPH and BPH (Appendix E). Also MPH and BPH for both TKW and DF were highly correlated with each other. For PH, the F1 hybrid performance had a negative correlation with mid-parent performance but positive correlations with MPH (r = 0.69) and BPH (r = 0.88) (Appendix F). MPH was highly correlated with BPH (r = 0.74) which was significant. In addition, mid-parent performance had significant negative correlations with both MPH and BPH. While F1 hybrid performance had a weak correlation with mid-parent performance for grain yield, it was highly correlated with both MPH (r = 0.69) and BPH (r = 0.74). Again MPH and BPH had a very strong correlation with each other (r = 0.89) (Appendix F)

Table 3.8 Pearson correlation coefficients (*r*) of genetic distance between parental lines with hybrid performance (F1), mid-parent heterosis (MPH) and better-parent heterosis (BPH) for hybrids developed from both closely and distantly parents.

Trait	Panicle length (cm)	Panicle weight (g)	Panicle yield (g)	Number of kernels panicle^{-1}	Thousand kernel weight (g)	Days to flowering	Plant height (cm)	Grain yield (kg ha^{-1})
F1	-0.02	-0.14	-0.09	-0.13	0.09	-0.17	-0.07	-0.02
MPH	-0.10	0.11	0.11	0.05	0.09	-0.28*	-0.14	-0.10
BPH	-0.07	0.04	0.02	-0.04	0.04	-0.16	-0.08	-0.07

F1 = F1 hybrid performance; MPH = Mid-parent heterosis; BPH = Better-parent heterosis; * Significant at $p \leq 0.05$

Discussion

Parental inbred for producing superior sorghum hybrids are traditionally identified by developing testcross hybrids among the potential parental lines for agronomic performance. Public programs are particularly interested in developing and releasing parental lines that combine well across other parents and hence test of combining ability for desired traits is a routine exercise. However, this remains the most expensive and time-consuming activity in hybrid breeding programs. Over the years, a number of molecular markers have been developed that can be used to assess the genetic diversity between parental inbred lines and be used for predicting hybrid performance (Singh, 1992; Jordan et al., 2003).

In the present study, some potential parental inbred lines were selected based on genetic distance and used to develop F1 hybrids to determine the correlation between genetic distance among parents and heterosis in sorghum. Heterosis was estimated for eight agronomic traits and the results revealed that sorghum hybrids exhibit both MPH and BPH for nearly every trait in almost all hybrids. Mean values of the hybrids were significantly larger for all traits than those of parental inbred lines, indicating that heterosis is largely positive. Comparatively, grain yield showed a greater level of heterosis unlike the other agronomic traits. In addition, some agronomic traits, for example, PW, PY, KN and TKW expressed average BPH in the desired direction indicating the presence of true heterosis, which indicate desirable genetic complementation between the parental inbred lines. In addition, most of the traits measured in this study exhibited positive BPH in over 90% of the evaluated hybrids. Although it has been reported that BPH is observed primarily for traits related to yield, in the present study we observed significant BPH for many other traits not directly related to yield such as PH.

In addition, the correlation between genetic distance between parental inbred lines and MPH and BPH for majority of the traits was either positive or negative depending on the trait but none of them were statistically significant except one between genetic distance and MPH for DF. This result does not agree with some studies conducted in few hybrid crops. But the lack of significant correlation may indicate that the SNP markers used in this study were not in LD with quantitative trait loci (QTLs) contributing to heterosis for the traits under study. Since different traits have just few of the numerous SNPs associated with them, diversity based on the overall SNP loci may not provide evidence of heterosis for a trait. Previous results by other authors agree with this argument by asserting that genetic distance cannot accurately predict hybrid performance unless the DNA markers used in the analysis are linked to the genes/QTLs affecting the trait (Charcosset and Essioux, 1994; Bernardo, 1993). Therefore, selecting markers that are in LD with QTLs that affect heterosis of the trait of interest in the materials under study may help to increase the prediction of heterosis based on genetic distance rather than simply increasing the number of markers (Bernado, 1993; Charcosset and Essioux, 1994).

The correlation between genetic distance between parental inbred lines and BPH for a given trait was often statistically non-significant (Table 3.8). Similar trends were observed for MPH as well suggesting that while genetic divergence is required for heterosis, it is a poor predictor of hybrid performance. The current findings corroborate with previous studies, for example, Chen et al. (2010) did not find any significant correlations, while Xangsayasane et al. (2010) found that the correlation of heterosis for yield and genetic distance between parents for yield was significant. Again grouping the hybrids into two groups, one from closely and the other from distantly related lines, showed that a larger number of hybrids from distantly related

parents exhibited positive heterosis for almost all traits suggesting that increase in the genetic diversity within certain range or group will lead to positive heterosis (Singh et al., 2011; Ajmone Marsan et al., 1998; Betran et al., 1997). However, genetic distance, in general, was poorly correlated with hybrid performance *per se* and heterosis for all traits measured. This result agrees with previous findings in various crop species such as rice (Kwon et al., 2002), wheat (Martin et al., 1995), maize (Benchimol et al., 2000) and alfalfa (Riday et al., 2003) that also showed low correlations of genetic distance with heterosis. Although, Singh et al. (2011) and Drinic et al. (2002) found significant correlations between genetic distance with grain yield, but they were generally weak to be used as good predictors of hybrid performance. In their study, they used SSR-based molecular markers to study the genetic divergence of maize inbred lines and correlate heterosis in derived maize crosses with genetic distance.

Conclusion

The results from the present study suggest that correlations of genetic distance between parental lines with hybrid performance and heterosis based on SNP markers were variable and dependent on the trait. Although most of the correlations found in this study were statistically not significant and very low to be used as predictor for hybrid performance and heterosis, majority of the hybrids developed from distantly related parents exhibited higher and positive heterosis for almost all traits compared to those hybrids developed from closely related parental lines. This suggests that a certain level of genetic distance or divergence between parents is needed to capture maximum heterosis and hybrid performance.

References

Ajmone, P. M., P. Castiglioni, and F. Fusari et al. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96:219-227.

Benchimol, L.L., C.L. Souza Junior, A.A.F. Garcia, P.M.S.Kono, C.A.Mangolin, and A.M.M. Barbosa et al. 2000. Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. Plant Breed. 119:491-496.

Bernardo, R. 1992. Relationship between single-cross performance and molecular marker heterozygosity. Theor. Appl. Genet. 83: 628-634.

Bernardo,R.1993. Estimation of coefficient of coancestry using molecular markers in maize. Theor. Appl. Genet. 85:1055-1062.

Betran, F., J. Ribaut, D. Beck, and D.G.de Leon.1997.Correlation between molecular marker data and hybrid performance in tropical maize. International Sympozium on Heterosis in Crops “Genetics and exploitation of heterosis in crops, pp.17-22.

Cai, J., and W. Lan. 2005. Using of AFLP marker to predict the hybrid yield and yield heterosis in rice. Chin Agric. Sci. Bull 21(4): 39-43.

Charcosset, A., and L. Essioux. 1994. The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor. Appl. Genet. 89: 336-343.

Chen, X., D. Sun, D.F.Rong, G. Sun, and J. Peng. 2010. Relationship of genetic diversity and hybrid performance in hybrids derived from a new photoperiod-thermo sensitive male sterile wheat line 337S. Euphytica 175(3): 365-371.

Cheng, S.H., J.Y. Zhuang, Y.Y. Fan, J.H. Du, and L.Y. Cao. 2007. Progress in research and development on hybrid rice: a super-domesticate in China. Ann. Bot. 100: 959-966.

Crow, J.F. 1998. 90 years ago: the beginning of hybrid maize. Genet. 148:923-928.

Drinic, S.M., S. Trifunovic, G.Drinic, and K. Kostantinov. 2002. Genetic divergence and its correlation to heterosis in maize as revealed by SSS-based Markers. Maydica 47(1):1-8.

Duvick, D.N. 1999. Heterosis: feeding people and protecting natural resources. In: Coors JG Pandey S (ed).The genetics and exploitation of heterosis in crops. Crop Sci. pp:19-30.

East, E.M.1908. Inbreeding in corn. pp. 419-428. Rep. Conn. Agric. Exp. Stn.

Fehr, W.R. 1987. Principle of cultivar development.Theory and technique. Vol. I.

Frei, O.M., C.W. Stuber , M.M.Goodman.1986. Use of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci. 26:37-42.

Hadjinov, M. I., V.S. Sherbak, and N.I. Benko et al.1980. Interrelationships between isozyme diversity combining ability in maize lines. Maydica 27:135-149.

Hallauer, A.R., M.J. Carena, and J.B. Miranda Filho. 2010. Quantitative Genetics in Maize Breeding, 3rd ed. Handbook of Plant Breeding Volume 6. Springer, New York. 663 pages.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative Genetics in Maize Breeding, 2nd ed. Iowa State University Press, Ames, IA. SB191.M2 H29.

Jordan, D.R., Y. Tao , I.D. Godwin, R.G. Henzell, M. Cooper, and C.L. McIntyre. 2003. Prediction of hybrid performance in grain sorghum using RFLP markers. Theor. Appl. Genet.106:559-567.

Lamkey, K. R., A.R. Hallauer, and A.L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. J. Hered. 78:231-234.

Liao, F.M., K.L. Zhou, H.H. Yang, and Q.S. Xu. 1998. Genetic difference of parents and its relation to heterosis in hybrid rice. Chin. J. Rice Sci. 12(4):193-199.

Martin, J.M., T.E. Talbert , S.P. Lanning, and N.K.Blake. 1995. Hybrid performance in wheat as related to parental diversity. Crop Sci. 35:104-108.

Munhoz, R.E.F., A.J. Prioli, A.T. Amaral Júnior, C.A. Scapim, and G.A. Simon. 2009. Genetic distances between popcorn populations based on molecular markers and correlations with heterosis estimates made by diallel analysis of hybrids. Genet. Mol. Res. 8(3): 951-962.

Nei, M.1972. Genetic distance between populations. Am. Nat. 106: 283-292.

Prado, S.A., L.G. Brenda, A. D. Novoa, D. Foster, M. L. Senior, C. Zinselmeier, M.E. Otegui, and L. Borrás. 2013. Correlations Between Parental Inbred Lines and Derived Hybrid Performance for Grain Filling Traits in Maize. Crop Sci. 53:1636-1645.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. De Bakker , M.J. Daly, and P.C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559-575.

Riday, H., E.C. Brummer, T.A. Cambell, and D. Luth. 2003. Comparison of genetic and morphological distance with heterosis between Medicago sativa and subsp. falcata. Euphytica 131:37-45.

Schwartz, D. 1960. Genetic studies of mutant isozymes in maize. Proc. Natl Acad. Sci. USA. 88:1202-1206.

Singh, A.K. 1992. RFLP based genetic diversity in relation to heterosis in crop

plant. In Abstr. Symp. Frontiers Plant Biotechnol. Nov.25-27, IARI, New Delhi, pp 43.

Singh, V.K., P. Upadhyay, P.Sinha, A.K. Mall, R.K. Ellur, A. Singh, S.K. Jaiswal, S. Biradar, S. Ramakrishna, R.M. Sundaram, and I. Ahmed. 2011. Prediction of hybrid performance based on the genetic distance of parental lines in two-line rice (Oryza sativa L.) hybrids. J. Crop Sci. Biotech. 14(1):1-10.

Shull, G.H., 1908.The composition of a field of maize. American Breeders Assoc. Rep. 296-301.

Shull, G.H. 1909. A pure line method of corn breeding. Am. Breed. Assoc. Rep. 5: 51-59.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J.Wall.1990. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield heterosis and RFLPs. Theor. Appl. Genet. 80:833-840.

Stuber, C. W., R.H. Moll, and M.M.Goodman. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genet.95: 225-236.

Xangsayasane, P., F. Xie, J.E. Hernandez, and T.H. Borromeo. 2010. Hybrid yield heterosis and genetic diversity of IRRI and Lao rice. Field Crop Res. 117:18-23.

Xiao, J., J.Li, L. Yuan, S.R. McCouch, and S.D. Tanksley. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor. Appl. Genet. 92:637-643.

Xu, S.X., J.I.E. Liu, and G.S. Liu. 2004. The use of SSRs for predicting the hybrid yield and yield heterosis in 15 key inbred lines of Chinese maize. Hereditas 141(3): 207-215.

Zhang, Q.F., Y.J. Gao, S.H. Yang, R.A. Ragab, M.A. Saghai-Maroof, and Z.B. Li. 1994. A half-diallel analysis of heterosis in elite hybrid rice based on RFLP and microsatellites. Theor. Appl. Genet. 89:185-192.

Zhang, T., L. Han, J.D. Xu, K.F. Jiang, X.J. Wu, X.D. Wang, and J.K. Zheng. 2006. Correlation between genetic distance and yield heterosis of hybrid aromatic rice. Sci. Agric. Sin 39(4): 831-835.

Zhang, Q.F., Y.J. Gao, M.A.S. Maroof, S.H. Yang and J.X. Li.1995. Molecular divergence and hybrid performance in rice. Mol. Breed. 1:133-142.

Zhang, T., X.L. Ni, K.F. Iang, H.F. Deng, H.E. Qing, Q.H. Yang, Y.A.N.G. Li, W.A.N Xian-Qi, Y.J. Cao, and J.K Kenyamn. 2010. Relationship between heterosis and parental genetic distance based on molecular markers for functional genes related to yield traits in rice. Rice Sci. 17(4): 288-295.

Zhao, Q.Y., Z. Zhu, Y.D. Zhang, L. Zhao, T. Chen, O.F. Zhang, and W.L Wang. 2008. Correlation analysis between genetic distance of SSR markers and heterosis japonica. The Fifth National Congress of Plant Molecular Breeding-cum-academic exchanges Proceedings.

Chapter 4 - Prediction of Hybrid Vigor Based on Inbred Line Performance in Sorghum [*Sorghum bicolor* (L.) Moench]

Abstract

The approach used to identify inbred lines that can produce superior hybrids is costly and time-consuming. It requires creation of testcrosses from potential parents and evaluation of the crosses to estimate hybrid performance and combining abilities of inbred lines for the desired traits. Predicting hybrid performance in any way possible may help to reduce the number of crosses to be made and evaluated. The objectives of this study were 1) to determine whether traits measured on parental lines can be used to predict hybrid performance in sorghum; and 2) to assess the combining ability of selected inbreds and determine the relationship between inbred performance and general combining ability (GCA) effects. Forty-six parental inbred lines and 75 F_1 hybrids generated from intercrossing the inbreds were evaluated in four environments. The experiments were laid in a randomized complete block design with three replicates. Data were collected on agronomic characteristics including plant height, days to flowering and grain yield as well as yield components namely panicle length, panicle weight, panicle yield, number of kernels per panicle and thousand kernel weight. Highly significant differences were observed for all traits in both parental inbred lines and hybrids. Generally the hybrids outperformed the parents for all traits. The average performance of the parents (mid-parent performance) was significantly correlated with hybrid performance for thousand kernel weight (r = 0.34), days to flowering (r = 0.55) and plant height (r = 0.57). Correlation for grain yield was positive but not significant. The ability to predict hybrid performance using inbred

line performance varied for the different traits. General combining ability (GCA) for both males and females as well as specific combining abilities (SCA) were also significant for most traits. The highly significant GCA effects observed for most traits and the greater relative importance of GCA as compared to SCA show the significance of additive gene effects in controlling the agronomic traits measured. Results show that information on parental inbred line performance could provide some clue about hybrid performance in sorghum.

Keys words: Sorghum; GCA, General combining ability; SCA, Specific combining ability; Correlation.

Introduction

Sorghum is one of the most important food, feed and fodder crops in the world. It is used as a food crop in most of the developing countries in Africa and Asia, and predominantly as feed grain in the developed countries. Unlike in the developing countries where farmers primarily grow local landraces or open pollinated varieties, this has been completely replaced by the hybrid technology in the commercialized western agriculture for the last six decades. Over the years, a number of sorghum hybrid breeding programs have been initiated in different parts of Africa where sorghum is one of the main sources of food. The programs were specifically aimed at developing hybrids with resistance to prevalent biotic and abiotic stresses, good adaptation to the local environment, good grain quality traits, good threshability, milling recovery and long storage capability (Reddy et al., 2006). A study assessing the relative advantage of growing hybrid sorghum over the local landraces conducted at the National Institute for Agriculture Research (INRAN) in Niger, West Africa showed that on average hybrids yielded 2t/ha, with the best hybrids yielding as high

as 6.5t/ha (Kapran et al., 1997). The national average yield of sorghum largely of local landraces in the same year was 1.53 ton/ha (FAO, 2014). A related study on feasibility of seed production and delivery indicated that hybrid seed business was shown to be more profitable than the open pollinated variety with the hybrid seed being sold at eight times more than the local grain sorghum in 1996 (Kapran et al., 1997).

The high performance of hybrids over their inbred parents (heterosis) attracted the interest of the industry as well as growers and has been successfully exploited to improve productivity. Heterosis or hybrid vigor is the phenomenon whereby the performance of the hybrid exceeds the performance of its inbred parents (East, 1908; Shull, 1908), and it is expressed in terms of increased growth rate, size and yield in the F1 hybrid relative to its inbred parents (Melchinger et al., 1998; Tollenaar et al., 2004). When the performance of the F1 hybrid is better than that of the average performance of its inbred parents, it is referred as mid-parent heterosis (MPH); whereas when the hybrid performance exceeds that of the better parent, it is called better-parent heterosis (BPH).

Heterosis has been exploited in many cereal crops including maize, rice and sorghum, and it is largely responsible for tremendous increase in maize yields in the United States between the 1930's and 1970's (Duvick, 2001). Hybrid breeding program in sorghum started in US in the 1950s, resulting in tremendous increase in yield (Stephens and Holland, 1954; Quinby and Martin, 1954). Heterosis in sorghum is expressed in the form of increased grain and forage yields, early flowering and maturity, increased plant height, larger stems and panicles (Quinby, 1963). Moreover, increased number of seeds per panicle and seed weight have been reported to be responsible for higher grain yield in hybrids (Kambal and Webster, 1966; Blum,

1969). Sorghum is a predominantly self-pollinated crop; the exploitation of the hybrid technology relies on the use of cytoplasmic-genetic male sterility for seed production and fertility restorer sources to regain fertility in hybrid crops. Thus sorghum hybrid breeding programs should run two parallel breeding activities: one for R-line (a fertility-restoring male parent) and the other for seed parent line (a sterile female parent, A-line and a maintainer, B-line) development. This makes sorghum hybrid breeding more challenging than corn and the development of parental lines has to be done without any knowledge of the potential of the inbreds as hybrid parents.

The next more challenging task is to evaluate the value of hundreds or even thousands of the lines developed as parents of commercial hybrid. The traditional approach to do this involves the rigorous task of test cross hybrid synthesis and testing of the hybrids at multiple locations over multiple years to estimate combining abilities of inbred parents for the desired traits. This approach is very time-consuming and costly because it requires a lot of resources, time, labor and finance. Hence, predicting hybrid performance in any way possible may help reduce the number of crosses to be made and hybrids to be evaluated. But trait evaluation at the inbred level has little value if the performance of parental inbred line is not correlated to the hybrid performance (Hallauer and Miranda, 1988). Therefore, any information on parental inbred lines that is indicative of the performance of the hybrid is highly desirable in order to reduce the need for developing a large number of testcross hybrids and conducting extensive multiple-location/year trials.

Results from past studies on relationship between hybrid and inbred performance have been inconsistent. Some have reported low correlations between inbred line and hybrid performance for grain yield in corn (Hallauer and Miranda, 1988) while others reported positive correlations (Flint Garcia et al., 2009; Prado et

al., 2013; Ertiro et al., 2013). Nevertheless, in both cases the correlations reported have been weak especially for more complex traits such as grain yield (Hallauer and Miranda, 1988). Other studies attribute the low correlation between inbred and hybrid performance to genotype-by-environment interaction effect which implies that the effect of environment on inbreds and hybrids may be different (Sadras and Slafer, 2012).

For sorghum no information is available whether there is any positive relationship between the performance of parental inbred lines and their derived hybrids for grain yield and other agronomic traits. Knowledge of such relationship between the parental inbred line performance with that of their hybrids may help to determine whether phenotypic traits expressed in the parents could be used to predict the performance of the derived hybrids in sorghum. The objectives of this study were 1) to determine whether traits measured on parental lines can be used to predict hybrid performance in sorghum; and 2) to assess the general and specific combining abilities of the parents and determine the relationship between inbred performances with their general combining ability.

Materials and methods

Genetic materials

A total of forty-six parental inbred lines and seventy-five derived F1 hybrids were used in this study. This collection included forty-three pollinator parents (R-lines/fertility-restorers) from Kansas State University (KSU) sorghum breeding program and three standard seed parents (AOK11, ATx399 and Tx3042). The three females are seed parents from the U.S public breeding programs .The pollinator parents were crossed to the three standard seed parents in a Design II mating scheme (Hallauer and Miranda, 1988). Some of the pollinators were either too early or too late

compared to the females and hence were not crossed while others produced small amount of seeds not sufficient for multi-location evaluation. Hence, only 75 out of the possible 129 F1 hybrids were evaluated along with three commercial checks (Seneca, Dekalb54-00 and Pioneer84G62).

Experimental design and field management

A randomized complete block design with three replications was used in this study. The plot size was 5m long 2 rows spaced at 0.75m apart. At planting, approximately 3g seeds were directly drilled into each row. Tests were conducted at Kansas State University (KSU) Research Farm near Manhattan, KS during the 2012, 2013 and 2014 summer seasons, and at the North East agricultural experiment station at Ottawa, KS during the 2014 summer season. The hybrids alongside their parental lines were planted on silt loam: fine silty, mixed superactive, mesic cumulic hapludolls soils at KSU Research Farm near Manhattan, KS and on woodson silt loam at North East agricultural experiment station at Ottawa, KS. On average Manhattan location receives annual precipitation of about 907mm (35.7inches) with the average minimum and maximum temperatures of -18°C and 32°C, respectively. At Ottawa location annual precipitation can reach as far as 1000mm (40 inches) on average. The mean annual minimum temperature for this location is -7°C and the annual average maximum temperature is about 32°C. In 2012, 2013 and 2014 at Manhattan location, planting was done on June 8, 7 and 17, respectively, while it was on June 12, 2014 at Ottawa location. The fields were fertilized with ammonium phosphate and ammonium nitrate at the rate of 31.38 kg ha^{-1} P_2O_5 and 112 kg ha^{-1} N, respectively. Weeds were controlled with Bicep Lite II Magnum (Syngenta Crop Prot. LLC) at 0.82 kg ha^{-1} a.i atrazine and 1.03 kg ha^{-1} a.i S-metolachlor and Calisto (Syngenta Crop Prot. LLC) at 0.22 kg ha^{-1} a.i. Mesotrione, applied pre-plant. Hand weeding was used to control

post-emergence weeds throughout the growing period. The average annual rainfall for the period from April to October at Manhattan location was 338mm in 2012, 539mm in 2013 and 576mm in 2014.

Data collection

Data collected included days to flowering (DF), plant height (PH), grain yield and yield components including panicle length (PL), panicle weight (PW), panicle yield (PY), number of kernels per panicle (KN) and thousand kernel weight (TKW).

- DF was determined by recording the number of days from planting to when 50% of plants in each plot reached half bloom.
- PH was recorded by measuring the distance from soil surface to the tip of the panicle at physiological maturity expressed in centimeters.
- Grain yield was measured as the weight of the kernels harvested at maturity from each plot recorded in kilograms per hectare.

After physiological maturity, three panicles from main plants were randomly sampled from each plot for measuring yield components. Mean of the three panicles was used to represent a plot and the moisture content was adjusted 12.5% for statistical analysis.

- PL was determined as the mean length of the panicles measured from the base to the tip of the panicle.
- PW was recorded as the weight of panicle from individual plant.
- PY was measured as the weight of grains threshed from a single panicle.
- KN was recorded by counting the kernels threshed from each panicle using a laboratory seed counter (Seed Counter Model 850-3, International Marketing and Design Corp, 13802 Lookout Road, San Antonio, TX, 78233, USA).

– TKW was determined by measuring the weight of 250 kernels from each panicle and multiplying by four.

Statistical analysis

Analysis of variance (ANOVA) was performed for each trait using PROC GLM of SAS 9.3 (SAS Institute, Cary NC). Pearson correlation analyses between parental and hybrid traits were performed using the procedure CORR in SAS. Broad-sense heritability (H) for each trait was estimated across environments and replicates according to the method by Hallauer et al. (2010).

$$H = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/e + \sigma_e^2/er}$$

where σ^2_g, σ^2_{ge} and σ^2_e are the genetic, genotype-by-environment interaction and residue variance components, respectively. H was estimated for means across environments, r and e are number of replications and environments, respectively. Better-parent heterosis and mid-parent heterosis were calculated for all traits measured using Fehr (1987) method as follows:

$$\text{Better Parent Heterosis (BPH)} = \frac{(F_1 - BP)}{BP} \times 100$$

$$\text{Mid Parent Heterosis (MPH)} = \frac{(F_1 - MP)}{MP} \times 100$$

where F_1 is the performance of the F_1 hybrid, BP is the performance of the better parent involved in a cross and MP is the average performance of the parents involved in the cross. Significance of heterosis estimate was tested using a *t* test at α=0.05: $t = (X_{DHM}-X_{MP})/SE(X_{DHM}-X_{MP})$, where X_{DHM} is the average of all derived hybrid means, X_{MP} is the average of all mean parental values and SE is the standard error of the difference between averages (Prado et al., 2013).

General combining ability (GCA) and specific combining ability (SCA) effects were analyzed using SAS version 9.3 (SAS Institute, Cary NC). Entry, male, female, and male × female interaction effects were determined for the combined data as well as for each environment. Entry, replications, environment, and their interactions with all other factors were treated as random effects, whereas female effects were treated as fixed. Random variables were specified using RANDOM statement in the GLM procedure. Entry effects were partitioned into inbred and hybrid components, and into inbred versus hybrid components. The hybrid effect was further partitioned into male, female, and male × female interaction effects representing general combining ability (GCA) for male, GCA for female, and specific combining ability (SCA) effects, respectively. The effects were tested for all parameters using appropriate error terms specified by the TEST option in the GLM. The GCA for each parental inbred line was calculated as the difference between the mean performance of the progeny of a given parental inbred line and the overall mean of the hybrids. A two-tailed t-test in SAS version 9.3 was used to determine the significance of GCA for each parent and was confirmed using the procedure outlined by Cox and Frey (1984) and Kearsey and Pooni (1996).

Results

Analysis of variance, performance of parental inbred lines and derived F1 hybrids

The analysis of variance (ANOVA) for DF, PH, GY and TKW is presented in Table 4.1. The genotype and genotype-by-environment interaction effects were highly significant for all traits in both the parents and the hybrids (Table 4.1). This suggests that the performance of the genotypes were variable depending on the environment,

thus there was differential response of the genotypes with respect to environment. The male, female and male × female interaction effects for these and the rest of the traits were highly significant indicating significant GCA for both male and female parents as well as significant SCA for all of the traits.

Results of the performance of the parental lines and their hybrids across environments for eight agronomic traits is presented in Table 4.2. All traits differed significantly among the parents and the hybrids. The hybrids outperformed the parents for all traits (Table 4.2). On average, the hybrids were 5.9, 28, 26.6 and 25.8% higher for PL, PW, PY and KN, respectively than the parents. Similarly, the hybrids flowered on average 9 d earlier than the parents and were 8.98 cm taller. Again the hybrids yielded 2900.9 kg ha^{-1} which is 44.1% more than the parents (Table 4.2). On average, the hybrids also had slightly higher TKW (3.7%) than the parents with the parents having 26.8 g compared to 27.8 g in the hybrids (Table 4.2). The range for TKW was 20.4 to 30.5g among the inbred parents and 23.9 to 31.3g among the hybrids. Inbred parents had relatively shorter stature ranging from 92.2 to 131.7cm with an average of 113.2 cm as compared to 106.3 to 138.5cm for the hybrids that averaged 122.2cm. Among the parents, DF ranged from 59 to 83 d, while the range was from 56 to 74 d in the hybrids. Similarly, grain yield was different between the parents and the hybrids with an average of 6572.1 kg ha^{-1} for the parents vs. 9473 kg ha^{-1} for the hybrids. The grain yield ranged from 4116 to 8574 kg ha^{-1} and from 6824 to 12443 kg ha^{-1} among the parents and the hybrids, respectively.

Table 4.1 Mean squares of the combined analysis of variance of four agronomic traits of sorghum evaluated at Manhattan and Ottawa, KS during 2012, 2013 and 2014 summer seasons.

Parental inbred lines					
Source of variation	*df*	Thousand kernel weight (g)	Plant height (cm)	Days to flowering	Grain yield (kg ha^{-1})
Environment (E)	3	579.1	36647.5	2284.9	664614007
Replication/E	9	34.2	102.3	74	6386730
Genotype (G)	48	48.9***	836.8***	380.9***	12378120***
G x E	142	22.57***	236.6***	111.3***	1148335***
Error	385	7.4	40.5	33.2	3161195
F1 hybrids					
Environment (E)	3	406.8	42174.8	800.4	403079069
Replication/E	8	21.1	182.7	36.3	11904593
Genotype (G)	72	29.1***	702.2***	103.8***	26091760***
Male (M)	8	123.6**	1446.8***	141.3**	317786.6**
Female (F)	2	6.0*	145.1**	69.2**	316891.0**
M x F	16	8.0**	234.9***	88.2**	9678747.1**
G x E	218	21.2***	363.1***	33.2***	15735220***
M x E	8	7.4*	209.3**	28.8	3166450.4
F x E	2	7.6	33.5	11.5	1250425.0
M x F x E	16	9.8***	242.9**	32.1	5869622.3
Error	576	7.3	71.6	12.6	3943385

*, ** and *** Statistically significant at $p \leq 0.05$, 0.01 and 0.001 levels of probability, respectively

Table 4.2 Across environment performance of parental inbred lines and their derived F1 hybrids evaluated at Manhattan and Ottawa, KS during 2012, 2013 and 2014 summer seasons.

	Parental inbred lines			Hybrids			Male parents		Female parents	
Trait	Mean	Range	H	Mean	Range	H	Mean	Range	Mean	Range
PL	28.3(±2.8)	23.6-35.1	0.87	30.0(±1.4)	25.0-30.8	0.27	28.6	27.5-35.1	24.5	23.6-26.2
PW	63(±9.1)	40.6-82.4	0.54	81.5(±7.1)	65.6-101.2	0.10	60.6	40.6-82.4	65.4	59.9-71.4
PY	40.0(±5.3)	26.7-52.2	0.19	50.6(±4.2)	43.1-61.0	0.19	38.9	26.7-52.2	41.8	37.4-46.6
KN	1461(±171)	1039-1874	0.21	1838(±153)	1471-2222	0.29	1444	1039-1727	1493	1276-1874
TKW	26.8±2	20.4 -30.5	0.49	27.8±1.5	23.9-31.3	0.13	26.2	20.4-30.5	28.4	27.1-29.1
PH	113.2±8	92.2-131.7	0.67	122.2±7.6	106.3-138.5	0.39	115.5	98.4-131.7	98.3	92.2-103.8
DF	73±6	59-83	0.66	64±2.9	59-74	0.40	73.4	59-83	67	64-73
GY	6572.1±1004	4116.4-8574	0.10	9473±1324	6824-12443	0.37	6762.3	4116.4-7956	7222	5700.5-8574.1

PL= Panicle length (cm); PW= Panicle weight (g); PY=Panicle yield (g); KN=Number of kernels per panicle; TKW=Thousand kernel weight (g); PH=Plant height (cm); DF=Days to flowering; GY=Grain yield (kg ha^{-1}); H = Broad-sense heritability; Numbers in parentheses are standard deviations.

Heterosis among traits for F1 sorghum hybrids

The mid-parent (MPH) and better-parent heterosis (BPH) results for all traits across environments are presented in Table 4.3. For most traits, heterossis ranged from negative values to higher positive values with overall average of positive heterosis except both MPH and BPH for DF and BPH for PL and TKW were negative (Table 4.3). MPH for PL was between -7.7 to 16.1 with an average of 4.6%. The range for BPH for the trait was from -20.5 to 8.5% but averaged lower (-2.5%). Average MPH for PW ranged from low negative of -1.9 to 61.6% with a positive mean of 25.6%. Similarly, the BPH values stretched from -8.2 to 50.3% with a mean of 18.2%. The mean MPH for PY and KN were 22.6 and 22.4%, respectively and mean BPH were 15 and 14.4%. Both PY and KN had a wider range for MPH than their BPH. TKW had the lowest mean positive MPH. Both MPH and BPH had narrower ranges of all the traits measured (Table 4.3). For PH, the mean MPH and BPH were 14.9 and 6.9%, ranging from 0.8 to 28.8 and from -10.5 to 20.5%, respectively. For DF, the average MPH and BPH were -8.5 and -2.5%, in that order. The range was from -14.8 to 11.6% for MPH and from -19.3 to 15.6% for BPH. Nevertheless, grain yield had the highest MPH (38.1%) and BPH (24.6%) of all the traits studied. The mid-parent heterosis (MPH) for grain spanned from low negative of -2.1% to a high of 93.0% and the BPH range was also wide (-20.4 to 67.5%).

Disaggregation of heterosis by the female parent allows a closer look to the dynamics of heterosis. Table 4.4 depicts BPH and proportion of crosses exhibiting positive BPH as sorted by female parents. While the overall trend is similar for all females, the degree of heterosis can be different based on the *per se* performance of the females especially whenever they are regarded as better parent which appears to be the case. While MPH was overwhelmingly positive for almost all traits in all

females, the BPH, a stronger indicator of hybrid performance, is quite different. For grain yield, the BPH was comparable for OK11 and Tx3042 females with an average of 28.9% and 29.0%, respectively, while it was only 13% for the Tx399 female. Likewise the proportion of hybrids with positive BPH were different between the females with OK11 and Tx3042 having 97% and 100% of their hybrids outperforming both of parents while only 70% of hybrids of Tx399 beating both parents (Table 4.4). Similarly, BPH for TKW were almost similar for OK11 and Tx3042 females. The mean BPH for Tx399 hybrids was -2.0 with the highest proportion of hybrids exhibiting positive BPH followed by OK11 and Tx3042 hybrids. All three females were different for DF with OK11 having BPH of -8.1 as compared to 5.0 for Tx3042 and -3.7 for Tx399 with 100% of the hybrids of OK11 being earlier than both of parents while all of the hybrids of Tx3042 were later than the earlier flowering parent. For Tx399 female about 50% of the hybrids were earlier than the earliest parent. On the other hand mean BPH for TKW was negative for all females with only 36, 24 and 65% of the hybrids for OK11, Tx3042 and Tx399, respectively, having positive better parent heterosis.

The correlation between inbred line performance and hybrid vigor was positive for all traits in all female parents except for TKW for hybrids developed using ATx3042 as female parent and grain yield for the hybrids developed using OK11 and Tx3042 (Table 4.4). However, only those PH and DF for the OK11 hybrids were significant with coefficients of $r = 0.45$ for PH and $r = 0.63$ for DF (Table 4.4). Correlation for grain yield, TKW, PL, PW and PY being moderate with coefficients of $r = 0.32$, $r = 0.27$, $r = 0.64$, $r = 0.35$ and $r = 0.39$, respectively, indicating that selection of these traits in inbred parents though not effectively predict heterosis for the traits may be of some benefit for enhancing hybrid performance.

Table 4.3 Across environment mean mid-parent heterosis (MPH), better-parent heterosis (BPH) and percent number of hybrids with positive heterosis for eight agronomic traits of sorghum [*Sorghum bicolor* (L.) Moench].

Trait	Heterosis (%)					
	Mid-parent heterosis (MPH)			Better-parent heterosis (BPH)		
	Mean	Range	£ Hybrids with "+" heterosis	Mean	Range	£ Hybrids with "+" heterosis
Panicle length (cm)	4.6(±4.9)	-7.7 – 16.1	83.0	-2.5(±4.9)	-20.5 – 8.5	41.4
Panicle weight (g)	25.6(±13.7)	-1.9 – 61.6	98.6	18.2(±12.7)	-8.2 – 50.3	92.9
Panicle yield (g)	22.6(±14.4)	-2.2 – 69.5	97.1	15.0(±14.6)	-7.4 – 48.4	87.1
Number of kernels panicle^{-1}	22.4(±12.5)	-3.4 – 62.0	97.1	14.4(±12.3)	-8.4 – 46.9	87.1
Thousand kernel weight (g)	1.6 (±6.0)	-11.4-17.6	55.7	-1.04(±6.8)	-14.5-26.0	34.0
Plant height (cm)	14.9(±5.9)	0.8-28.8	100.0	6.9(±7.9)	-10.5-20.5	76.0
Days to flowering	-8.5(±4.3)	-14.8-11.56	4.0	-2.5(±6.9)	-19.3-15.6	39.0
Grain yield (kg ha^{-1})	38.1(±21.5)	-2.1-93.0	97.0	24.6(±19.4)	-20.41-67.54	91.0

Numbers in parentheses are standard deviations; £ = proportion of hybrids exhibiting positive heterosis for the trait.

Table 4.4 Mean better-parent heterosis (BPH), percent number of hybrids with positive better-parent heterosis and pearson correlation coefficients (*r*) for four agronomic traits of sorghum hybrids developed using different seed parents (i.e AOK11, ATx3042 and ATx399).

	AOK11			ATx3042			ATx399		
Trait	Mean	Hybrids with "+" BPH[a]	Inb-Hybr corr (*r*)	Mean	Hybrids with "+" BPH[a]	Inb-Hybr corr(*r*)	Mean	Hybrids with "+" BPH[a]	Inb-Hybr corr(*r*)
Thousand kernel weight (g)	-0.6(±8.4)	36	0.22	-0.8(±6.9)	24	-0.16	-2.0(±3.2)	65	0.27
Plant height (cm)	9.6(±6.9)	88	0.45**	10.6(±5.6)	94	0.29	-0.8(±5.6)	40	0.26
Days to flowering	-8.1(±4.4)	0	0.63**	5.0(±3.7)	0	0.32	-3.7(±4.5)	50	0.03
Grain yield (kg ha^{-1})	28.9(±18.6)	97	-0.06	29.0(±18.9)	100	-0.01	13(±17.3)	70	0.32

* and ** Significant at $p \leq 0.05$ and $p \leq 0.01$, respectively; [a] values in this column refer to the proportion of hybrids that express positive better-parent heterosis; Inb-Hyb refers to the correlation between inbred line performance with hybrid performance; Numbers in parentheses are standard deviations.

Correlations among traits within parental inbreds and hybrids

Pearson correlation coefficients among agronomic traits for parental inbred lines and their hybrids are presented in Table 4.5. Many of the traits were correlated with each other both in the parents and hybrids. Among the inbred parents, PL was not significantly correlated with grain yield and yield components but was positively and significantly correlated with PH (r = 0.41) and DF (r = 0.64). Similarly, PW was significantly correlated with PY (r = 0.83), KN (r = 0.79) and TKW (r = 0.44) but not significantly correlated with grain yield, PH and DF. PY on the other hand was significantly correlated with KN (r = 0.85) and TKW (r = 0.51), but not correlated with the other traits. All other correlations were not significant except between DF and PH (Table 4.5).

Contrary to the inbreds, correlation between PL and all yield components and agronomic traits was significant. Accordingly, PL positively and significantly correlated with PW (r = 0.37), PY (r = 0.23) and KN (r = 0.45). It was also significantly correlated with grain yield (r = 0.23) as well as with PH (r = 0.62) and DF (r = 0.31). PL also significantly but negatively correlated with TKW (r = 0.39) (Table 4.5). Likewise, PY and KN were significantly correlated with each other and all other traits measured except correlation of PH and DF with PY and DF with KN were not significant. PW was positively and significantly correlated only with PL, PY and KN. TKW was negatively correlated with PH and all other remaining correlations were not significant.

Table 4.5 Pearson correlation coefficients (r) among eight agronomic traits in the parental inbred lines (above diagonal) and hybrids (below diagonal) evaluated at Manhattan and Ottawa, KS during 2012, 2013 and 2014 summer seasons.

Trait	Panicle length (cm)	Panicle weight (g)	Panicle yield (g)	Number of kernels panicle^{-1}	Thousand kernel weight (g)	Plant height (cm)	Days to flowering	Grain yield (g)
Panicle length (cm)	-	0.10	-0.04	0.01	-0.10	0.41**	0.64***	0.12
Panicle weight (g)	0.37**	-	0.83***	0.79***	0.44**	-0.27	-0.10	0.06
Panicle yield (g)	0.23*	0.66***	-	0.85***	0.51**	-0.19	-0.20	0.22
Number of kernels panicle^{-1}	0.45***	0.66***	0.79***	-	0.08	-0.32	-0.11	0.23
Thousand kernel weight (g)	-0.39**	-0.02	0.25*	-0.26*	-	-0.14	-0.32	0.01
Plant height (cm)	0.62***	0.12	0.13	0.34**	-0.34**	-	0.31*	-0.12
Days to flowering	0.31*	0.07	0.09	0.21	-0.18	0.01	-	0.003
Grain yield (kg ha^{-1})	0.23*	0.14	0.34*	0.29*	-0.11	-0.08	0.16	-

*,** and *** Significant at $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, respectively.

Correlations between mid-parent performance and hybrid performance, mid- and better-parent heterosis

Across environment pearson correlation between the average performance of parents (mid-parent performance) and hybrid performance, MPH and BPH for eight agronomic traits are presented in Tables 4.6-4.9. The mid-parent performance had significant and positive correlation with hybrid performance for PL (r = 0.52), DF (r = 0.55), PH (r = 0.57) and TKW (r = 0.34) (Tables 4.6, 4.8 and 4.9) but not correlated with the other traits. However, mid-parent performance was negatively and significantly correlated with MPH for PL (r = -0.49), PW (r = -0.61), PY (r = -0.69), KN (r = -0.60), DF (r = -0.57), TKW (r = -0.49) and grain yield (r = -0.47) (Tables 4.6-4.9). Correlation with PH was not significant (Table 4.8). Similarly, mid-parent performance was significantly correlated with BPH for PL (r = -0.59), PW (r = -0.41), PY (r = -0.54), KN (r = -0.49), DF (r = -0.64), TKW (r = -0.46) and grain yield (r = -0.30) (Tables 4.6-4.9).

Moreover, hybrid performance was significantly and positively correlated with MPH for PL (r = 0.49), PW (r = 0.69), PY (r = 0.76), KN (r = 0.65), DF (r = 0.37), PH (r = 0.81), TKW (r = 0.65) and grain yield (r = 0.76) (Tables 4.6-4.7). Similarly, the hybrid performance had positive correlation with BPH for PL (r = 0.25), PW (r = 0.73), PY (r = 0.78), KN (r = 0.62), PH (r = 0.65), TKW (r = 0.34) and grain yield (r = 0.77) (Tables 4.6-4.9).

Table 4.6 Pearson correlation coefficients (*r*) among hybrid performance, mid-parent value, mid-parent heterosis and better-parent heterosis for panicle length (above diagonal) and panicle weight (below diagonal).

Trait	F1 hybrid	Mid-parent value	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.52**	0.49**	0.25*
Mid-parent value	0.15	-	-0.49**	-0.59**
Mid-parent heterosis	0.69**	-0.61**	-	0.85**
Better-parent heterosis	0.73**	-0.41**	0.88**	-

*, ** Significant at $p \leq 0.05$, 0.01, respectively; Mid-parent value = the average performance of the parents

Table 4.7 Pearson correlation coefficients (*r*) among hybrid performance, mid-parent value, mid-parent heterosis and better-parent heterosis for panicle yield (above diagonal) and number of kernels per panicle (below diagonal).

Trait	F1 hybrid	Mid-parent value	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	-0.07	0.76**	0.78**
Mid-parent value	0.22	-	-0.69**	-0.54**
Mid-parent heterosis	0.65**	-0.60**	-	0.91**
Better-parent heterosis	0.62**	-0.49**	0.89**	-

** Significant at $p \leq 0.01$; Mid-parent value= the average performance of the parents

Table 4.8 Pearson correlation coefficients (*r*) among hybrid performance, mid-parent value, mid-parent heterosis and better-parent heterosis for days to flowering (above diagonal) and plant height (below diagonal).

Trait	F1 hybrid	Mid-parent value	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.55**	0.37**	-0.04
Mid-parent value	0.57**	-	-0.57**	-0.64**
Mid-parent heterosis	0.81**	-0.03	-	0.66**
Better-parent heterosis	0.65**	-0.18	0.91**	-

** Significant at $p \leq 0.01$; Mid-parent value= the average performance of the parents.

Table 4.9 Pearson correlation coefficients (*r*) among hybrid performance, mid-parent value, mid-parent heterosis and better-parent heterosis for thousand kernel weight (above diagonal) and grain yield (below diagonal).

Trait	F1 hybrid	Mid-parent value	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.34**	0.65**	0.34**
Mid-parent value	0.21	-	-0.49**	-0.46**
Mid-parent heterosis	0.76**	-0.47**	-	0.68**
Better-parent heterosis	0.77**	-0.30*	0.89**	-

*, ** Significant at $p \leq 0.05$, 0.01, respectively; Mid-parent value = the average performance of the parents

General (GCA) and specific combining ability (SCA)

The combined analysis for four major agronomic traits (i.e TKW, PH, DF and grain yield) was performed on 27 hybrids tested in two environments. The analysis of variance for the combined data is presented in Table 4.10. The analysis showed highly significant entry and entry × location interaction effects for all traits measured. Partitioning of the entry effect into parental inbred and hybrid components also revealed significant effects for all traits except for TKW, DF and grain yield for hybrid effect. The inbred vs. hybrid component was also significant for all the traits (Table 4.10).

Further partitioning of the hybrid effect into male, female and male × female interaction components also revealed significant effect for all components and all traits (Table 4.10). This indicates that GCA for male and female, and SCA effects have significantly impacted all traits. However, from the mean squares presented in Table 4.10, it was clear that male parents significantly contributed for much of the variation observed among the hybrids. Comparison of marginal means for both male and female for four major traits show the magnitude of the variation among male and female parents. For DF, variation among the females ranged from 62 to 65 d as compared to 60 to 65 d for males. For PH, the range was 114.6 to 123cm for males and 113.8 to 122.8 among females. Similarly variation for grain yield among the males is relatively large with the lowest yield being 6578 kg ha^{-1} in PR11/12-1435 and the highest (9876 kg ha^{-1}) recorded in PR11/12-526. Among the females the range was 8.4 to 9.9t/ha (Tables 4.11). Similarly, the difference for TKW among the females ranged only from 25.7 to 28.4g as compared to males where wider range of 24.7 to 29.8g was recorded. Variation for all other yield components including PL,

PW, PY and KN was consistently higher among the males than females (data not shown)

Table 4.10 Mean squares from the combined analysis of variance for four major agronomic traits of sorghum [*Sorghum bicolor* (L.) Moench] genotypes evaluated at Manhattan and Ottawa during the 2014 summer season.

Source of variation	*df*	Thousand kernel weight (g)	Plant height (cm)	Days to flowering	Grain yield (kg ha^{-1})
Location (L)	1	0.06	18287.8***	2312.8***	353901636.7***
Location (Rep)	4	8.9	27.1	19.7	7600739.2
Entry	38	20.7**	716.7***	175.6***	18346303.6***
Inbred (I)	11	25.8*	139.9*	846.2***	9528474.6**
Hybrid (H)	26	16.8	86.5**	300.5	9327346
Female (F)	2	6.0*	145.1**	69.2**	316891.0**
Male (M)	8	123.6**	1446.8***	141.3**	317786.6***
F × M	16	8.0**	234.9***	88.2**	9678747.1**
H vs I	1	65.5*	10113.2***	2885.2***	349835313***
Entry × L	38	16.6*	212.9***	175.6***	18346303.6***
I × L	11	26.0	202.8*	55.5	2957326
H × L	26	8.9***	216.2***	29.5	5548233
F × L	2	7.6	33.5	11.5	1250425
M × L	8	7.4*	209.3**	28.8**	3166450.4***
M × F × L	18	10.7	48.0	11.7	10246824.1*
I vs H × L	1	112.9**	231.3	276.4*	50657.7*
Error	226	12.6	166.4	46.8	5173625
CV		12.8	11.2	10.4	26.9

*,**, and *** Significant at $p \leq 0.05$, 0.01 and 0.001, respectively.

Correlation between performance of parental inbred line and GCA estimates

The means for four agronomic traits and GCA effects are presented in Table 4.11. Among the female parents, AOK11 was the best general combining female parent for DF and grain yield with significant and positive GCA effects of 2 d and 440.7 kg ha^{-1}, respectively (Table 4.11). ATx3042 was the best general combining female parent for PH with significant and positive GCA effect of 3.5cm. As for the male parents, PR11/12-564 was the best general combining male parent for PH and PR11/12-526 for grain yield. The correlation between GCA effects and inbred line performance varied among the traits. The inbred line performance was significantly and positively correlated with GCA effects for TKW ($r = 0.97$, $p < 0.01$) and grain yield ($r = 0.83$, $p < 0.01$) (data not shown)

Table 4.11 Mean and general combining ability (GCA) of sorghum [*Sorghum bicolor* (L.) Moench] parental lines for four agronomic traits evaluated at Manhattan and Ottawa, KS during 2014 summer season.

	Thousand kernel weight (g)		Days to flowering		Plant height (cm)		Grain yield(kg/ha)	
Female parents	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA
AOK11	25.7	-1.7*	65	2.00**	121.8	2.50*	9915.9	440.7**
ATx3042	28.1	0.7	62	-1.0	122.8	3.50*	8425.1	-1050.1*
ATx399	28.4	1.0	63	0	113.8	-5.5*	9484	8.8*
Mean	27.4	-	63	-	119.5	-	9275.2	-
LSD	0.7	-	1.8	-	3.5	-	605.0	-
Male parents								
PR11/12-1420	28.0	0.6	65	2.0*	117.9	-1.4*	9606.9	131.7**
PR11/12-1426	24.7	-2.7*	62	-1.0	122.5	3.2*	9398.8	-76.4
PR11/12-1435	27.3	-0.1	60	-3.0*	114.6	-4.7*	6578	-2897.2**
PR11/12-505	26.8	-0.6	64	1.0	121.4	2.1*	9031	-444.2**
PR11/12-526	27.3	-0.1	65	2.0*	117.7	-1.6*	9876.3	401.1**
PR11/12-533	27.3	-0.1	65	2.0*	118.7	-0.6	8866.8	-608.4**
PR11/12-564	27.2	-0.2	63	0	122.6	3.3*	8747.1	-728.1*
R-45	29.8	-2.4*	65	2.0*	121.7	2.4*	9798.9	323.7*
Tx2737	27.4	0	62	-1.0	116.8	-2.5*	9649	173.8*
Mean	27.3	-	64	-	119.3	-	9061	-
LSD	1.2	-	1.9	-	2.8	-	240.5	-

* and ** Significant at $p \leq 0.05$ and 0.01, respectively; GCA= General combining ability; LSD= Least significant difference.

Discussion

The significance of heterosis in sorghum hybrid breeding programs cannot be overemphasized. Over the years, hybrid improvement based on exploiting heterosis has contributed significantly to the enormous increase in grain and forage yields. However, practical selection of parental inbred lines that can produce superior hybrids is a challenge because testing all potential parental lines in hybrid combination is very expensive and time-consuming. The current procedure involves sampling of the parents, creation of testcross hybrids for field evaluation in multiple locations and years to assess hybrid performance and combining ability of the parents for the desired traits. Then all lines that produced the most promising hybrids are included in further test crosses and their hybrids evaluated at multiple environments. Hybrids in the order of thousands are evaluated this way to come up with few most promising hybrids for commercial consideration. These make hybrid breeding very expensive and a time consuming task. The present study was conducted to determine whether agronomic traits expressed in the parental inbred lines are transferable to their hybrids and determine if inbred line performance can be used as predictor of hybrid performance in sorghum.

In this study, tests conducted across environments showed hybrids exhibiting variable levels of both MPH and BPH for almost all traits studied (Tables 4.3 and 4.4). Across hybrids, average estimates of MPH for different traits was different with TKW showing the least heterosis of 1.6% and grain yield being 38.1%. The range for grain yield, however, was from -2.1 to 93% which is wider than previously reported for grain yield in sorghum (Kirby and Atkins 1968; Liang and Walter 1968, 1969). Better-parent heterosis also had similar trend with maximum heterosis of 67% (Table

4.3) which was also higher than previously reported for sorghum (52.1%) and maize (50.5%) (Bunphan et al., 2015; Flint-Garcia et al., 2009). Furthermore, the sorghum hybrids exhibited MPH and BPH in almost every trait. For majority of the traits measured, positive MPH and BPH was observed in over 80% of the hybrids evaluated. Similar results were reported in maize (Flint-Garcia, 2009) that most of the traits studied exhibited BPH in over 90% of the hybrids tested and at higher levels compared to other crop species. However, contrary to previous observation that heterosis is prevalent in reproductive traits (Flint-Garcia et al., 2009), some of such traits in the current study such as thousand kernel weight had lower heterosis while other reproductive traits such grain yield remained high. It is not clear why heterosis is so variable for different traits. This may has to do with the fact that some of these traits are interrelated and can be affected by increase or decrease in heterosis for other traits. For example, the low heterosis for TKW in this study can be explained by the effect of heterosis on kernel number which is about 26% (Table 4.3) that even if photosynthetic efficiency was improved by that much, then heterosis for TKW will still remain low. Moreover, this agrees with the general fact that excessive increase in grain size which contributes to TKW is not common not only in sorghum but also in other hybrid crops.

Moreover, some of the traits measured in the parents were shown to have significant correlation with those measured in the hybrids which suggests that selecting those traits in the parental lines may enhance prediction of hybrid performance. For example, we observed that PW was significantly and positively correlated with PY and KN in both the hybrids and parents. Additionally, correlation analyses showed that the average performance of the parental inbred lines (mid-parent value) was significantly and positively correlated with the hybrid performance for DF

(r = 0.55), plant height (r = 0.57) and TKW (r = 0.34). These results suggest that selection of early flowering parental lines may enhance the performance of their hybrids with regard to flowering time. In addition, the correlation between mid-parent performance and hybrid performance for PH was moderately positive indicating that tall parental lines generally tend to produce tall hybrids, which means that the performance of the hybrid for PH can somewhat be predicted based on the stature of the parents. From a genetic point of view, the positive correlation between hybrids and inbred lines indicates that a large amount of additive gene action is affecting the derived hybrid performance (Rojas and Sprague, 1952). Although the correlation found in this study for TKW was not strong enough to be a predictor of hybrid performance, it appears the parents with large seed size (high TKW) tend to produce hybrids with large seed size as well. However, no significant correlation (r = 0.21, p = 0.09) was observed between mid-parent and hybrid performance for grain yield, indicating that high yielding parents do not automatically produce high yielding hybrids. These results corroborate with previous findings (Ertiro et al., 2013; Samanci 1996). Ertiro et al. (2013) used maize inbred lines and their hybrids to study the relationship between the parental inbred line performance and hybrids for food-feed related traits in maize (*Zea mays* L.). In their study, they found no relationship between mid-parent and hybrid performance for grain yield (r = 0.18, p = 0.16).

Furthermore, we found negative correlation between the average performance of the parents (mid-parent performance) with both MPH and BPH for all traits except PH (Tables 4.3 and 4.4). It appears that as the average performance of the parents increases, the mid-parent value tends to be closer or similar to the hybrid value, and hence reducing MPH. Similarly, as the performance of the better-parent increases, the hybrid value for the trait tends to be closer or equal to the better-parent value, thus

reducing the difference between hybrid and better-parent values resulting in low BPH. Similar results have been reported in maize (Flint-Garcia et al., 2009).

Over the years, the value of parental line for a particular trait of interest has been assessed by analysis of its GCA effects. However, the assessment of GCA effects is very expensive and time-consuming as it requires the formation and evaluation of the crosses at multiple locations and years. Therefore, the correlation between inbred line performance and its GCA effects may help to predict the performance of the hybrids without extensive field evaluation. Ideally, GCA effect of a line shows its potential for generating superior hybrids, implying that high GCA effect indicates that the parental mean is greater than the grand mean, suggesting not only strong evidence of favorable flow of genes from parents to offspring at high frequency but also indicates the concentration of predominantly additive gene effects. Besides that GCA and SCA effects are indicators of relative importance of additive and non-additive gene action, they may also help to determine the breeding procedure to be used to improve the performance of the desired traits.

Therefore, in this study, we further used some parental inbred lines that had all male and female combinations to study the GCA and SCA effects of the parents for all the traits measured, and also investigate whether there is any relationship between inbred line performance *per se* and GCA effects. Significant GCA mean squares for most traits (Tables 4.10 and 4.11) were observed, highlighting the importance of the additive contribution of genes underlying major agronomic traits on hybrid performance. In addition, significant and positive correlations between inbred line performance *per se* and GCA effects were observed for almost all traits, indicating

that traits are under additive gene action in the parents and may be used as predictors for hybrid performance or heterosis in sorghum.

Conclusion

This study was conducted to determine whether agronomic traits measured on parental inbred lines could be used to predict the performance of their hybrids in sorghum. As expected, the hybrids outperformed their parents in all traits. Most of the traits studied expressed both mid-parent and better-parent heterosis in almost all hybrids evaluated. Correlations between mid-parental inbred line performance and single cross hybrid performance for all traits were significant with low to medium correlation values. The ability to predict hybrid performance using inbred line performance varied for the different traits. Results show that studying parental inbred line performance could generate important information for predicting hybrid performance in sorghum.

References

Blum, A. 1969. Nature of heterosis in grain production by the sorghum panicle. Crop Sci. 10:28-31.

Bunphan, D., P. Jaisil, J. Sanitchon, J.E. Knoll, and W.F. Anderson. 2015. Heterosis and Combining Ability of F1 Hybrid Sweet Sorghum in Thailand. Crop Sci. 55(1): 178-187.

Cox, D.J., and K.J. Frey. 1984. Combining ability and the selection of parents for interspecific oat mating. Crop Sci. 24:963-967.

Duvick, D.N. 2001. Biotechnology in the 1930s: The development of hybrid maize. Nat. Rev. Genet. 2:69-74.

East, E.M. 1908. Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. 1907:419-428.

Ertiro, B.T., H, Zeleke, D. Friesen, M. Blummel, and S.T. Afriyie. 2013. Relationship between the performance of parental inbred lines and hybrids for food-feed traits in maize (*Zea mays* L.) in Ethiopia. Field Crops Res.153: 86-93.

Fehr, W.R. 1987. Principle of Cultivars Development. Macmillan publishing company. A division of Macmillan Inc. New York pp.1: 1- 465.

Flint-Garcia, S.A., E.S., Buckler, P. Tiffin, E. Ersoz and N. M. Springer. 2009. Heterosis Is Prevalent for Multiple Traits in Diverse Maize Germplasm. PLoS One 4(10):e7433.

Hallauer, A.R., M.J. Carena, and J.B. Miranda Filho. 2010. Quantitative Genetics in Maize Breeding, 3rd ed. Handbook of Plant Breeding Volume 6. Springer, New York. 663 pages.

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative Genetics in Maize Breeding, 2nd ed. Iowa State University Press, Ames, IA. SB191.M2 H29.

Holland, J.B., D.V. Uhr, D. Jeffers, and M.M. Goodman. 1998. Inheritance of resistance to southern corn rust in tropical-by-corn-belt maize populations. Theor. Appl. Genet. 96:232-241.

Kambal, A.E., and O.J. Webster. 1966. Manifestation of hybrid vigor in grain sorghum and relations among the components of yield, weight per bushel and height. Crop Sci. 6:513-516.

Kapran, I., M. Amadou, M. Abdou, S. Souley, N. Kondo, J.D. Axtell, G. Ejeta, and T.Tyler. 1997. Heterosis and prospects for marketing sorghum hybrids in Niger. In Book of Abstracts. The Genetics and Exploitation of Heterosis in Crops. An International Symposium, CIMMYT, Mexico, pp.17-22.

Kearsey, M.J., and H.S. Pooni. 1996. The Genetical Analysis of Quantitative Traits. Chapman and Hall: London.

Kirby, J.A., and R.E. Atkins. 1968. Heterotic response for vegetative and mature plant characters in grain sorghum [*Sorghum bicolor* (L.) Moench]. Crop Sci. 8:335-339.

Liang, G.H., and T.L. Walter. 1968. Heritability estimates and gene effects for agronomic traits in grain sorghum, *Sorghum vulgate* Pers. Crop Sci. 8(1):77-81.

Liang, G.H., C.B. Overley, and A.J. Casady.1969. Interrelations among agronomic characters in grain sorghum [*Sorghum bicolor* (L.) Moench]. Crop Sci. 9(3): 299-302.

Melchinger A.E., H.F. Utz, and C.C. Schon.1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genet.149:383-403.

Prado, S.A., L.G. Brenda, A. D. Novoa, D. Foster, M. L. Senior, C. Zinselmeier, M.E. Otegui, and L. Borrás. 2013. Correlations Between Parental Inbred Lines and Derived Hybrid Performance for Grain Filling Traits in Maize. Crop Sci. 53:1636-1645.

Reddy, B.V., H.C. Sharma, R.P. Thakur, S. Ramesh, F. Rattunde, and M. Mgonja. 2006. Sorghum hybrid parents research at ICRISAT-strategies, status and impacts. J. SAT Agric. Res. 2(1): 1-24.

Rojas, B.A., and G.F. Sprague. 1952. A comparison of variance components in corn yield trials: III. General and Specific combining ability and their interaction with locations and years. Agron. J. 44:462-466.

Sadras, V.O., and G.A. Slafer. 2012. Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. Field Crops Res. 127:215-224.

Samanci, B.1996. Phenotypic correlations between maize inbreds and their single cross hybrids in short season areas. Euphytica 89: 291-296.

SAS Institute. 2011. The SAS system for Windows. v.9.3. SAS Inst., Cary, NC.

Shull, G.H. 1908. The composition of a field of maize. Rep. Am. Breed. Assoc. 4:296-301.

Stephens, J.C., and R.F. Holland. 1954. Cytoplasmic male sterility for hybrid sorghum seed production. Agron. J. 46:20-23.

Tollenaar, M., A. Ahmadzadeh, and E. A. Lee. 2004. Physiological basis for grain yield improvement in maize. Crop Sci. 44: 2086-2094.

Quinby, J.R. 1963. Manifestation of hybrid vigor in sorghum. Crop Sci. 3:288-291.

Quinby, J.R. and J.H. Martin. 1954. Sorghum improvement. Adv. Agron. 6:305-359.

Chapter 5 - Genomic Prediction of Hybrid Performance Based on Hybrid Phenotype and Inbred Genotype in Sorghum [*Sorghum bicolor* (L.) Moench]

Abstract

Genomic selection is a new breeding method in which genome-wide markers are used to predict the phenotypes of untested lines thus improving breeding efficiency and genetic gain. The objective of this study was to assess the efficacy of predicting the performance of untested sorghum hybrids using a genomic prediction model. One hundred and two public parental inbred lines were genotyped with the genotyping-by-sequencing (GBS) platform generating 66,265 SNP markers. These markers were used to predict the performance of 204 F1 hybrids resulted from crosses between the parental lines and this was validated using a five-fold cross-validation procedure. Various training population size, cross-validation procedures and genetic effects (additive versus additive and dominance) were used to build the genomic prediction (GP) model in order to determine their effects on prediction accuracy of hybrid performance. Increasing training population size from 41 to 163 F1 hybrids increased prediction accuracies in all traits with the effect being different for different traits. Considering the additive marker effects alone in genomic prediction model, the five-fold cross validated prediction accuracies ranged from 0.03 for thousand kernel weight (TKW) to 0.58 for grain yield. When both additive and dominance effects were considered in the model, the prediction accuracies improved ranging from 0.06 for TKW to 0.67 for grain yield. Prediction accuracy show similar trend in both scenarios with the full model seemingly providing better prediction at least for some

of the traits. The results suggest that genomic prediction could become an effective tool for predicting the performance of untested sorghum hybrids.

Key words: Genomic selection; GP, Genomic prediction; Prediction accuracy; GBS, Genotyping-by-sequencing; SNPs, Single nucleotide polymorphisms; Additive marker effects; Dominance marker effects.

Introduction

The use of conventional breeding schemes such as the pedigree method, which involves phenotypic selection and trait screening over several successive generations is the most common method used in breeding programs. In hybrid sorghum breeding, several parental inbred lines are developed and intercrossed every season and their F1 hybrids evaluated across environments for estimating combining ability and selecting the top hybrids. The parental lines that produce superior hybrids undergo further testing in as many environments as possible to establish their combining ability and determine the stability of the performance of the hybrids for the desired traits. However, after such rigorous field evaluations, only the top few are advanced while the majority of the hybrids discarded. Because the procedure involves creation and evaluation of very large number of crosses in different locations and years just to select few hybrids, it is not efficient with regard to both cost and time.

The development of molecular marker techniques and quantitative trait loci (QTL) mapping over the last two decades has led to the use of marker-assisted selection (MAS) as an alternative to phenotypic selection to improve the efficiency of the system. This approach has been successfully used in different crops to incorporate QTLs controlling abiotic stresses, for example, submergence, salinity and drought tolerance traits have been successfully incorporated into new varieties in several crops

(Gregorio et al., 2013; Tuberosa et al., 2007; Araus et al., 2008). But MAS has been shown to be more effective in capturing large effect QTLs (Xu and Crouch, 2008; Castro et al., 2003) and thus only a few significant markers with large effects can be utilized. QTLs with small effects are hard to capture using MAS and hence its impact for improving efficiency for breeding complex traits such as yield has become limited (Bernardo, 2010). Moreover, many QTLs detected by MAS are usually specific to a particular genetic background. Hence, MAS has only limited applicability to quantitatively inherited traits, and its effect becomes even much less in hybrid selection. Therefore, a more efficient (less expensive and faster) method that allows selection of inbred parents with enhanced hybrid performance is needed. Such method should provide a clue about hybrid performance without expensive field testing. Since hybrid performance is the result of interaction between alleles at several loci to influence the expression of the trait, it should be possible to predict the performance of a hybrid simply by studying the genotypes of the inbred parents. Predicting hybrid performance can ultimately reduce the number of hybrids to be evaluated in the field and hence reduce costs associated with phenotyping a large number of crosses.

The next generation sequencing technologies have provided tools for scanning the entire genome of species instead of few selected genomic regions (QTLs) and capture single nucleotide polymorphisms throughout the genome. Such polymorphisms may be responsible for a change in gene functions. Thus, selection approach that takes into account all SNPs across the genome known as genomic selection may be more powerful than other indirect selection schemes used in the past. Genomic selection (GS) has a potential to replace MAS and in conjunction with phenotypic selection can lead to improved gain per cycle and thus enhancing breeding efficiency. In GS, genome-wide molecular markers with both major and minor effects

on the traits are used to build the prediction model that is used to predict the phenotypes of untested individuals (Meuwissen et al., 2001). Theoretically, very large number of genotypes can be included in selection schemes based on the genomic prediction of their phenotypic performance. Phenotypes are predicted from the genome information using appropriate prediction models which will provide genome estimated breeding values (GEBVs) for each genotype. Prediction is made based on phenotypic values of a set of individuals (training population) randomly drawn from the larger set and marker information of the entire population (Meuwissen et al., 2001). Predictions are outputs from a model of the relationship of the genome-wide markers with phenotypes of the individuals in the training set.

Genomic selection (GS) has been successfully conducted in cattle (Hayes, 2009; Haber et al., 2010) and several crops (Windhausen et al., 2012; Sallam et al., 2015; Spindel et al., 2015). The main advantage of incorporating all molecular markers in the model is that it makes it possible to capture both major and minor QTLs for important agronomic traits. When GEBV accuracy is high enough, GP can reduce breeding time because the proportion of superior genotypes in a breeding population may increase, and hence accelerate selection gain (Bernado, 2010; Heffner et al., 2010). To date, several studies have found high GEBV accuracies for grain yield and other quantitative traits in maize and wheat using experimental cross-validation (Lorenzana and Bernado, 2009; Guo et al., 2012). Again GP in at least one genomic prediction for single-cross hybrid performance in maize has been shown to outperform marker-assisted recurrent selection (Massman et al., 2013). Furthermore, moderate cross-validation prediction accuracies have also been reported for yield and other traits in diverse germplasm and breeding populations of maize, wheat and barley (Crossa et al., 2014; Heffner et al., 2011; Lorenz et al., 2012).

This approach may be extended to hybrid breeding to replace the extensive hybrid synthesis and evaluation schemes by genome based prediction. In the present study, genomic prediction was applied to predict the performance of 204 untested hybrids based on the genotype of the parental lines. The objective was to determine whether hybrid performance in sorghum can be predicted using a genomic prediction model with a reasonable accuracy that warrants its application in hybrid breeding program.

Materials and methods

Genetic materials

A total of 102 parental inbred lines including 99 pollinator lines and 3 seed parents from the Kansas State and Texas A &M Universities sorghum breeding programs were used in this study. Of these, 59 lines were Acetolactate synthase (ALS) inhibitor herbicide resistant sorghum pollinator parents (R-lines), 16 Acetyl co-enzyme-A Carboxylase (ACCase) pollinator parents and 24 regular (non-herbicide resistant) pollinator parents. The pollinator parents were crossed to three standard seed parents, ATx399, ATx3042 and AOK11 to develop 204 F1 hybrids which formed three populations based on the female parent. Population 1 consisted of crosses between 77 pollinator parents and AOK11 as a female parent, while population 2 comprised hybrids of crosses between 59 pollinator parents and ATx3042 as the female parent. Population 3 was made up of F1 hybrids between 68 pollinator parents and ATx399 as female parent. Forty-four of the pollinator lines were common across the three populations.

Genomic DNA extraction and genotyping-by-sequencing (GBS)

Sorghum seeds of parental lines were planted in the greenhouse at Kansas State University (KSU) using 96-cell flat trays filled with Metro-mix 360 (Sun Gro)

growing medium. Ten to fourteen days after planting, young leaf tissues were harvested from each line for genomic DNA extraction using the standard cetyltrimethylammonium bromide (CTAB) method (Doyle, 1987). The Quant-*i*T PicoGreen *ds*DNA Assay Kit (Invitrogen) was used to quantify the concentrations of the DNA samples. SNP genotyping was carried out using the genotyping-by-sequencing (GBS) platform at the Institute of Genomic Diversity at Cornell University. The DNA samples were digested with *ApeKI* restriction enzyme (recognition site: G|CWCG) and 96-plex GBS libraries were constructed. DNA sequencing was done using either the Illumina Genome Analyzer IIx or Hiseq2000. The Illumina sequencing reads were aligned to the sorghum reference genome v2.1 (http://phytozome.jgi.doe.gov/pz/portal.html). SNP calling was conducted using TASSEL 3.0 GBS pipeline www.maizegenetics.net/tassel/. GBS generated 282,536 SNPs with >1% minor allele frequency (MAF) and < 20% missing data. The GBS SNP markers were filtered to ≥ 5% minor allele frequency (MAF) and < 20% missing data using PLINK v1.07 (Purcell et al., 2007), resulting in 66,265 SNPs remaining for analysis. Before analysis, the missing data were imputed using BEAGLE 4.1 (Browning, 2007).

Experimental design and data collection

The F1 hybrids were evaluated at KSU Agronomy Research Farm Ashland Bottoms near Manhattan and at the North East experimental station near Ottawa, KS during 2012, 2013 and 2014 summer seasons. Planting dates were June 8, 7 and 17 for 2012, 2013 and 2014 summer seasons, respectively at KSU Agronomy Research Farm Ashland Bottoms near Manhattan, KS while it was on June 17, 2014 for Ottawa location. The Ashland Bottoms had silt loam: fine silty, mixed superactive, mesic cumulic hapludolls soils while the Ottawa location had woodson silt loam soils. On

average Manhattan location receives annual precipitation of about 907 mm (35.7inches) with the average minimum and maximum temperatures of -18°C and 32°C, respectively. The annual precipitation for Ottawa location can reach as far as 1000mm (40 inches) on average. The mean annual minimum temperature for this location is -7°C and the annual average maximum temperature is about 32°C.The experiment was laid in a randomized complete block design with three replications. The gross plot size was 2 rows, 5m long spaced at 0.75cm. On average the annual precipitation for KSU Agronomy Research Farm Ashland Bottoms near Manhattan, KS was 338, 539 and 576mm for 2012, 2013 and 2014, respectively.

Data collected included days to flowering (DF), plant height (PH), grain yield and yield components including panicle length (PL), panicle weight (PW), panicle yield (PY), number of kernels per panicle (KN), and thousand kernel weight (TKW).

- DF was determined by recording the number of days from planting to when 50% of plants in each plot reached half bloom.
- PH was recorded by measuring the distance from soil surface to the tip of the panicle at physiological maturity expressed in centimeters.
- Grain yield was measured as the weight of the kernels harvested at maturity from each plot recorded in kilograms per hectare.

After physiological maturity, three panicles from main plants were randomly sampled from each plot for measuring yield components. Mean of the three panicles was used to represent a plot and the moisture content was adjusted 12.5% for statistical analysis.

- PL was determined as the mean length of the panicles measured from the base to the tip of the panicle.

- PW was recorded as the weight of panicle from individual plant.
- PY was measured as the weight of grains threshed from a single panicle.
- KN was recorded by counting the kernels threshed from each panicle using a laboratory seed counter (Seed Counter Model 850-3, International Marketing and Design Corp, 13802 Lookout Road, San Antonio, TX, 78233, USA).
- TKW was determined by measuring the weight of 250 kernels from each panicle and multiplying by four.

Statistical analysis

Variance components and heritability

The variance components were calculated using SAS v.9.3 (SAS Institute, Cary NC) treating all effects as random effects. Broad-sense heritability (H) for each trait was estimated based on across environment and replicate data using the equation described by Hallauer et al. (2010) as:

$$H = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_{ge}/e + \sigma^2_e/er}$$

Where σ^2_g, is genetic variance, σ^2_{ge} is genotype-by-environment interaction variance, σ^2_e is residual variance, r is the number of replicates and e is the total number of environments. Broad-sense heritability (H) was estimated for means across environments.

Principal component (PCA), familial relatedness and linkage disequilibrium (LD) analyses.

Population structure can cause biased estimation of breeding values and hence can affect prediction accuracy **(**Riedelsheimer et al., 2013; Lipka et al., 2014). Before running genomic prediction analyses, principal component analysis of the parental

inbred lines was performed in R program using *prcomp* package (Becker et al., 1988). Also, the familiar relatedness among the parental inbred lines was assessed by calculating a kinship matrix using the VanRaden method (VanRaden, 2008) in TASSEL 5.2.14 (Bradbury et al., 2007) based on the "scaled Identity by state (IBS)" (Endelman and Jannink, 2012). Also, the linkage disequilibrium (LD) analysis was performed in TASSEL 5.2.14 (Bradbury et al., 2007).

Genomic prediction of hybrid performance

The performance of the untested F1 hybrids was predicted for DF, PH, grain yield and yield components (TKW, PL, PW, PY and KN). Genomic predictions were estimated using ridge regression best linear unbiased prediction model (RR-BLUP) in R program (Endelman, 2011). BLUPs of allelic effects were estimated by assuming that all marker effects are distributed with the same variance N $(0,\sigma^2_a)$ and shrinking them toward zero (Whittaker et al., 2000).

The hybrid performance was first predicted by considering only additive marker effects using the following reduced model:

$$Y = 1_n\mu + K_A a + e$$

Further prediction was made by both additive and dominance marker effects in the prediction model to see if the combined genetic effects would improve the prediction accuracy of hybrid performance. Hence the prediction was rerun using the following full model:

$$Y = 1_n\mu + K_A a + K_D d + e$$

Where 1_n = a vector of ones, and n and μ represent the number of hybrids and the across environment mean, respectively. K_A is the design matrix ($n \times m$) for the additive marker effects, in which m indicates the number of markers. The K_A was additively coded as -1, 0 and 1, where "-1" and "1" representing homozygous

genotypic classes A_2A_2 and $A_1A_1$and "0" representing heterozygous genotypic class A_1A_2 for each SNP locus. K_D is the design matrix for the dominance marker effects coded as 0, 1, 0 with score "0" representing homozygous genotypic classes A_2A_2 and A_1A_1 and "1" for the heterozygous genotypic class A_1A_2. The additive and dominance effects of the i^{th} marker were represented as *a* and *d,* respectively, in the prediction model while *e* represents the residual effect for the j^{th} hybrid.

The additive and dominance marker effects were assumed to be normally distributed, N (0, σ_a^2) and N (0, σ_d^2) with constant variance of additive effects (σ_a^2) and dominance effects (σ_d^2), respectively. To estimate the prediction accuracy [r_{GS}], the observed phenotype was correlated with the predicted phenotype and divided by the square root of heritability of the trait evaluated across environments.

Cross validation procedure

Prediction accuracy (r_{GS}) of hybrid performance for days to flowering (DF), plant height (PH), grain yield and yield components was estimated using five-fold cross validation procedure with random sampling. To determine the effect of size of the training population, the training set sample size was varied (n_{TP} = 41, 82, 122 and 163), and run with 100 iterations. In addition, the five-fold cross-validated prediction accuracy results were obtained by subdividing the 204 F1 hybrids into five random subsets such that one subset was used as a validation set while the other four sets were used as training set. Marker effects were estimated in the training set to predict the performance of F1 hybrids in the validation set. Furthermore, the effect of evaluating training and validation sets containing less related individuals on genomic prediction accuracies of hybrid performance was also investigated. Prediction accuracy was also investigated within each population and across populations (F1 hybrids developed using different seed parents). Again, prediction accuracies were determined, whereby

the hybrids assigned to training and validation sets had overlapping males but no common seed parent. Moreover, in each cross-validation procedure, the predicted phenotypes (GEBVs) for individuals in the validation population (VP) were calculated using the marker effects estimated from the training population (TP).

Results

Hybrid performance, variance components and heritability

Table 5.1 summarizes the across environment performance of hybrids for days to flowering (DF), plant height (PH), grain yield and yield components including thousand kernel weight (TKW), panicle length (PL), panicle weight (PW), panicle yield (PY) and number of kernels per panicle (KN). All traits measured on the 204 hybrids were normally distributed as revealed by the histograms and normal Q-Q plots in Appendices V-Y. The mean DF was 65 d ranging from 53 to 85 d. On average, PH was 110.7cm with the range spanning from 79.3 to 164 cm while the grain yield ranged from 4014 to 14475.5 kg ha^{-1} with an average of 7894.5 kg ha^{-1}. Mean PL, PW and PY were 25.5cm, 68.8 and 47.7g, respectively. The range of values for these traits were 19.1 to 32.7cm for PL, 28.4 to 97.3 for PW and 26.3 to 71g for PY. Furthermore, mean KN and TKW were 1640 and 29.1g, respectively. The range was 1029 to 2324 for KN and 23.3 to 38.8g for TKW (Table 5.1).

Partitioning of the total variance to component sources indicated that the genotype × environment interaction component was consistently larger than the genotypic variance for almost all of the traits except KN and DF. This shows that genes controlling most of the yield components are quantitative in nature and sensitive to environmental variation. Broad-sense heritability estimates varied among all traits with the highest estimate of 0.81 recorded for DF and the lowest (0.23) for PW and TKW.

Table 5.1 Across environment performance results of sorghum hybrids for eight agronomic traits evaluated at Manhattan and Ottawa during 2012, 2013 and 2014 summer seasons.

Trait	Mean	Range	$\sigma^2_{g,}$	$\sigma^2_{ge,}$	$\sigma^2_{e,}$	H
Panicle length (cm)	25.5 (±2.6)	19.1-32.7	1.45	2.3	2.4	0.55
Panicle weight (g)	68.8(±14.7)	28.4-97.3	17.8	67.9	21.4	0.23
Panicle yield (g)	47.7(±9.1)	26.3-71.0	25.8	30.5	11.5	0.47
Number of kernels panicle^{-1}	1640(±307.3)	1029-2324	33.6	28.2	12.9	0.52
Thousand kernel weight (g)	29.1(±2.4)	23.3-38.8	0.43	2.27	3.9	0.23
Days to flowering	65(±5.3)	53-85	10.2	3.93	6.8	0.81
Plant height (cm)	110.7(±14.6)	79.3-164	40.9	69.8	52.5	0.47
Grain yield (kg ha^{-1})	7894.5(±2331.3)	4014-14475.5	41.1	49.33	31.2	0.23

$\sigma^2_{g,}$ genetic variance; σ^2_{ge}, genotype-by-environment variance; σ^2_{e}, residual variance; *H*, broad-sense heritability; Numbers in parentheses are standard errors.

Population structure, familiar relatedness and linkage disequilibrium (LD) analyses

The principal component analysis (PCA) result is presented in Figure 5.1. PCA results revealed three major subgroups generally based on pedigree information with the first three principal components jointly accounting for 25.1% of the total molecular variation (Figure 5.1). Although about 97% of the lines were from the KSU sorghum breeding program, there was clear pattern of genetic structure in this collection. Furthermore, across the whole population, the mean kinship value was 0.01, ranging from 0 to 1.5. Almost 98% of the parental inbred lines had kinship values of less than 0.5, indicating that majority of the lines were unrelated (data not shown). The program intends to expand the parental sources in its inbred development activities and the relatively low kinship values among the collections may be the result of the deliberate effort to diversify the parental sources.

A number of factors including linkage disequilibrium (LD) between the SNP markers with quantitative trait loci (QTLs) associated with the desired trait have shown to affect prediction accuracy in genotypic prediction or selection studies. Theoretically, the genetic basis of genomic prediction or selection is that the genetic variance of every QTL for a desired trait can be captured by SNPs because of LD between the QTL and SNPs. In the present study we also assessed the LD of the SNP markers across the entire inbred parents. Across the whole population, the average LD which is the squared allele frequency (r^2) was 0.38 which was higher than previously reported threshold for genomic prediction studies. Genome-wide LD decayed gradually as the physical distance increased (Figure 5.2).

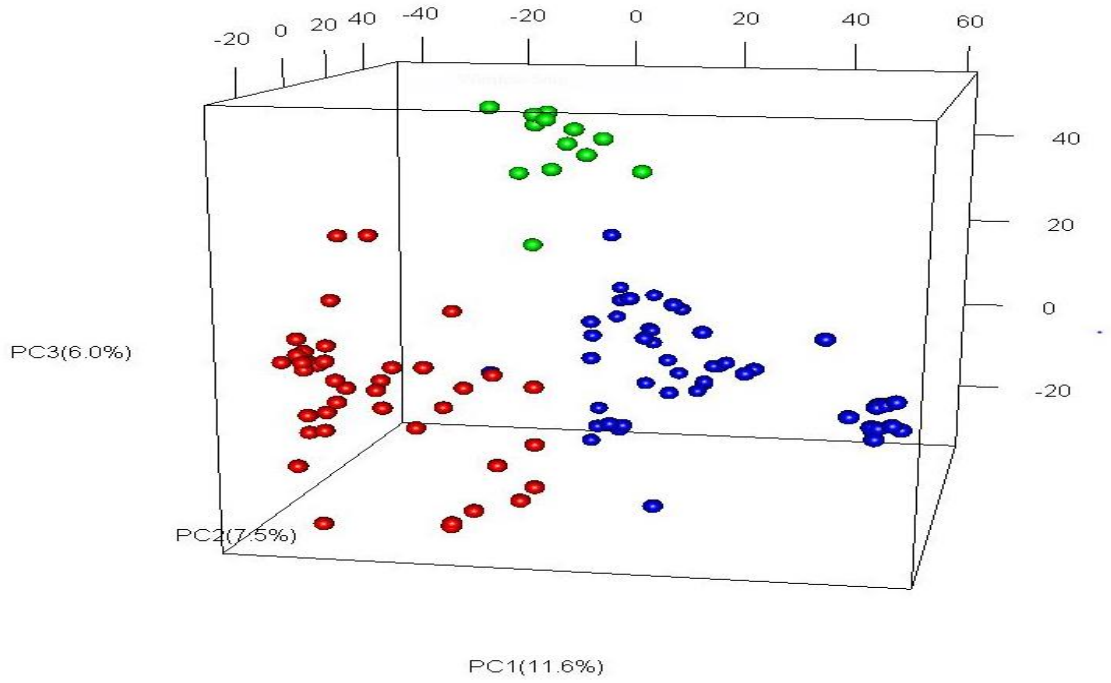


Figure 5.1 Principal component analysis (PCA) results of 102 parental inbred lines estimated using 66265 single nucleotide polymorphism markers (SNPs). Subgroup, G1 = Red; G2 = green and G3 = blue.

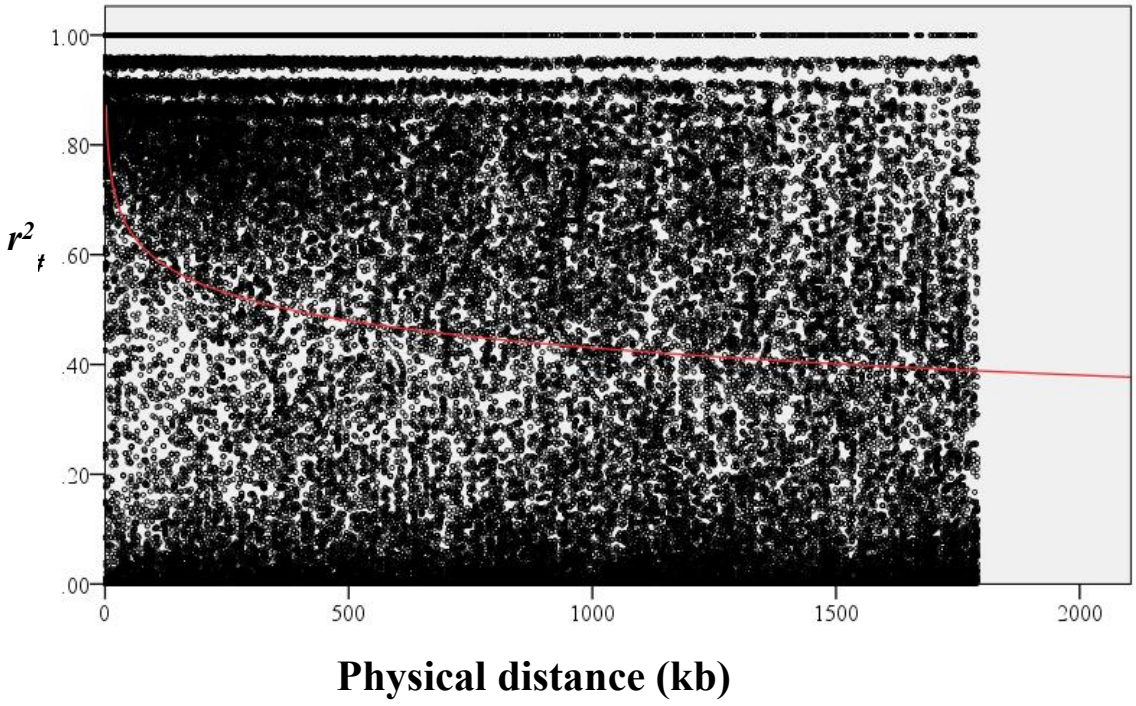


Figure 5.2 Scatter plot and estimated genome-wide linkage disequilibrium (LD) decay curve. The *y*-axis is the squared allele frequency(r^2) of genome-wide SNP pairs and the *x*-axis is the physical distance (kb) across chromosomes.

Genomic prediction accuracy of hybrid performance

The effect of training population size on prediction accuracy

Prediction of hybrid performance was studied for various training population sizes considering additive marker effects alone as well as for combined additive and dominance effects. Summary of the results are presented in Tables 5.2 and 5.3. The prediction accuracies (r_{GS}) of hybrid performance for yield and yield components based on additive effects alone increased as the number of individuals assigned to the training set increased for all traits (Table 5.2). Increasing the training population size from 41 (20%) to 163 (80%) increased the prediction accuracy for PL, PW, PY and KN by 20, 100, 175 and 89%, respectively. Other traits including DF, PH and grain

yield also had their prediction accuracies increased by 156, 65 and 28%, respectively, when the training population sizes were increased (Table 5.2). Prediction accuracy (r_{GS}) for different traits based on additive effects model was markedly different with grain yield and other traits such as PL, KN and PH having higher prediction accuracies while TKW showing the lowest prediction accuracy. Similarly, the prediction accuracy (r_{GS}) of hybrid performance under both additive and dominance model was similar to when only the additive effects were considered and for all traits the accuracy increased as the number of individuals assigned to the training set increased (Table 5.3).

Table 5.2 Prediction accuracy (r_{GS}) of hybrid performance for eight agronomic traits as affected by training population size considering additive effects of the markers alone in the model.

Trait	Prediction accuracy (r_{GS})			
	Traininng population size (TP)			
	n_{TP} =41	n_{TP} =82	n_{TP} = 122	n_{TP} =163
Panicle length (cm)	0.25	0.28	0.28	0.30
Panicle weight (g)	0.19	0.26	0.33	0.38
Panicle yield (g)	0.08	0.12	0.17	0.22
Number of kernels panicle^{-1}	0.18	0.26	0.29	0.34
Thousand kernel weight (g)	0.01	0.02	0.04	0.12
Days to flowering	0.09	0.12	0.14	0.23
Plant height (cm)	0.23	0.28	0.33	0.38
Grain yield (kg ha^{-1})	0.46	0.53	0.56	0.59

Table 5.3 Prediction of hybrid performance of eight agronomic traits considering both additive and dominance effects of the markers in the model.

	Prediction accuracy (r_{GS})			
	Training population size (TP)			
Trait	n_{TP} =41	n_{TP} =82	n_{TP} = 122	n_{TP} =163
Panicle length (cm)	0.20	0.24	0.25	0.28
Panicle weight (g)	0.15	0.18	0.21	0.28
Panicle yield (g)	0.09	0.15	0.17	0.27
Number of kernels panicle^{-1}	0.17	0.22	0.24	0.29
Thousand kernel weight (g)	0.03	0.02	0.02	0.18
Days to flowering	0.06	0.10	0.13	0.14
Plant height (cm)	0.26	0.30	0.33	0.34
Grain yield (kg ha^{-1})	0.49	0.52	0.56	0.58

Genomic prediction accuracy (r_{GS}) of hybrid performance under five-fold cross-validation

Five-fold cross validation results of prediction accuracy (r_{GS}) of hybrid performance are presented in Figures 5.3 and 5.4. Both models gave moderate to high prediction accuracies of hybrid performance for all traits with the highest accuracy observed for grain yield and the lowest for thousand kernel weight under both reduced and full prediction models.

Prediction accuracy based on additive marker effects alone was slightly different from when both additive and dominance effects were considered for all traits except for KN. For KN, the model considering additive and dominance effects together had the same level of prediction accuracy with the one based on additive effects alone. For other traits including PL, PW, TKW and grain yield the use of the combined additive and dominance model (full model) marginally improved prediction accuracy whereas accuracy for PH and DF prediction was higher when the additive model alone was used than the full model. For grain yield, which showed an overall higher prediction accuracy, the additive model alone gave r_{GS} of 0.58 versus 0.67 obtained when the full model was used (Figures 5.3 and 5.4). Similarly, for the trait with lowest prediction accuracy, TKW, the r_{GS} increased from 0.03 under the reduced model to 0.06 for the full model. Other traits PW and PL also displayed similar trend. On the other hand, the use of the full model decreased the prediction accuracy from 0.24 to 0.17 for panicle yield, from 0.18 to 0.14 for DF and from 0.36 to 0.3 for PH (Figures 5.3 and 5.4).

Prediction accuracy (r_{GS}) of hybrid performance using five-fold cross-validation where training and validation sets are related by common males or females are presented in Table 5.4. When relatedness was only due to common male parental

lines in the training and validation sets, the prediction accuracy of hybrid performance for different traits ranged from 0.06 for TKW to 0.59 for grain yield. On the other hand, when relatedness was due to common female parents, the average prediction accuracy (r_{GS}) ranged from 0.17 for panicle weight to 0.56 for grain yield (Table 5.4).

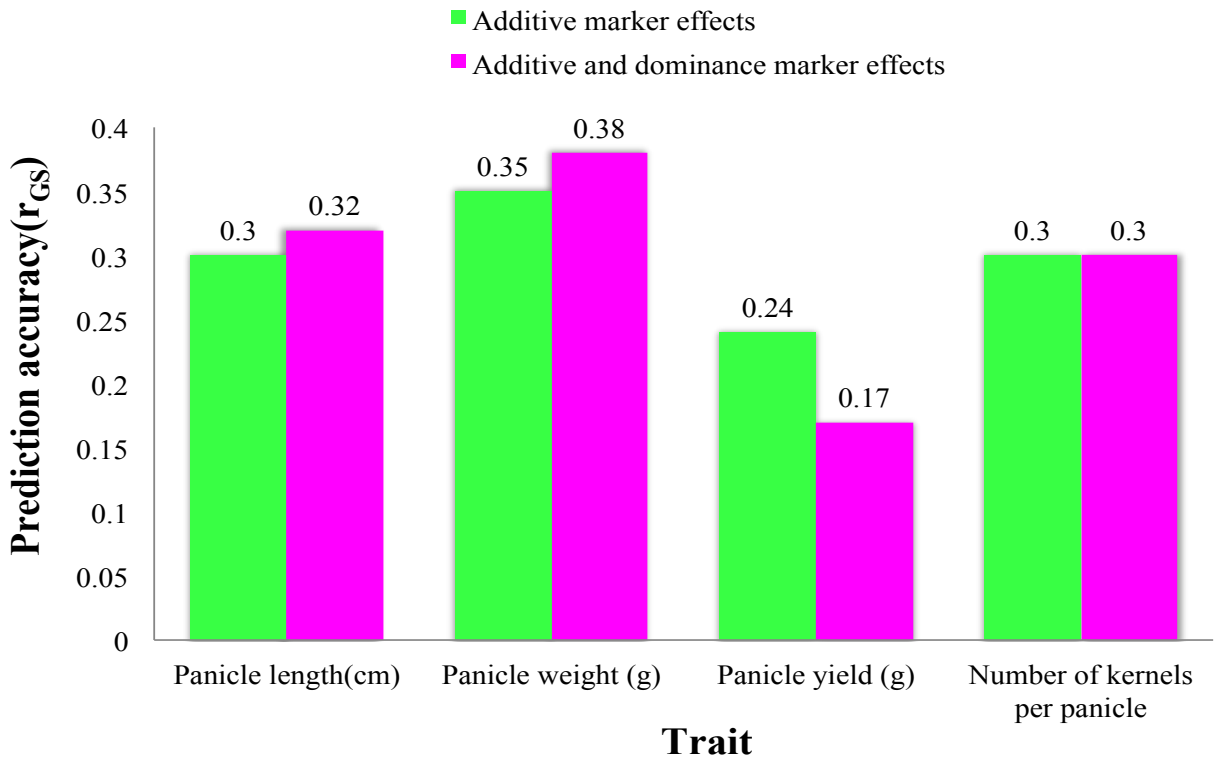


Figure 5.3 Five-fold cross-validated prediction accuracy (r_{GS}) of hybrid performance for four agronomic traits considering additive marker effects alone versus additive and dominance effects.

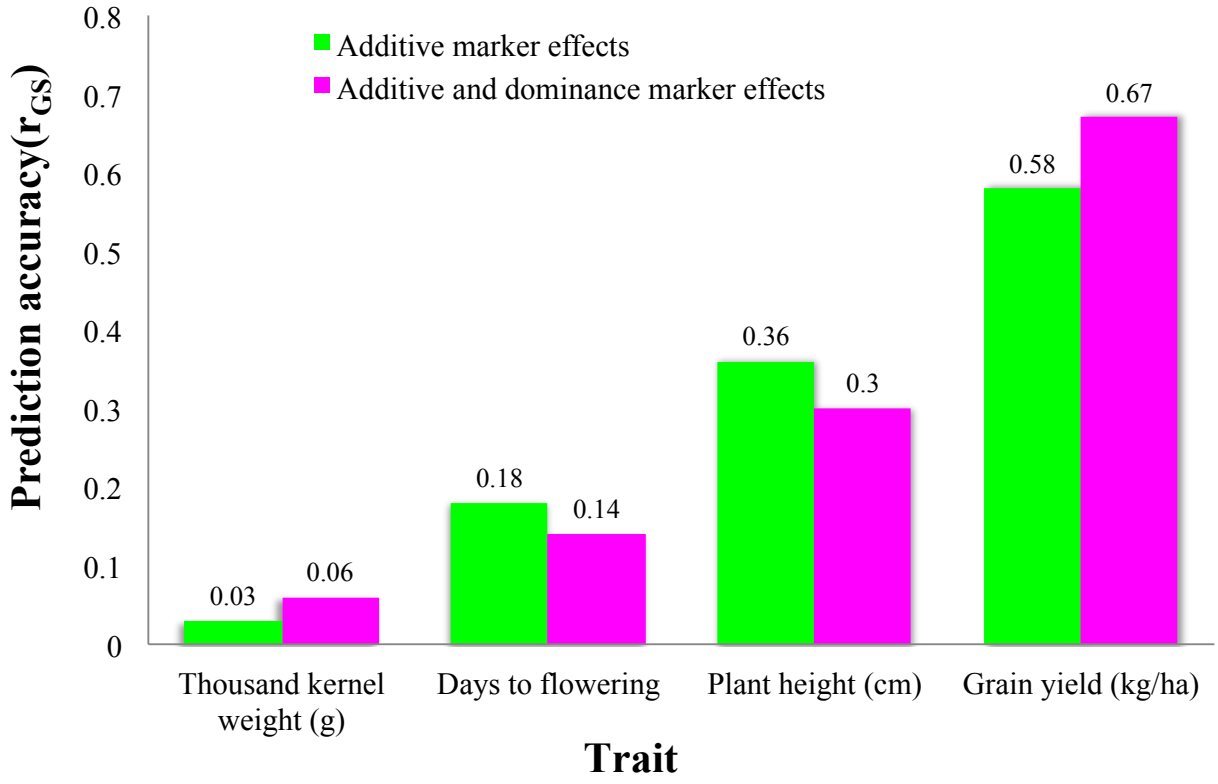


Figure 5.4 Five-fold cross-validated prediction accuracy of hybrid performance for four agronomic traits considering additive marker effects alone versus both additive and dominance effects.

Table 5.4 Prediction accuracy (r_{GS}) of hybrid performance using five-fold cross validation where training and validation sets are related by common males and females.

Trait	Related by common males (r_{GS})	Related by common females (r_{GS})
Panicle length (cm)	0.28	0.33
Panicle weight (g)	0.35	0.17
Panicle yield (g)	0.18	0.19
Number of kernels panicle^{-1}	0.26	0.23
Thousand kernel weight (g)	0.06	0.22
Days to flowering	0.16	0.27
Plant height (cm)	0.34	0.31
Grain yield (kg ha^{-1})	0.59	0.56

Discussion

The recent breakthrough in marker development and bioinformatics tools relating DNA markers with phenotypes have expanded the knowledge of gene functions and opened way for MAS to enhance breeding efficiency. While the applicability of MAS was limited to QTLs with large effect, a further development based on next-generation sequencing has provided another tool known as genomic selection. Because it accounts for all loci with both major and minor effects on the trait, genomic selection is expected to address some of the shortcomings of MAS.

In the present study, marker effects estimated on 204 sorghum hybrids were used to predict hybrid performance with respect to eight different traits namely, days to flowering, plant height, grain yield and yield components including panicle length, panicle weight, panicle yield, number of kernels per panicle and thousand kernel weight. Genomic selection utilizes phenotype data on a subset of a population (training population) to predict the performance of the entire population based on their genotype only, and phenotypic and genotypic data from the training population. So for genomic selection to be effective, it is very important that high quality genotype data is obtained on the entire population and also good quality phenotype data on sub samples of the population. Moreover, this study also looked at the effect of training population size on prediction accuracy of hybrid performance and compared two prediction models, one based on additive marker effects only and the other considering both additive and dominance effects, to predict the performance of 204 sorghum hybrids. The additive and dominance allelic effects were estimated for each marker and used to calculate predicted phenotypes (GEBVs) for untested hybrids using RR-BLUP genomic prediction model based on an infinitesimal model where all predictors are maintained in the analysis. This model was chosen because previous

studies have shown that it can give higher prediction accuracy than other genomic prediction models (Habier et al., 2007; Zhao et al., 2013). Also, it is suitable for situations where significant amount of pedigree relatedness among genotypes can be exploited efficiently (Zhao et al., 2013; Habier et al., 2007).

Previous studies have shown that in cross-validation schemes, prediction accuracy can be overestimated if both TP and VP sets contain related lines. Therefore, in this study, principal component analysis (PCA) was performed on the parental lines to determine the genetic structure of the lines before genomic prediction analysis was performed. The results show that the parental lines are structured into three subgroups to some extent based on pedigree information. Following the PCA results, an alternative cross-validation was considered in which the prediction accuracy of hybrid performance was assessed by assigning F1 hybrids in the training and validation sets with either common male or female parents. The linkage disequilibrium (LD) analysis results show that on average the LD squared Pearson correlation coefficient (r^2) between adjacent markers was 0.38, which was higher than reported by Wurschum et al. (2012). Nevertheless, the LD found in the present study is above 0.2, which has been reported as suitable threshold for genome-wide approaches for genomic selection/prediction (Hayes et al., 2009).

The genomic prediction accuracy was markedly different for different traits with grain yield having more than 50% accuracy and thousand kernel weight consistently the lowest. Increase in training population size improved prediction accuracy for all traits but the extent of the increase was different for different traits. Similar results have been reported in previous studies in other crops (Lorenz et al., 2012; Asoro et al., 2011; Heffner et al., 2011; Crossa et al., 2014; Jan et al., 2016).

Jan et al. (2016) reported increased prediction accuracies in canola with increase in training population size and no significant increase in accuracy was observed after assigning more than 70% of hybrids in the training set.

Again, grain yield consistently had the highest five-fold cross-validated prediction accuracy (r_{GS}) among the traits assessed in this study. This result corroborates with previous studies that have also reported high prediction accuracy of grain yield in wheat (Zhao et al., 2013; Crossa et al., 2010; Heffner et al., 2011; Heslot et al., 2012) and biomass yield for maize hybrids (Albrecht et al., 2011; de los Campos et al., 2009; Gonzalez-Camacho et al., 2012; Crossa et al., 2010, 2011). Furthermore, higher prediction accuracies of hybrid performance were observed for many of the traits with the full model (both additive and dominance effects) than when the reduced model (additive effects only). The result agrees with previous simulation study on maize (Technow et al., 2012) where higher prediction accuracy was reported when dominance effects of the markers were considered in the model. Some other studies have reported contrasting results such as in hybrid wheat by Zhao et al. (2013) where higher prediction accuracies of hybrid performance was observed when dominance effects were not considered in the model. They attributed this to small population size (90 hybrids) used in their study arguing that dominance model is more sensitive to the size of available data for training, suggesting that the dominance effects on prediction accuracy can be captured when the population size is large. In the present study, 204 sorghum hybrids were used which was high compared to 90 hybrids studied by Zhao et al. (2013), and perhaps that is why higher prediction accuracies were observed for some traits when dominance effects were considered in the model.

Conclusion

This study has shown that it is possible to predict the performance of untested sorghum hybrids for important agronomic traits such as grain yield using a genomic prediction model. Based on the present results, we believe that the use of genomic prediction model in the hybrid breeding program in sorghum is likely to become a viable strategy in the near future for predicting the performance of sorghum hybrids prior to phenotyping, hence significantly reducing the number of hybrids to be evaluated and also costs associated with phenotyping a large number of the hybrids in the field. The advantage is that over the years, genotyping and sequencing costs have been decreasing such that it is now possible for even small, public breeding programs to obtain high density marker information at an affordable cost.

References

Albrecht, T., V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, and M. Ouzunova et al. 2011.Genome-based prediction of testcross values in maize. Theor. Appl. Genet.123:339-350.

Araus, J. L., G. A. Slafer, C. Royo, and M. D. Serret. 2008. Breeding for yield potential and stress adaptation in cereals. Crit. Rev. Plant Sci. 27:377-412.

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Gen. 4:132-144.

Becker, R.A., J. M. Chambers, and A. R. Wilks. 1988. The New S Language. Wadsworth & Brooks/Cole, Pacific Grove, PA.

Bernardo, R. 2010. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, and Y. Ramdoss et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Browning, S.R., and B.L .Browning. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084-1097.

Castro, A.J., F. Capettini, A.E. Corey, T. Filichkina, P.M. Hayes, A. Kleinhofs, D. Kudrna, K. Richardson, S. Sandoval-Islas, C. Rossi, and H. Vivar. 2003. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor. Appl. Genet. 107:922-930.

Crossa, J., G. De Los Campos, P. Pérez, D. Gianola, J. Burgueño, and J. L. Araus, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genet.186:713-724.

Crossa, J., P. Pérez, G. de los Campos, G. Mahuku, and S. Dreisigacker et al. 2011. Genomic selection and prediction in plant breeding. J. Crop Improv. 25: 239-261.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang,S. Dreisigacker, R. Babu, Y. Li, and D. Bonnett. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112(1): 48-60.

De los Campos G., H.Naya, D. Gianola, J. Crossa, and A.Legarra et al. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigrees. Genet. 182:375-385.

Doyle, J.J.1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull 19:11-15.

Endelman, J.B.2011.Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4(3):250-255.

Endelman, J.B., J. L. Jannink. 2012. Shrinkage estimation of the realized matrix. G3 2:1405-1413.

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. Cairns et al. 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. Theor. Appl. Genet. 125: 759-771.

Gregorio, G.B., M.R.Islam, G.V.Vergara, and S.Thirumeni.2013.Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. SABRAO J. Breed. Genet. 45(1): 31-41.

Guo, Z., D.M. Tucker, J.Lu, V. Kishore, and G. Gay. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124 (2):261-275.

Habier, D., R.L. Fernando, and J.C.M, Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genet.177(4):2389-2397.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010.The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:1-12.

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho. 2010. Quantitative Genetics in Maize Breeding. Iowa State University Press, Ames, IA.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009.Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41:1-9.

Heffner, E.L., A.J. Lorenz, J. Jannink, and M.E. Sorrells. 2010. Plant breeding with

genomic selection: Gain per unit time and cost. Crop Sci. 50:1681-1690.

Heffner, E.L., J.L.Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome 4(1): 65-75.

Heslot, N., H.P. Yang, M.E. Sorrels, and J.L. Jannink. 2012. Genomic selection in plant breeding: a comparison of models. Crop Sci. 52:146-160.

Jan, H.U., A. Abbadi, S. Lücke, R.A. Nichols, and R.J.Snowdon. 2016. Genomic prediction of testcross performance in canola (*Brassica napus*). PLoS One 11(1): e0147769.

Lipka, A.E., F. Lu, J.H. Cherney, E.S. Buckler, M.D. Casler, and D.E. Costich. 2014. Accelerating the Switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. PLoS One 9(11): e112227.

Lorenz, A.J., K.P. Smith, and J.L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52:1609-1621.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151-161.

Massman, J. M., A. Gordillo, R. E. Lorenzana, and R. Bernardo. 2013. Genome-wide predictions from maize single-cross data. Theor. Appl. Genet. 126:13-22.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic values using genome-wide dense marker maps. Genet.157:1819-1829.

Purcell, S., B. Neale, K.Todd-Brown, L.Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. De Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set

for whole-genome association and population-based linkage analyses. American J. Human Genet. 81(3): 559-575.

Riedelsheimer, C., J.B. Endelman, M.Stange, M.E. Sorrells, J.L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected bi-parental maize populations. Genet. 194(2): 493-503.

Sallam, A.H., J.B. Jannink, and K.P.Smith.2015.Assessing Genomic Selection Prediction in a Dynamic Barley Breeding Population. Plant Genome 8(1):1-15.

SAS Institute. 2011. The SAS system for Windows. v.9.3. SAS Inst., Cary, NC.

Spindel, J., H. Begum, D. Akdemir , P. Virk , B. Collard, E. Redona, G.Atlin, J.L. Jannink, and S.R. McCouch. 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 Genet.11: e1004982.

Technow, F., C. Riedelsheimer, T. A. Schrag, and A. E. Melchinger.2012.Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. 125:1181-1194.

Tuberosa, R., S. Salvi, S. Giuliani, M. C. Sanguineti, and M. Bellotti et al. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47: 120-141.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423.

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker- assisted selection using ridge regression. Genet. Res. 75:249-252.

Windhausen, V. S., G. N. Atlin, J. Crossa, J. M. Hickey, and P. Grudloyma et al. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda) 2:1427-1436.

Würschum, T. 2012. Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet.125(2): 201-210.

Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Sci. 48:391-407.

Zhao, Y., J. Zeng, R. Fernando, and J.C. Reif. 2013. Genomic prediction of hybrid wheat performance. Crop Sci. 53(3): 802-810.

Appendix A - Determination of number of principal components to use for clustering the inbred lines

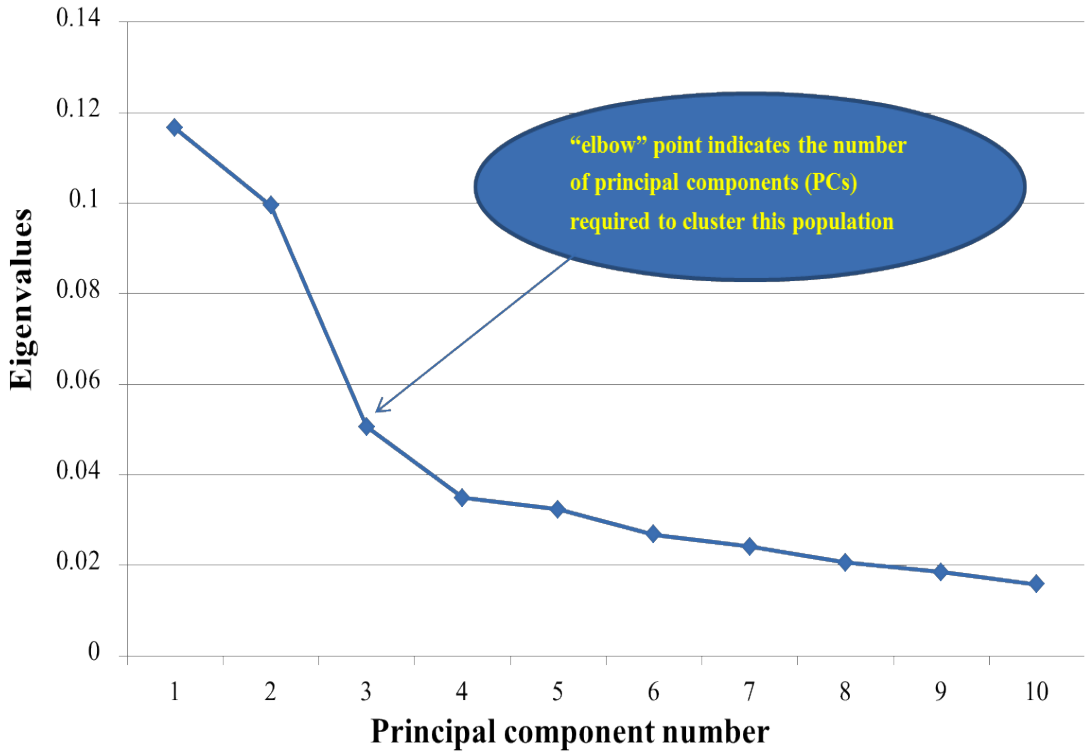


Figure A.1 Scree plot of principal components (x-axis) and their contribution to variance (y-axis).

Appendix B - Scatterplots and estimated linkage disequilibrium (r^2) decay curves.

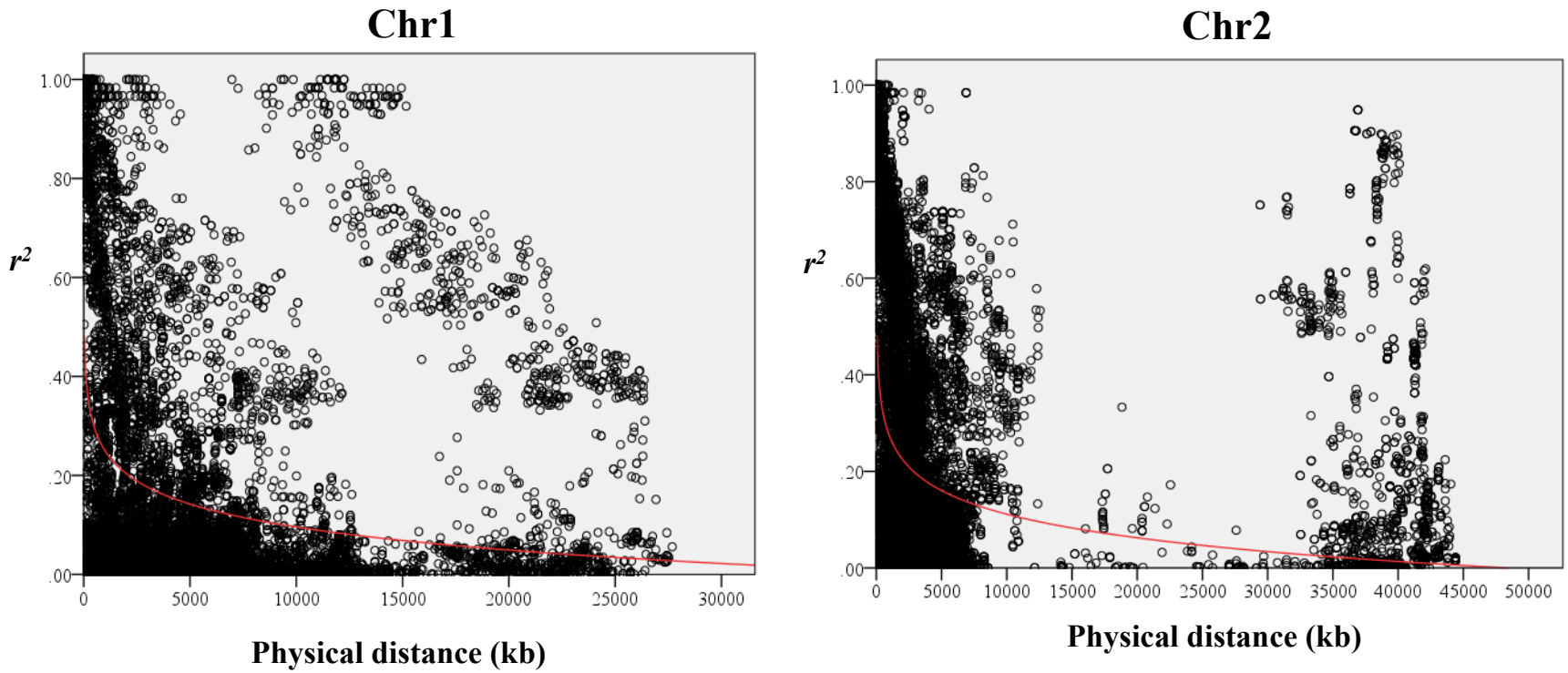


Figure B.1 Scatterplots of linkage disequilibrium (r^2) against physical distance (kb) and estimated LD decay curves for chromosomes 1 and 2.

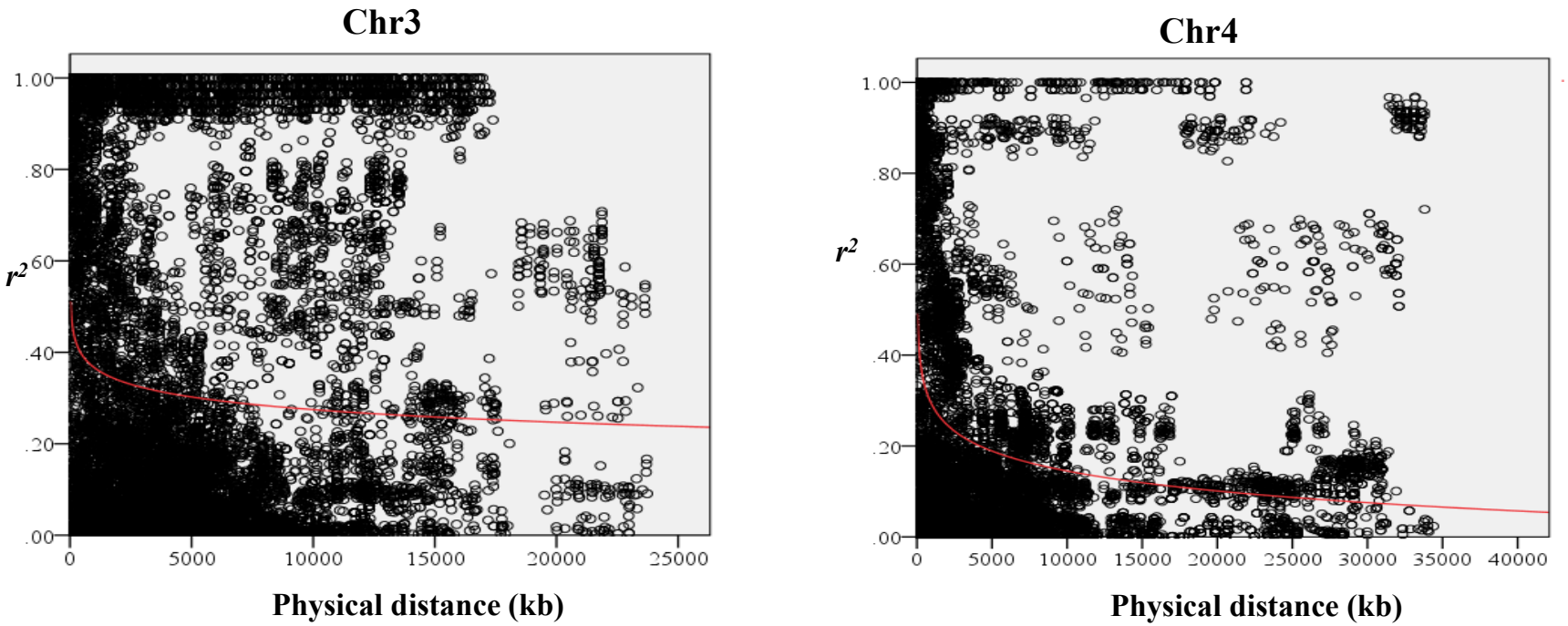


Figure B.2 Scatterplots of linkage disequilibrium (r^2) against physical distance (kb) and estimated LD decay curves for chromosomes 3 and 4.

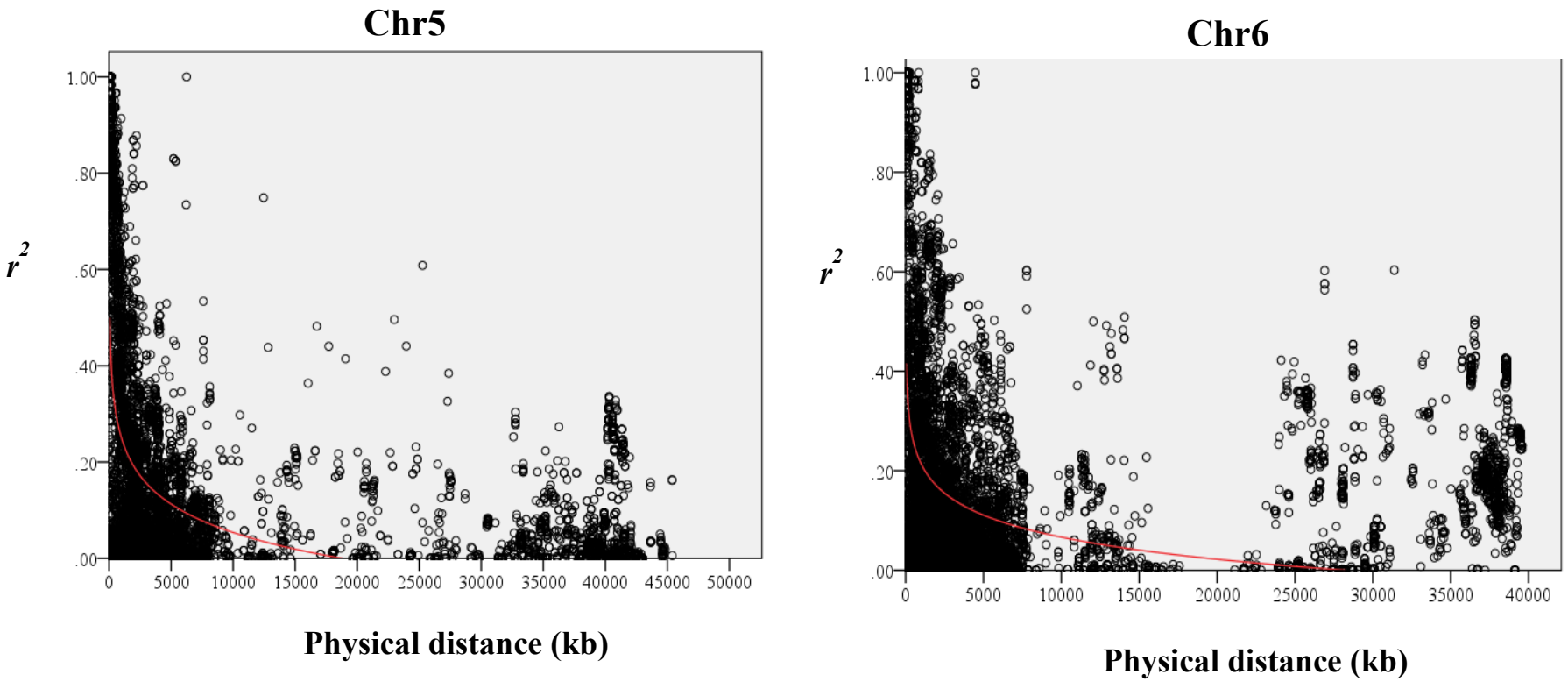


Figure B.3 Scatterplots of linkage disequilibrium (r^2) against physical distance (kb) and estimated LD decay curves for chromosomes 5 and 6.

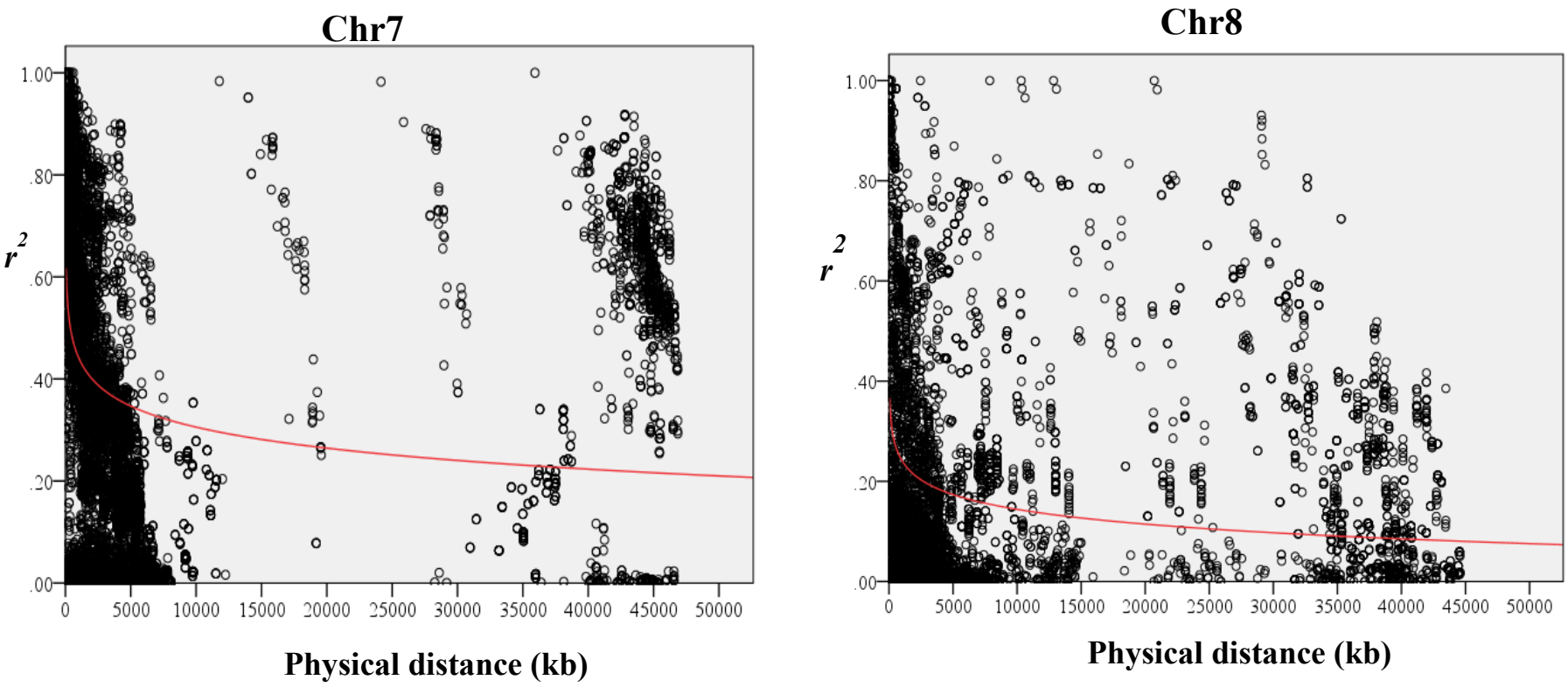


Figure B.4 Scatterplots of linkage disequilibrium (r^2) against physical distance (kb) and estimated LD decay curves for chromosomes 7 and 8.

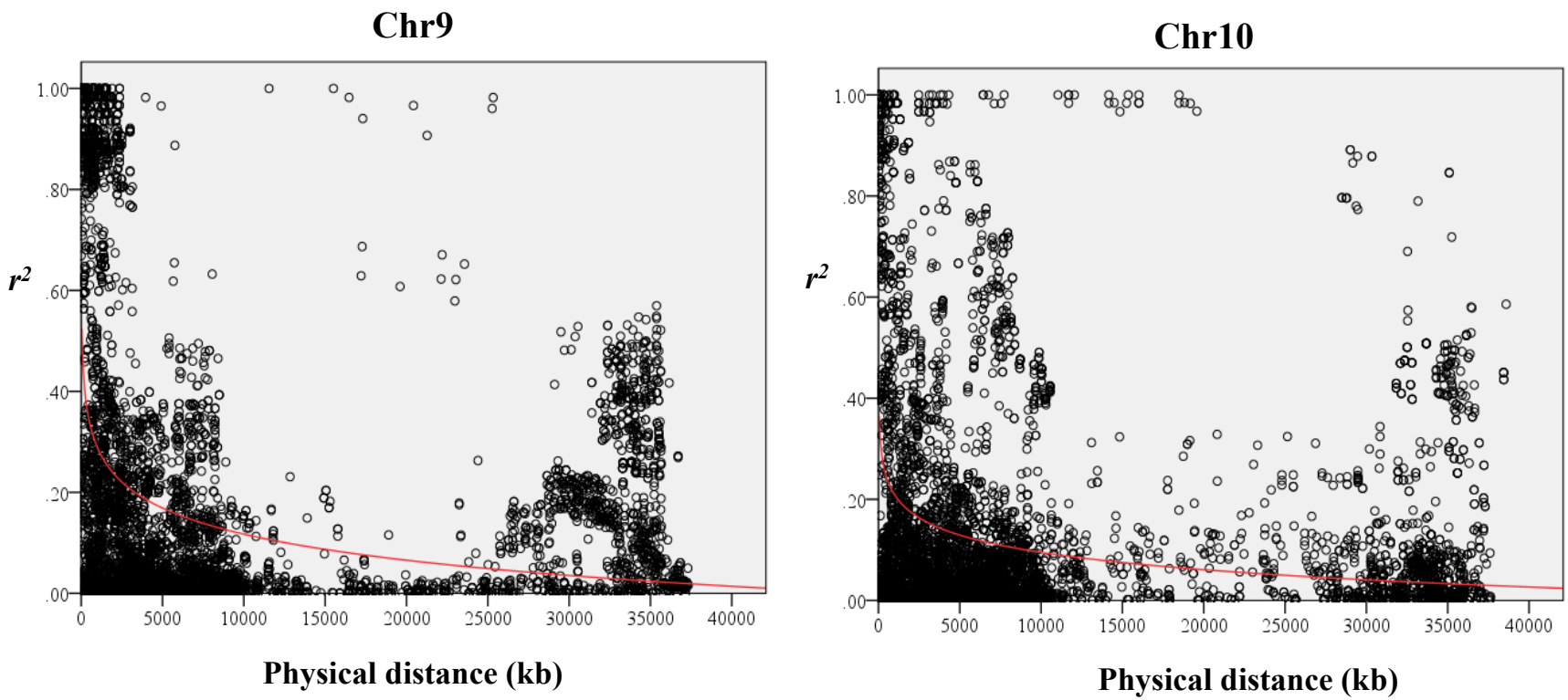


Figure B.5 Scatterplots of linkage disequilibrium (r^2) against physical distance (kb) and estimated LD decay curves for chromosomes 9 and 10.

Appendix C - Pearson correlation coefficients among F1 hybrid, mid-parent, mid-parent heterosis (MPH) and better-parent heterosis (BPH) for panicle length (above diagonal) and panicle weight (below diagonal)

Trait	F1 hybrid	Mid-parent	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.83**	0.43**	0.35*
Mid-parent	0.56*	-	-0.14	-0.19
Mid-parent heterosis	0.62**	-0.29*	-	0.93**
Better-parent heterosis	0.52**	-0.32*	0.90**	-

Appendix D - Pearson correlation coefficients among F1 hybrid, mid-parent, mid-parent heterosis (MPH) and better-parent heterosis (BPH) for panicle yield (above diagonal) and number of kernels per panicle (below diagonal)

Trait	F1 hybrid	Mid-parent	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.64**	0.56**	0.94**
Mid-parent	0.46**	-	-0.38**	-0.41**
Mid-parent heterosis	0.70**	-0.31**	-	0.94**
Better-parent heterosis	0.62**	0.95**	-0.35*	-

Appendix E - Pearson correlation coefficients among F1 hybrid, mid-parent, mid-parent heterosis (MPH) and better-parent heterosis (BPH) for thousand kernel weight (above diagonal) and days to flowering (below diagonal)

Trait	F1 hybrid	Mid-parent	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	0.35*	0.74**	0.52**
Mid-parent	0.32*	-	-0.38**	-0.44**
Mid-parent heterosis	0.68**	-0.48**	-	0.83**
Better-parent heterosis	0.74**	-0.19	0.83**	-

Appendix F - Pearson correlation coefficients among F1 hybrid, mid-parent, mid-parent heterosis (MPH) and better-parent heterosis (BPH) for plant height (above diagonal) and grain yield (below diagonal)

Trait	F1 hybrid	Mid-parent	Mid-parent heterosis	Better-parent heterosis
F1 hybrid	-	-0.26	0.69**	0.88**
Mid-parent	0.07	-	-0.87**	-0.43**
Mid-parent heterosis	0.69**	-0.66**	-	0.74**
Better-parent heterosis	0.74**	-0.47**	0.89**	-

Appendix G - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle length and panicle weight of 102 hybrids.

Source	Panicle length (cm)	Mid-parent (cm)	Better-parent (cm)	Panicle weight (g)	Mid-parent (g)	Better-parent (g)	Genetic distance
PR14/15-1509 x PR14/15-1637	30.43	29.38	30.08	109.96	100.37	101.35	0.38
PR14/15-1509 x PR14/15-1650	32.66	29.59	30.50	113.29	94.21	99.39	0.53
PR14/15-1501 x PR14/15-1633	26.78	29.17	30.08	96.51	100.50	102.43	0.33
PR14/15-1505 x PR14/15-1635	29.65	28.96	29.92	112.58	102.04	106.74	0.12
PR14/15-1549 x PR14/15-1625	33.83	29.17	30.67	125.71	96.71	102.21	0.30
PR14/15-1505 x PR14/15-1638	30.46	27.75	28.00	109.86	96.37	97.33	0.33
PR14/15-1505 x PR14/15-1637	29.08	29.04	30.08	122.89	99.34	101.35	0.38
PR14/15-1569 x PR14/15-1627	31.67	29.96	33.17	117.16	103.62	121.35	0.46
PR14/15-1549 x PR14/15-1627	34.08	30.42	33.17	117.13	106.28	121.35	0.48
PR14/15-1549 x PR14/15-1623	32.17	29.80	31.92	129.32	95.17	99.14	0.49
PR14/15-1569 x PR14/15-1630	26.17	27.63	28.50	98.24	89.46	93.04	0.50
PR14/15-1505 x PR14/15-1650	31.07	29.25	30.50	110.69	93.18	97.33	0.51
PR14/15-1569 x PR14/15-1641	29.67	28.04	29.33	114.12	91.86	97.83	0.55
PR14/15-1545 x PR14/15-1622	29.50	30.74	31.67	116.13	94.28	107.19	0.47
PR14/15-1545 x PR14/15-1631	33.13	29.57	29.80	137.91	82.19	81.36	0.50
PR14/15-1501 x PR14/15-1628	31.48	27.88	28.25	126.71	103.49	104.55	0.39
PR14/15-1537 x PR14/15-1625	34.08	32.54	34.40	140.73	102.17	102.21	0.45
PR14/15-1537 x PR14/15-1622	30.83	33.04	34.40	131.28	104.66	107.19	0.47
PR14/15-1501 x PR14/15-1622	31.13	29.96	31.67	115.86	104.81	107.19	0.50
PR14/15-1501 x PR14/15-1634	31.95	28.09	28.25	119.43	87.53	102.43	0.51
PR14/15-1501 x PR14/15-1636	30.72	30.13	32.00	123.59	104.07	105.70	0.52
PR14/15-1501 x PR14/15-1650	32.88	29.38	30.50	132.84	95.73	102.43	0.56
PR14/15-1537 x PR14/15-1647	28.75	34.24	34.40	103.96	109.60	117.08	0.57

PR14/15-1501 x PR14/15-1649	30.00	28.04	28.25	108.45	97.77	102.43	0.70
PR14/15-1565 x PR14/15-1622	32.92	30.96	31.67	129.33	104.30	107.19	0.24

Appendix H - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle length and panicle weight of 102 hybrids.

Source	Panicle length (cm)	Mid-parent (cm)	Better-parent (cm)	Panicle weight (g)	Mid-parent (g)	Better-parent (g)	Genetic distance
PR14/15-1589 x PR14/15-1635	35.31	29.46	29.92	110.03	117.06	127.38	0.29
PR14/15-1593 x PR14/15-1635	35.45	28.63	29.92	135.78	113.98	121.22	0.30
PR14/15-1553 x PR14/15-1622	32.33	31.30	31.67	140.37	98.70	107.19	0.33
PR14/15-1521 x PR14/15-1635	33.84	29.00	29.92	141.20	101.64	106.74	0.37
PR14/15-1517 x PR14/15-1635	31.04	29.50	29.92	119.62	110.37	114.00	0.38
PR14/15-1565 x PR14/15-1628	31.58	28.88	30.25	138.01	102.98	104.55	0.39
PR14/15-1553 x PR14/15-1624	35.83	30.38	30.50	124.05	96.52	101.40	0.46
PR14/15-1565 x PR14/15-1625	34.50	30.46	30.67	116.38	101.81	101.21	0.46
PR14/15-1553 x PR14/15-1623	33.25	31.42	31.92	116.38	94.68	99.14	0.47
PR14/15-1553 x PR14/15-1631	33.08	30.13	30.92	134.68	86.62	90.21	0.47
PR14/15-1565 x PR14/15-1621	33.58	29.67	30.25	122.23	91.70	101.40	0.48
PR14/15-1565 x PR14/15-1630	33.42	29.38	30.25	114.16	97.22	101.40	0.50
PR14/15-1553 x PR14/15-1630	33.58	29.71	30.92	141.89	91.63	93.04	0.50
PR14/15-1565 x PR14/15-1645	30.83	29.09	30.25	121.08	98.79	101.40	0.56
PR14/15-1521 x PR14/15-1650	33.36	29.29	30.50	127.90	92.78	96.53	0.27
PR14/15-1521 x PR14/15-1633	29.80	29.08	30.08	155.34	97.55	98.57	0.30
PR14/15-1593 x PR14/15-1636	34.00	29.67	32.00	127.47	113.46	121.22	0.30
PR14/15-1589 x PR14/15-1638	31.76	28.25	29.00	107.47	111.39	127.38	0.37

PR14/15-1581 x PR14/15-1625	33.00	30.13	30.67	133.08	107.97	113.72	0.37
PR14/15-1593 x PR14/15-1638	31.11	27.42	27.50	152.63	108.31	121.22	0.38
PR14/15-1593 x PR14/15-1628	31.21	27.42	27.50	141.46	112.89	121.55	0.38
PR14/15-1593 x PR14/15-1633	31.05	28.71	30.08	136.25	109.90	121.22	0.39
PR14/15-1521 x PR14/15-1648	30.95	28.38	28.67	138.58	92.10	96.53	0.46
PR14/15-1581 x PR14/15-1621	34.00	29.33	29.58	132.74	97.86	113.72	0.46
PR14/15-1521 x PR14/15-1636	32.61	30.04	32.00	120.78	101.12	105.70	0.48

Appendix I - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle length and panicle weight of 102 hybrids.

Source	Panicle length (cm)	Mid-parent (cm)	Better-parent (cm)	Panicle weight (g)	Mid-parent (g)	Better-parent (g)	Genetic distance
PR14/15-1521 x PR14/15-1634	33.72	28.00	28.08	145.10	84.58	96.53	0.50
PR14/15-1593 x PR14/15-1648	34.47	28.00	28.67	127.01	104.45	121.22	0.51
PR14/15-1593 x PR14/15-1650	34.81	28.92	30.50	112.72	105.12	121.22	0.52
PR14/15-1593 x PR14/15-1634	32.60	27.63	27.92	140.26	96.92	121.22	0.54
PR14/15-1561 x PR14/15-1622	32.79	31.47	31.67	116.12	104.93	107.19	0.31
PR14/15-1589 x PR14/15-1637	34.67	29.54	30.08	124.20	114.37	127.38	0.37
PR14/15-1561 x PR14/15-1625	32.42	30.97	31.27	128.13	102.44	102.67	0.46
PR14/15-1573 x PR14/15-1624	30.50	29.34	30.50	137.81	86.56	91.63	0.47
PR14/15-1561 x PR14/15-1644	30.42	29.64	31.27	110.54	91.60	102.67	0.54
PR14/15-1561 x PR14/15-1646	30.17	30.18	31.27	111.52	93.20	102.67	0.57
PR14/15-1557 x PR14/15-1646	31.08	29.78	30.48	129.89	89.87	96.01	0.57
PR14/15-1533 x PR14/15-1627	34.17	31.26	33.17	175.97	108.12	121.35	0.35
PR14/15-1589 x PR14/15-1628	35.50	28.25	29.00	145.22	115.97	127.38	0.37
PR14/15-1589 x PR14/15-1636	33.95	30.50	32.00	142.45	116.54	127.38	0.38
PR14/15-1589 x PR14/15-1623	34.63	30.46	31.92	116.50	113.26	127.38	0.49
PR14/15-1589 x PR14/15-1648	34.29	28.84	29.00	126.21	107.53	127.38	0.52

PR14/15-1589 x PR14/15-1650	32.98	29.75	30.50	122.84	108.20	127.38	0.53
PR14/15-1557 x PR14/15-1622	31.75	31.08	31.67	117.16	101.60	107.19	0.27
PR14/15-1557 x PR14/15-1625	35.50	30.58	30.67	133.80	99.11	102.21	0.34
PR14/15-1557 x PR14/15-1621	32.08	29.78	30.48	99.97	89.01	96.01	0.48
PR14/15-1585 x PR14/15-1625	35.50	29.96	30.67	121.54	97.44	102.21	0.47
PR14/15-1529 x PR14/15-1625	34.50	29.89	30.67	159.32	98.43	102.21	0.39
PR14/15-1529 x PR14/15-1627	33.33	31.14	33.17	129.16	108.00	121.35	0.47
PR14/15-1609 x PR14/15-1648	33.25	.	.	115.52	.	.	0.34
Dekalb54-00	28.83	.	.	109.03	.	.	.

Appendix J - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle length and panicle weight of 102 hybrids.

Source	Panicle length (cm)	Mid-parent (cm)	Better-parent (cm)	Panicle weight (g)	Mid-parent (g)	Better-parent (g)	Genetic distance
Pioneer84G62	31.00	.	.	109.76	.	.	.
PR14/15-1609 x PR14/15-1649	34.08	.	.	128.94	.	.	0.35
Seneca	29.17	.	.	102.40	.	.	.
PR14/15-1609 x PR14/15-1635	33.83	.	.	145.15	.	.	0.38
PR14/15-1577 x PR14/15-1625	33.67	.	.	121.08	.	.	0.38
PR14/15-1609 x PR14/15-1650	36.83	.	.	123.42	.	.	0.39
PR14/15-1597 x PR14/15-1635	35.02	.	.	129.50	.	.	0.40
PR14/15-1597 x PR14/15-1628	32.26	.	.	144.55	.	.	0.41
PR14/15-1597 x PR14/15-1634	34.14	.	.	132.04	.	.	0.44
PR14/15-1597 x PR14/15-1638	31.87	.	.	128.58	.	.	0.45
PR14/15-1601 x PR14/15-1638	30.39	.	.	127.88	.	.	0.46
PR14/15-1597 x PR14/15-1650	32.15	.	.	117.75	.	.	0.47
PR14/15-1597 x PR14/15-1649	33.20	.	.	127.87	.	.	0.47
PR14/15-1597 x PR14/15-1633	30.98	.	.	129.08	.	.	0.47
PR14/15-1597 x PR14/15-1648	33.81	.	.	141.45	.	.	0.47
PR14/15-1601 x PR14/15-1637	30.83	.	.	113.78	.	.	0.47

PR14/15-1601 x PR14/15-1649	28.53	.	.	109.48	.	.	0.49
PR14/15-1605 x PR14/15-1635	31.79	.	.	165.01	.	.	0.51
PR14/15-1601 x PR14/15-1650	30.81	.	.	121.47	.	.	0.56
PR14/15-1601 x PR14/15-1633	31.07	.	.	116.18	.	.	0.59
PR14/15-1565 x PR14/15-1626	34.42	30.25	.	130.10	101.40		0.34
PR14/15-1545 x PR14/15-1626	31.75	29.80	.	117.95	81.36	.	0.34
PR14/15-1609 x PR14/15-1638	33.17	.	.	133.83	.	.	0.37
PR14/15-1609 x PR14/15-1636	34.83	.		128.98	.		0.37
PR14/15-1605 x PR14/15-1648	32.50	.	.	131.29	.		0.38

Appendix K - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle yield and number of kernels per panicle of 102 hybrids.

Source	Panicle yield (g)	Mid-parent (g)	Better-parent (g)	No. kernels panicle^{-1}	Mid-parent	Better-parent	Genetic distance
PR14/15-1509 x PR14/15-1637	64.40	57.97	58.06	2204.83	1816.00	1896.00	0.38
PR14/15-1509 x PR14/15-1650	65.88	56.00	58.06	1772.75	1735.50	1736.00	0.53
PR14/15-1501 x PR14/15-1633	64.73	61.48	64.42	2249.67	2050.00	2209.00	0.33
PR14/15-1505 x PR14/15-1635	64.03	58.75	61.51	2020.58	1913.00	2000.00	0.12
PR14/15-1549 x PR14/15-1625	69.70	57.21	63.06	2311.17	1958.50	2090.00	0.30
PR14/15-1505 x PR14/15-1638	67.65	56.65	57.32	2004.08	1878.50	1931.00	0.33
PR14/15-1505 x PR14/15-1637	77.47	56.93	57.88	2415.33	1861.00	1896.00	0.38
PR14/15-1569 x PR14/15-1627	71.38	59.58	68.67	2290.58	1962.00	2005.00	0.46
PR14/15-1549 x PR14/15-1627	81.42	60.01	68.67	2334.58	1916.00	2005.00	0.48
PR14/15-1549 x PR14/15-1623	79.47	53.06	54.77	2608.00	1872.50	1918.00	0.49
PR14/15-1569 x PR14/15-1630	55.45	51.98	53.46	1857.83	1828.50	1919.00	0.50
PR14/15-1505 x PR14/15-1650	69.84	54.96	55.98	2174.75	1780.50	1826.00	0.51
PR14/15-1569 x PR14/15-1641	68.67	52.39	54.28	2061.83	1738.50	1919.00	0.55
PR14/15-1545 x PR14/15-1622	66.36	52.84	59.25	2132.83	1779.00	1955.00	0.47
PR14/15-1545 x PR14/15-1631	79.50	43.75	46.43	2624.00	1627.00	1651.00	0.50
PR14/15-1501 x PR14/15-1628	73.40	61.08	64.42	2469.50	2040.00	2209.00	0.39
PR14/15-1537 x PR14/15-1625	88.35	61.11	63.06	2784.33	2007.00	2090.00	0.45
PR14/15-1537 x PR14/15-1622	81.80	59.20	59.25	2599.75	1939.50	1955.00	0.47
PR14/15-1501 x PR14/15-1622	73.30	61.84	64.42	2335.00	2082.00	2209.00	0.50
PR14/15-1501 x PR14/15-1634	71.73	52.07	64.42	2411.58	1820.00	2209.00	0.51
PR14/15-1501 x PR14/15-1636	78.33	62.67	64.42	2707.83	2143.50	2209.00	0.52
PR14/15-1501 x PR14/15-1650	87.02	59.18	64.42	2569.33	1972.00	2209.00	0.56
PR14/15-1537 x PR14/15-1647	61.64	65.80	72.45	1892.50	2149.00	2374.00	0.57

PR14/15-1501 x PR14/15-1649	64.84	58.72	64.42	2152.42	1939.50	2209.00	0.70
PR14/15-1565 x PR14/15-1622	80.67	58.54	59.25	2707.42	1920.50	1955.00	0.24

Appendix L - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle yield and number of kernels per panicle of 102 hybrids.

Source	Panicle yield (g)	Mid-parent (g)	Better-parent (g)	No. kernels panicle^{-1}	Mid-parent	Better-parent	Genetic distance
PR14/15-1589 x PR14/15-1635	65.88	67.72	73.93	2064.75	2204.50	2409.00	0.29
PR14/15-1593 x PR14/15-1635	79.18	63.36	65.20	2429.92	2136.50	2273.00	0.30
PR14/15-1553 x PR14/15-1622	87.48	53.61	59.25	2820.17	1782.00	1955.00	0.33
PR14/15-1521 x PR14/15-1635	85.79	57.90	61.51	2828.42	1870.50	2000.00	0.37
PR14/15-1517 x PR14/15-1635	70.37	64.55	67.58	2273.67	2141.00	2282.00	0.38
PR14/15-1565 x PR14/15-1628	82.04	57.78	57.83	2493.42	1878.50	1886.00	0.39
PR14/15-1553 x PR14/15-1624	79.47	47.27	57.83	2375.92	1816.50	1886.00	0.46
PR14/15-1565 x PR14/15-1625	69.72	60.45	63.06	2212.17	1988.00	2090.00	0.46
PR14/15-1553 x PR14/15-1623	67.08	51.37	54.77	2100.67	1763.50	1918.00	0.47
PR14/15-1553 x PR14/15-1631	72.68	44.52	47.97	2470.08	1630.00	1651.00	0.47
PR14/15-1565 x PR14/15-1621	75.61	53.61	57.83	2332.50	1728.00	1886.00	0.48
PR14/15-1565 x PR14/15-1630	69.53	55.65	57.83	2356.25	1812.00	1886.00	0.50
PR14/15-1553 x PR14/15-1630	87.87	50.72	53.46	2781.83	1673.50	1738.00	0.50
PR14/15-1565 x PR14/15-1645	71.07	55.47	57.83	2097.33	1798.50	1886.00	0.56
PR14/15-1521 x PR14/15-1650	78.13	54.11	54.28	2289.58	1738.00	1741.00	0.27
PR14/15-1521 x PR14/15-1633	94.62	56.41	58.54	3038.08	1816.00	1891.00	0.30
PR14/15-1593 x PR14/15-1636	76.03	63.06	65.20	2524.17	2175.50	2273.00	0.30
PR14/15-1589 x PR14/15-1638	66.88	65.63	73.93	2032.75	2170.00	2409.00	0.37
PR14/15-1581 x PR14/15-1625	85.65	65.00	66.93	2780.92	2066.50	2090.00	0.37
PR14/15-1593 x PR14/15-1638	96.26	61.26	65.20	2597.67	2102.00	2273.00	0.38

PR14/15-1593 x PR14/15-1628	89.00	61.47	65.20	2506.47	2072.00	2273.00	0.38
PR14/15-1593 x PR14/15-1633	85.56	61.87	65.20	2682.00	2082.00	2273.00	0.39
PR14/15-1521 x PR14/15-1648	82.38	51.91	54.28	2402.67	1640.50	1741.00	0.46
PR14/15-1581 x PR14/15-1621	84.09	58.16	66.93	2443.08	1806.50	2043.00	0.46
PR14/15-1521 x PR14/15-1636	69.75	57.60	60.91	2329.75	1909.50	2078.00	0.48

Appendix M - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle yield and number of kernels per panicle of 102 hybrids.

Source	Panicle yield (g)	Mid-parent (g)	Better-parent (g)	No. kernels panicle^{-1}	Mid-parent	Better-parent	Genetic distance
PR14/15-1521 x PR14/15-1634	86.43	47.00	54.28	2734.25	1586.00	1741.00	0.50
PR14/15-1593 x PR14/15-1648	78.49	57.37	65.20	2104.92	1906.50	2273.00	0.51
PR14/15-1593 x PR14/15-1650	71.13	59.57	65.20	2125.00	2004.00	2273.00	0.52
PR14/15-1593 x PR14/15-1634	90.58	52.46	65.20	3008.00	1852.00	2273.00	0.54
PR14/15-1561 x PR14/15-1622	68.84	60.13	61.00	2249.79	1978.50	2002.00	0.31
PR14/15-1589 x PR14/15-1637	72.45	65.91	73.93	2173.67	2152.50	2409.00	0.37
PR14/15-1561 x PR14/15-1625	77.47	62.35	63.06	2495.50	2046.00	2090.00	0.46
PR14/15-1573 x PR14/15-1624	89.17	42.05	47.38	2715.75	1670.00	1747.00	0.47
PR14/15-1561 x PR14/15-1644	64.87	53.82	61.64	2010.83	1752.50	2002.00	0.54
PR14/15-1561 x PR14/15-1646	68.27	53.86	61.64	2054.17	1787.00	2002.00	0.57
PR14/15-1557 x PR14/15-1646	79.90	50.62	55.15	2497.75	1672.50	1773.00	0.57
PR14/15-1533 x PR14/15-1627	114.75	60.98	68.67	3563.83	1885.00	2005.00	0.35
PR14/15-1589 x PR14/15-1628	90.24	65.83	73.93	2667.33	2140.00	2409.00	0.37
PR14/15-1589 x PR14/15-1636	86.82	67.42	73.93	2808.92	2243.50	2409.00	0.38
PR14/15-1589 x PR14/15-1623	71.13	64.35	73.93	2165.58	2163.50	2409.00	0.49
PR14/15-1589 x PR14/15-1648	83.70	61.74	73.93	2255.33	1974.50	2409.00	0.52
PR14/15-1589 x PR14/15-1650	78.68	63.94	73.93	2307.50	2072.00	2409.00	0.53

PR14/15-1557 x PR14/15-1622	69.72	57.20	59.25	2248.92	1864.00	1955.00	0.27
PR14/15-1557 x PR14/15-1625	83.88	59.11	63.06	2762.17	1931.50	2090.00	0.34
PR14/15-1557 x PR14/15-1621	57.88	52.27	55.15	1721.75	1671.50	1773.00	0.48
PR14/15-1585 x PR14/15-1625	79.87	58.22	63.06	2623.20	1956.50	2090.00	0.47
PR14/15-1529 x PR14/15-1625	96.70	55.33	63.06	2894.83	1793.50	2090.00	0.39
PR14/15-1529 x PR14/15-1627	91.02	58.14	68.67	2747.17	1751.00	2005.00	0.47
PR14/15-1609 x PR14/15-1648	75.40	.	.	2117.17	.	.	0.34
Dekalb54-00	66.93	.	.	2187.67	.	.	.

Appendix N - Mean F1 hybrid performance, mid-parent value and better-parent value for panicle yield and number of kernels per panicle of 102 hybrids.

Source	Panicle yield (g)	Mid-parent (g)	Better-parent (g)	No.kernels panicle^{-1}	Mid-parent	Better-parent	Genetic distance
Pioneer84G62	65.00	.	.	2138.67	.	.	.
PR14/15-1609 x PR14/15-1649	80.93	.	.	2607.33	.	.	0.35
Seneca	50.34	.	.	1976.75	.	.	.
PR14/15-1609 x PR14/15-1635	86.07	.	.	2716.67	.	.	0.38
PR14/15-1577 x PR14/15-1625	70.62	.	.	2376.42	.	.	0.38
PR14/15-1609 x PR14/15-1650	76.97	.	.	2334.83	.	.	0.39
PR14/15-1597 x PR14/15-1635	80.78	.	.	2426.92	.	.	0.40
PR14/15-1597 x PR14/15-1628	89.83	.	.	2754.58	.	.	0.41
PR14/15-1597 x PR14/15-1634	82.77	.	.	2737.42	.	.	0.44
PR14/15-1597 x PR14/15-1638	77.97	.	.	2271.17	.	.	0.45
PR14/15-1601 x PR14/15-1638	83.89	.	.	2629.50	.	.	0.46
PR14/15-1597 x PR14/15-1650	70.47	.	.	2263.17	.	.	0.47
PR14/15-1597 x PR14/15-1649	77.93	.	.	2604.50	.	.	0.47
PR14/15-1597 x PR14/15-1633	87.44	.	.	2695.00	.	.	0.47

PR14/15-1597 x PR14/15-1648	86.25	.	.	2509.83	.	.	0.47
PR14/15-1601 x PR14/15-1637	69.18	.	.	2434.58	.	.	0.47
PR14/15-1601 x PR14/15-1649	62.09	.	.	2133.58	.	.	0.49
PR14/15-1605 x PR14/15-1635	97.46	.	.	2962.33	.	.	0.51
PR14/15-1601 x PR14/15-1650	72.97	.	.	2278.83	.	.	0.56
PR14/15-1601 x PR14/15-1633	76.08	.	.	2584.00	.	.	0.59
PR14/15-1565 x PR14/15-1626	77.02	57.83		2259.58	1886.00		0.34
PR14/15-1545 x PR14/15-1626	74.83	46.43		2296.33	1603.00		0.34
PR14/15-1609 x PR14/15-1638	85.43	.		2445.33	.		0.37
PR14/15-1609 x PR14/15-1636	76.91	.		2694.92	.		0.37
PR14/15-1605 x PR14/15-1648	84.97	.	.	2476.40	.	.	0.38

Appendix O - Mean F1 hybrid performance, mid-parent value and better-parent value for thousand kernel weight and plant height of 102 hybrids.

Source	Thousand kernel weight (g)	Mid-parent (g)	Better-parent (g)	Plant height (cm)	Mid-parent (cm)	Better-parent (cm)	Genetic distance
PR14/15-1509 x PR14/15-1637	29.61	32.15	33.59	128.69	120.97	127.21	0.38
PR14/15-1509 x PR14/15-1650	37.35	33.07	33.59	148.59	122.56	130.39	0.53
PR14/15-1501 x PR14/15-1633	29.17	29.93	30.37	139.91	116.72	121.50	0.33
PR14/15-1505 x PR14/15-1635	31.84	30.64	30.71	132.29	121.08	125.31	0.12
PR14/15-1549 x PR14/15-1625	30.19	29.65	30.42	150.92	127.74	129.75	0.30
PR14/15-1505 x PR14/15-1638	34.15	30.83	31.10	135.25	117.98	119.12	0.33
PR14/15-1505 x PR14/15-1637	32.13	30.63	30.70	119.59	122.03	127.21	0.38
PR14/15-1569 x PR14/15-1627	30.96	30.96	34.68	140.97	117.90	122.77	0.46
PR14/15-1549 x PR14/15-1627	35.22	31.78	34.68	173.35	119.38	125.73	0.48
PR14/15-1549 x PR14/15-1623	30.58	28.70	28.87	135.04	125.10	125.73	0.49
PR14/15-1569 x PR14/15-1630	29.80	29.04	30.84	124.25	125.31	127.85	0.50
PR14/15-1505 x PR14/15-1650	32.17	31.56	32.55	131.66	123.62	130.39	0.51
PR14/15-1569 x PR14/15-1641	33.34	31.26	35.29	121.92	122.35	122.77	0.55
PR14/15-1545 x PR14/15-1622	31.72	30.21	30.54	148.59	125.99	132.51	0.47
PR14/15-1545 x PR14/15-1631	30.44	26.84	29.88	135.47	121.50	132.51	0.50
PR14/15-1501 x PR14/15-1628	30.37	31.44	33.40	127.21	112.69	113.45	0.39
PR14/15-1537 x PR14/15-1625	31.47	30.70	30.98	153.67	130.29	130.82	0.45
PR14/15-1537 x PR14/15-1622	31.93	30.76	30.98	133.56	125.15	130.82	0.47
PR14/15-1501 x PR14/15-1622	31.64	30.01	30.54	136.31	115.70	119.47	0.50
PR14/15-1501 x PR14/15-1634	30.70	30.97	32.46	134.12	115.23	118.53	0.51
PR14/15-1501 x PR14/15-1636	29.24	29.56	29.63	130.81	120.96	129.98	0.52
PR14/15-1501 x PR14/15-1650	33.76	31.02	32.55	136.95	121.16	130.39	0.56
PR14/15-1537 x PR14/15-1647	32.67	30.86	30.98	164.04	125.95	130.82	0.57

PR14/15-1501 x PR14/15-1649	30.64	31.16	32.84	126.15	111.32	111.93	0.70
PR14/15-1565 x PR14/15-1622	30.01	30.48	30.54	120.88	122.81	126.15	0.24

Appendix P - Mean F1 hybrid performance, mid-parent value and better-parent value for thousand kernel weight and plant height of 102 hybrids.

Source	Thousand kernel weight (g)	Mid-parent (g)	Better-parent (g)	Plant height (cm)	Mid-parent (cm)	Better-parent (cm)	Genetic distance
Pioneer84G62	30.51	.	.	129.54	.	.	.
PR14/15-1609 x PR14/15-1649	31.45	.	.	147.32	.	.	0.35
Seneca	25.59	.	.	115.99	.	.	.
PR14/15-1609 x PR14/15-1635	31.86	.	.	155.36	.	.	0.38
PR14/15-1577 x PR14/15-1625	29.81	.	.	146.26	.	.	0.38
PR14/15-1609 x PR14/15-1650	33.25	.	.	140.55	.	.	0.39
PR14/15-1597 x PR14/15-1635	33.31	.	.	149.01	.	.	0.40
PR14/15-1597 x PR14/15-1628	32.85	.	.	137.58	.	.	0.41
PR14/15-1597 x PR14/15-1634	30.44	.	.	146.47	.	.	0.44
PR14/15-1597 x PR14/15-1638	34.30	.	.	152.40	.	.	0.45
PR14/15-1601 x PR14/15-1638	31.99	.	.	148.59	.	.	0.46
PR14/15-1597 x PR14/15-1650	31.53	.	.	147.32	.	.	0.47
PR14/15-1597 x PR14/15-1649	30.18	.	.	140.55	.	.	0.47
PR14/15-1597 x PR14/15-1633	32.36	.	.	146.05	.	.	0.47
PR14/15-1597 x PR14/15-1648	35.80	.	.	161.71	.	.	0.47
PR14/15-1601 x PR14/15-1637	28.60	.	.	138.64	.	.	0.47
PR14/15-1601 x PR14/15-1649	29.43	.	.	127.42	.	.	0.49
PR14/15-1605 x PR14/15-1635	32.91	.	.	138.85	.	.	0.51
PR14/15-1601 x PR14/15-1650	32.48	.	.	133.98	.	.	0.56

PR14/15-1601 x PR14/15-1633	29.62	.	.	138.85	.	.	0.59
PR14/15-1565 x PR14/15-1626	33.62	30.42		138.85	126.15		0.34
PR14/15-1545 x PR14/15-1626	32.70	29.88		118.11	132.51		0.34
PR14/15-1609 x PR14/15-1638	35.05	.		151.13	.		0.37
PR14/15-1609 x PR14/15-1636	29.04	.		143.51	.		0.37
PR14/15-1605 x PR14/15-1648	34.13	.	.	153.42	.	.	0.38

Appendix Q - Mean F1 hybrid performance, mid-parent value and better-parent value for thousand kernel weight and plant height of 102 hybrids.

Source	Thousand kernel weight (g)	Mid-parent (g)	Better-parent (g)	Plant height (cm)	Mid-parent (cm)	Better-parent (cm)	Genetic distance
PR14/15-1521 x PR14/15-1634	31.75	31.97	32.46	142.24	120.01	121.49	0.50
PR14/15-1593 x PR14/15-1648	37.12	30.78	32.83	159.60	114.09	115.57	0.51
PR14/15-1593 x PR14/15-1650	33.49	30.64	32.55	151.98	121.50	130.39	0.52
PR14/15-1593 x PR14/15-1634	30.28	30.60	32.46	147.32	115.57	118.53	0.54
PR14/15-1561 x PR14/15-1622	31.03	30.80	31.06	136.07	120.06	120.65	0.31
PR14/15-1589 x PR14/15-1637	33.42	31.04	31.38	136.52	91.76	127.21	0.37
PR14/15-1561 x PR14/15-1625	31.13	30.74	31.06	139.28	125.20	129.75	0.46
PR14/15-1573 x PR14/15-1624	33.00	27.20	30.14	138.22	124.25	132.08	0.47
PR14/15-1561 x PR14/15-1644	32.31	31.97	32.87	135.47	122.13	123.60	0.54
PR14/15-1561 x PR14/15-1646	32.88	30.18	31.06	140.97	122.98	125.31	0.57
PR14/15-1557 x PR14/15-1646	32.14	30.27	31.24	139.28	125.52	125.73	0.57
PR14/15-1533 x PR14/15-1627	32.11	32.55	34.68	150.28	118.26	123.48	0.35
PR14/15-1589 x PR14/15-1628	34.33	32.39	33.40	154.09	84.88	113.45	0.37
PR14/15-1589 x PR14/15-1636	30.96	30.51	31.38	145.63	93.14	129.98	0.38
PR14/15-1589 x PR14/15-1623	32.40	29.95	31.38	143.30	90.38	124.46	0.49

PR14/15-1589 x PR14/15-1648	38.42	32.11	32.83	162.98	85.94	115.57	0.52
PR14/15-1589 x PR14/15-1650	34.34	31.97	32.55	155.79	93.35	130.39	0.53
PR14/15-1557 x PR14/15-1622	31.29	30.89	31.24	135.89	122.60	125.73	0.27
PR14/15-1557 x PR14/15-1625	30.40	30.83	31.24	138.43	127.74	129.75	0.34
PR14/15-1557 x PR14/15-1621	33.55	32.13	33.01	134.41	121.92	125.73	0.48
PR14/15-1585 x PR14/15-1625	30.48	30.28	30.42	146.90	126.98	129.75	0.47
PR14/15-1529 x PR14/15-1625	31.71	31.33	32.24	143.51	125.92	129.75	0.39
PR14/15-1529 x PR14/15-1627	32.76	33.46	34.68	166.58	117.56	122.08	0.47
PR14/15-1609 x PR14/15-1648	35.66	.	.	153.67	.	.	0.34
Dekalb54-00	30.72	.	.	138.43	.	.	.

Appendix R - Mean F1 hybrid performance, mid-parent value and better-parent value for thousand kernel weight and plant height of 102 hybrids.

Source	Thousand kernel weight (g)	Mid-parent (g)	Better-parent (g)	Plant height (cm)	Mid-parent (cm)	Better-parent (cm)	Genetic distance
Pioneer84G62	30.51	.	.	129.54	.	.	.
PR14/15-1609 x PR14/15-1649	31.45	.	.	147.32	.	.	0.35
Seneca	25.59	.	.	115.99	.	.	.
PR14/15-1609 x PR14/15-1635	31.86	.	.	155.36	.	.	0.38
PR14/15-1577 x PR14/15-1625	29.81	.	.	146.26	.	.	0.38
PR14/15-1609 x PR14/15-1650	33.25	.	.	140.55	.	.	0.39
PR14/15-1597 x PR14/15-1635	33.31	.	.	149.01	.	.	0.40
PR14/15-1597 x PR14/15-1628	32.85	.	.	137.58	.	.	0.41
PR14/15-1597 x PR14/15-1634	30.44	.	.	146.47	.	.	0.44
PR14/15-1597 x PR14/15-1638	34.30	.	.	152.40	.	.	0.45
PR14/15-1601 x PR14/15-1638	31.99	.	.	148.59	.	.	0.46

PR14/15-1597 x PR14/15-1650	31.53	.	.	147.32	.	.	0.47
PR14/15-1597 x PR14/15-1649	30.18	.	.	140.55	.	.	0.47
PR14/15-1597 x PR14/15-1633	32.36	.	.	146.05	.	.	0.47
PR14/15-1597 x PR14/15-1648	35.80	.	.	161.71	.	.	0.47
PR14/15-1601 x PR14/15-1637	28.60	.	.	138.64	.	.	0.47
PR14/15-1601 x PR14/15-1649	29.43	.	.	127.42	.	.	0.49
PR14/15-1605 x PR14/15-1635	32.91	.	.	138.85	.	.	0.51
PR14/15-1601 x PR14/15-1650	32.48	.	.	133.98	.	.	0.56
PR14/15-1601 x PR14/15-1633	29.62	.	.	138.85	.	.	0.59
PR14/15-1565 x PR14/15-1626	33.62	30.42		138.85	126.15		0.34
PR14/15-1545 x PR14/15-1626	32.70	29.88		118.11	132.51		0.34
PR14/15-1609 x PR14/15-1638	35.05	.		151.13	.		0.37
PR14/15-1609 x PR14/15-1636	29.04	.		143.51	.		0.37
PR14/15-1605 x PR14/15-1648	34.13	.	.	153.42	.	.	0.38

Appendix S - Mean F1 hybrid performance, mid-parent value and better-parent value for days to flowering and grain yield of 102 hybrids.

Source	Days to flowering	Mid-parent	Better-parent	Grain yield (kg ha^{-1})	Mid-parent (kg ha^{-1}	Better-parent (kg ha^{-1})	Genetic distance
PR14/15-1509 x PR14/15-1637	59.17	64.00	60	11683.78	8718.70	9605.90	0.38
PR14/15-1509 x PR14/15-1650	54.67	65.50	60	11594.63	7781.19	7831.49	0.53
PR14/15-1501 x PR14/15-1633	61.00	67.00	61	13041.17	11444.37	11970.74	0.33
PR14/15-1505 x PR14/15-1635	62.00	63.50	62	15766.03	7712.41	8368.80	0.12
PR14/15-1549 x PR14/15-1625	63.67	70.00	62	11680.02	9941.44	11598.91	0.30
PR14/15-1505 x PR14/15-1638	66.83	65.00	62	10563.10	8134.94	8368.80	0.33
PR14/15-1505 x PR14/15-1637	62.67	65.00	62	12703.21	8987.35	9605.90	0.38
PR14/15-1569 x PR14/15-1627	65.67	69.00	62	14792.66	9906.02	9115.03	0.46
PR14/15-1549 x PR14/15-1627	65.33	69.00	62	9899.17	8699.50	9115.03	0.48
PR14/15-1549 x PR14/15-1623	61.83	66.00	62	11745.25	8393.56	8503.15	0.49
PR14/15-1569 x PR14/15-1630	62.50	64.00	62	11878.36	10981.94	.	0.50
PR14/15-1505 x PR14/15-1650	57.67	66.50	62	12835.39	8049.85	7730.89	0.51
PR14/15-1569 x PR14/15-1641	62.67	64.50	62	8791.63	9232.15	10697.29	0.55
PR14/15-1545 x PR14/15-1622	60.33	70.50	63	13507.59	8762.47	9268.71	0.47
PR14/15-1545 x PR14/15-1631	65.83	70.00	63	10739.88	10402.15	12268.71	0.50
PR14/15-1501 x PR14/15-1628	63.67	67.00	64	12377.92	11419.00	11919.99	0.39
PR14/15-1537 x PR14/15-1625	61.83	71.00	64	12256.44	11683.78	11768.65	0.45
PR14/15-1537 x PR14/15-1622	62.00	71.00	64	11441.31	8512.44	11768.65	0.47
PR14/15-1501 x PR14/15-1622	65.67	71.00	64	11031.93	8087.12	9918.00	0.50
PR14/15-1501 x PR14/15-1634	64.00	72.00	64	12353.14	9481.59	9879.00	0.51
PR14/15-1501 x PR14/15-1636	60.00	67.00	64	11589.75	10142.49	10918.00	0.52
PR14/15-1501 x PR14/15-1650	62.83	67.50	64	8579.19	9324.45	10918.00	0.56
PR14/15-1537 x PR14/15-1647	60.83	66.00	64	10980.07	11557.85	11768.65	0.57
PR14/15-1501 x PR14/15-1649	60.50	64.50	64	12158.59	11121.95	11325.89	0.70
PR14/15-1565 x PR14/15-1622	59.50	71.50	65	9229.67	8019.56	10782.88	0.24

Appendix T - Mean F1 hybrid performance, mid-parent value and better-parent value for days to flowering and grain yield of 102 hybrids.

Source	Days to flowering	Mid-parent	Better-parent	Grain yield (kg ha^{-1})	Mid-parent (kg ha^{-1}	Better-parent (kg ha^{-1})	Genetic distance
PR14/15-1589 x PR14/15-1635	57.33	67.00	65	11892.44	8322.11	9588.20	0.29
PR14/15-1593 x PR14/15-1635	59.33	66.00	65	10041.89	8072.84	9089.65	0.30
PR14/15-1553 x PR14/15-1622	66.20	71.50	65	10913.57	7295.21	9334.18	0.33
PR14/15-1521 x PR14/15-1635	60.33	66.00	65	11353.43	7599.06	8142.10	0.37
PR14/15-1517 x PR14/15-1635	61.33	67.00	65	9889.15	8662.40	10268.77	0.38
PR14/15-1565 x PR14/15-1628	60.17	67.50	65	11660.38	11351.44	11919.99	0.39
PR14/15-1553 x PR14/15-1624	60.83	70.50	65	9098.02	10203.42	10782.88	0.46
PR14/15-1565 x PR14/15-1625	62.33	71.50	65	12427.00	11190.90	11598.91	0.46
PR14/15-1553 x PR14/15-1623	56.00	67.50	65	11261.67	8918.67	9334.18	0.47
PR14/15-1553 x PR14/15-1631	64.17	71.00	65	11889.91	8934.89	9334.18	0.47
PR14/15-1565 x PR14/15-1621	57.83	72.50	65	9544.61	9933.94	10782.88	0.48
PR14/15-1565 x PR14/15-1630	53.50	65.50	65	9054.07	11024.88	11266.88	0.50
PR14/15-1553 x PR14/15-1630	53.50	65.50	65	12773.81	10300.53	11266.88	0.50
PR14/15-1565 x PR14/15-1645	53.33	67.50	65	9824.72	8328.53	10782.88	0.56
PR14/15-1521 x PR14/15-1650	60.67	69.00	67	11577.12	7936.50	8142.10	0.27
PR14/15-1521 x PR14/15-1633	59.83	68.50	67	11164.47	10056.05	11970.00	0.30
PR14/15-1593 x PR14/15-1636	59.83	68.50	67	11777.39	9228.32	9366.98	0.30
PR14/15-1589 x PR14/15-1638	65.67	68.50	67	14835.97	8744.64	9588.20	0.37
PR14/15-1581 x PR14/15-1625	65.00	72.50	67	9642.91	10012.78	11598.91	0.37
PR14/15-1593 x PR14/15-1638	66.83	67.50	67	11795.28	8495.36	9089.65	0.38
PR14/15-1593 x PR14/15-1628	65.83	68.50	67	11497.39	10504.82	11919.99	0.38
PR14/15-1593 x PR14/15-1633	57.67	68.50	67	12011.52	10530.20	11970.74	0.39

PR14/15-1521 x PR14/15-1648	63.00	69.50	67	12565.97	8798.98	9455.86	0.46
PR14/15-1581 x PR14/15-1621	64.17	73.50	67	15395.68	8755.95	9085.24	0.46
PR14/15-1521 x PR14/15-1636	59.50	68.50	67	13328.96	8754.54	9366.98	0.48

Appendix U - Mean F1 hybrid performance, mid-parent value and better-parent value for days to flowering and grain yield of 102 hybrids.

Source	Days to flowering	Mid-parent	Better-parent	Grain yield (kg ha^{-1})	Mid-parent (kg ha^{-1}	Better-parent (kg ha^{-1})	Genetic distance
PR14/15-1521 x PR14/15-1634	63.00	73.50	67	10787.74	8093.64	8142.10	0.50
PR14/15-1593 x PR14/15-1648	62.83	69.50	67	13660.06	9272.76	9455.86	0.51
PR14/15-1593 x PR14/15-1650	58.33	69.00	67	11783.98	8410.27	9089.65	0.52
PR14/15-1593 x PR14/15-1634	60.33	73.50	67	13950.64	8567.42	9089.65	0.54
PR14/15-1561 x PR14/15-1622	64.29	73.00	68	9084.52	6975.91	8695.59	0.31
PR14/15-1589 x PR14/15-1637	61.00	68.50	68	13374.75	9597.05	9605.90	0.37
PR14/15-1561 x PR14/15-1625	62.17	73.00	68	11824.48	10147.25	11598.91	0.46
PR14/15-1573 x PR14/15-1624	65.33	72.00	68	12728.90	9186.92	9623.96	0.47
PR14/15-1561 x PR14/15-1644	60.67	69.50	68	10594.08	9237.41	9779.22	0.54
PR14/15-1561 x PR14/15-1646	58.33	68.00	68	11544.68	8573.74	8695.59	0.57
PR14/15-1557 x PR14/15-1646	61.00	69.00	68	11750.51	8692.27	8932.64	0.57
PR14/15-1533 x PR14/15-1627	64.67	72.50	69	11415.89	9832.51	10549.98	0.35
PR14/15-1589 x PR14/15-1628	58.33	69.50	69	12462.35	10754.10	11919.99	0.37
PR14/15-1589 x PR14/15-1636	57.17	69.50	69	14828.72	9477.59	9588.20	0.38
PR14/15-1589 x PR14/15-1623	59.83	69.50	69	13892.45	9045.68	9588.20	0.49
PR14/15-1589 x PR14/15-1648	63.67	70.50	69	13608.38	9522.03	9588.20	0.52
PR14/15-1589 x PR14/15-1650	57.17	70.00	69	11980.55	8659.55	9588.00	0.53
PR14/15-1557 x PR14/15-1622	61.50	74.00	70	12171.97	7094.44	8932.64	0.27
PR14/15-1557 x PR14/15-1625	63.67	74.00	70	9755.16	10265.78	11598.91	0.34

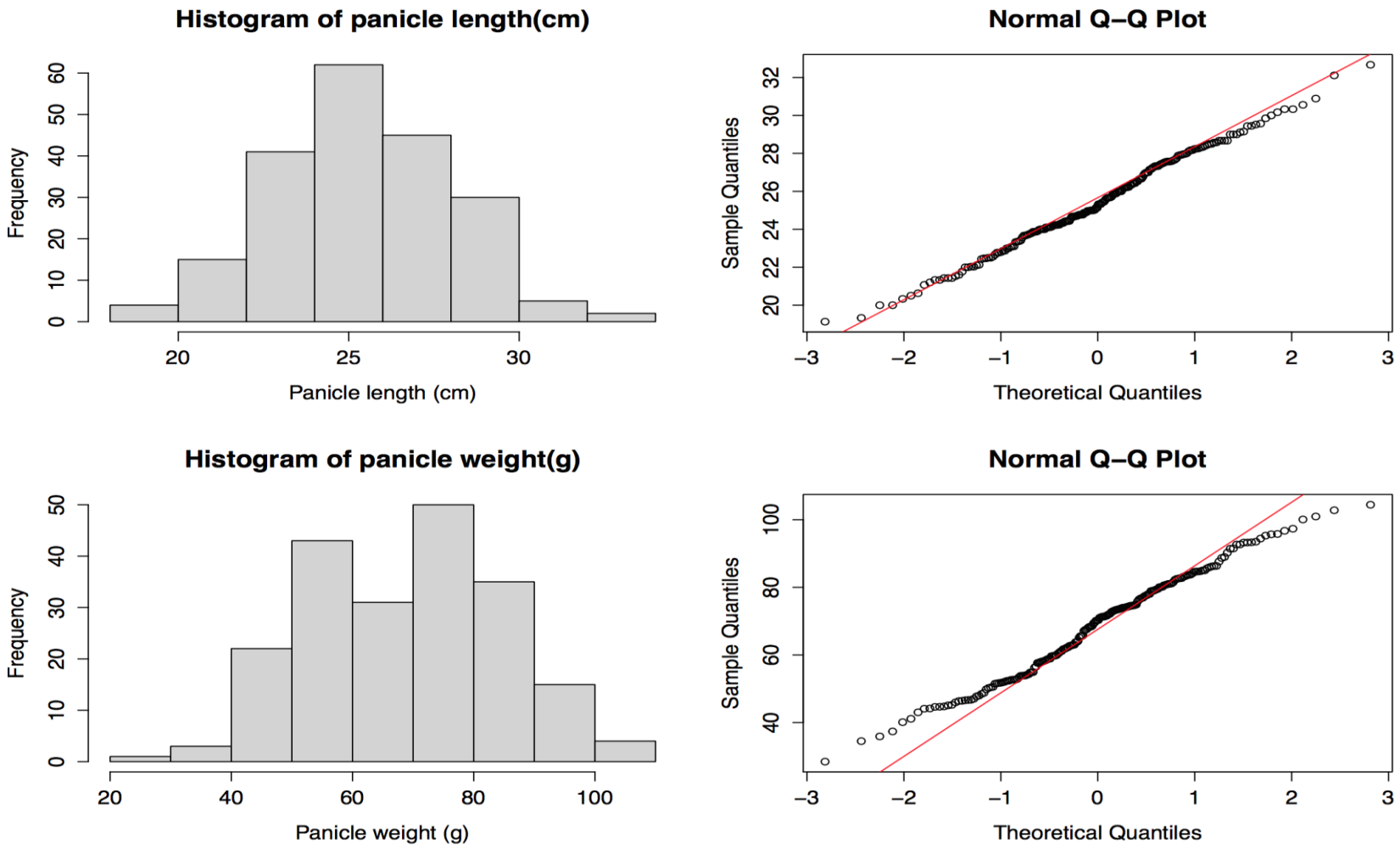
PR14/15-1557 x PR14/15-1621	58.33	75.00	70	8471.81	9008.94	9085.24	0.48
PR14/15-1585 x PR14/15-1625	64.67	75.00	72	9256.25	9905.23	9598.91	0.47
PR14/15-1529 x PR14/15-1625	64.00	76.00	74	8497.59	11242.95	11598.91	0.39
PR14/15-1529 x PR14/15-1627	63.67	75.00	74	12482.87	10001.01	10886.98	0.47
PR14/15-1609 x PR14/15-1648	64.50	.	.	11605.31	.	.	0.34
Dekalb54-00	62.17	.	.	10207.44	.	.	.

Appendix V - Mean F1 hybrid performance, mid-parent value and better-parent value for days to flowering and grain yield of 102 hybrids.

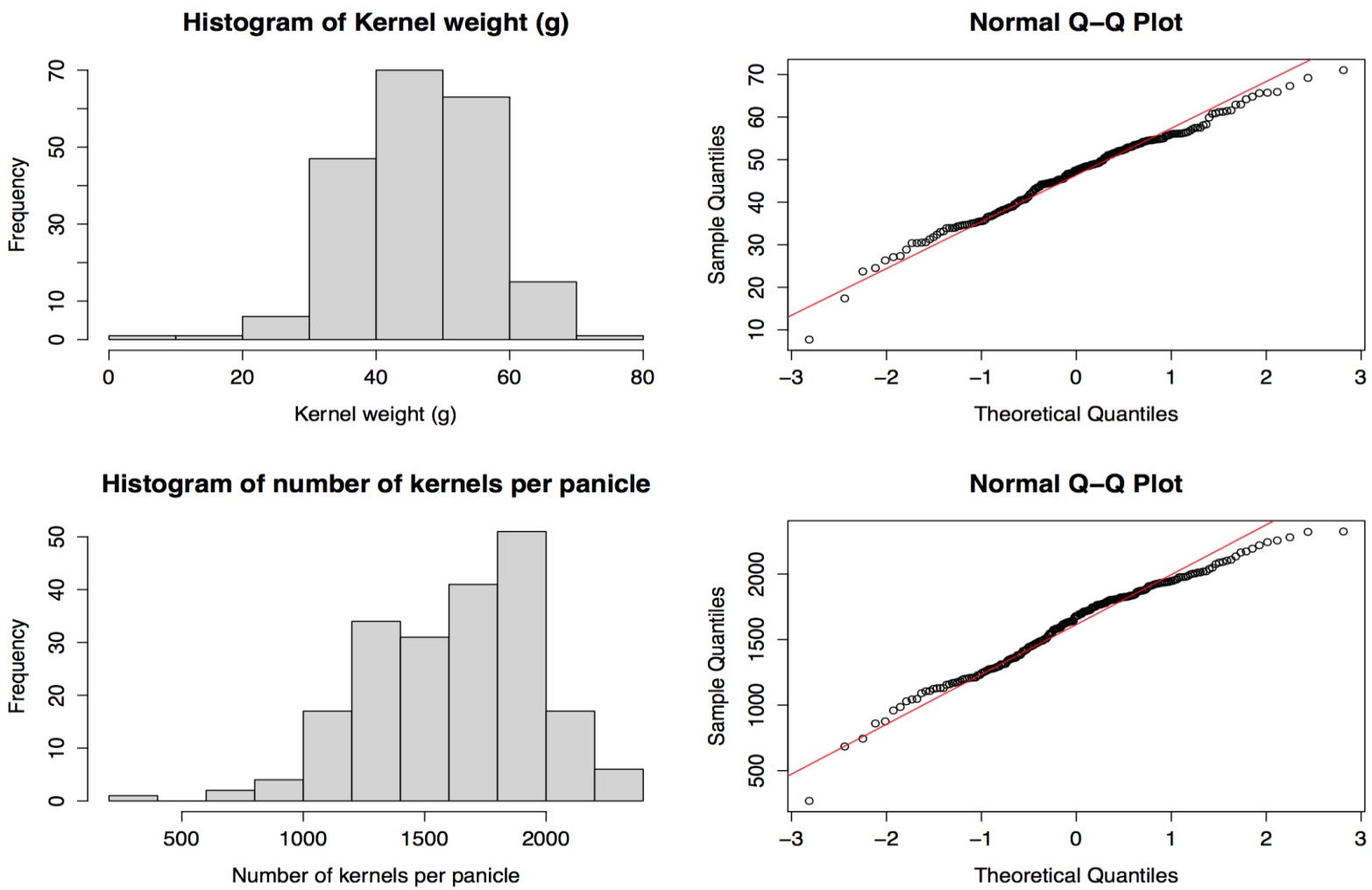
Source	Days to flowering	Mid-parent	Better-parent	Grain yield (kg ha^{-1})	Mid-parent (kg ha^{-1}	Better-parent (kg ha^{-1})	Genetic distance
Pioneer84G62	56.67	.	.	10254.70	.	.	.
PR14/15-1609 x PR14/15-1649	59.17	.	.	12807.65	.	.	0.35
Seneca	53.00	.	.	11196.45	.	.	.
PR14/15-1609 x PR14/15-1635	61.33	.	.	9120.76	.	.	0.38
PR14/15-1577 x PR14/15-1625	65.67	.	.	11063.40	.	.	0.38
PR14/15-1609 x PR14/15-1650	59.83	.	.	10347.97	.	.	0.39
PR14/15-1597 x PR14/15-1635	60.00	.	.	11591.98	.	.	0.40
PR14/15-1597 x PR14/15-1628	62.17	.	.	11893.47	.	.	0.41
PR14/15-1597 x PR14/15-1634	62.50	.	.	11225.08	.	.	0.44
PR14/15-1597 x PR14/15-1638	65.83	.	.	11928.71	.	.	0.45
PR14/15-1601 x PR14/15-1638	66.17	.	.	10245.29	.	.	0.46
PR14/15-1597 x PR14/15-1650	59.40	.	.	9316.45	.	.	0.47
PR14/15-1597 x PR14/15-1649	61.00	.	.	11366.55	.	.	0.47
PR14/15-1597 x PR14/15-1633	59.67	.	.	11744.69	.	.	0.47
PR14/15-1597 x PR14/15-1648	63.83	.	.	11410.35	.	.	0.47
PR14/15-1601 x PR14/15-1637	60.50	.	.	11830.39	.	.	0.47
PR14/15-1601 x PR14/15-1649	58.83	.	.	12134.81	.	.	0.49
PR14/15-1605 x PR14/15-1635	63.33	.	.	9631.85	.	.	0.51
PR14/15-1601 x PR14/15-1650	60.17	.	.	11492.50	.	.	0.56

PR14/15-1601 x PR14/15-1633	59.67	.	.	11692.93	.	.	0.59
PR14/15-1565 x PR14/15-1626	62.50	65.00		9384.89	10782.88		0.34
PR14/15-1545 x PR14/15-1626	59.00	63.00		11127.73	12268.71	.	0.34
PR14/15-1609 x PR14/15-1638	64.50	.		9803.67	.	.	0.37
PR14/15-1609 x PR14/15-1636	64.33	.		9618.25	.		0.37
PR14/15-1605 x PR14/15-1648	64.83	.		10487.26	.	.	0.38

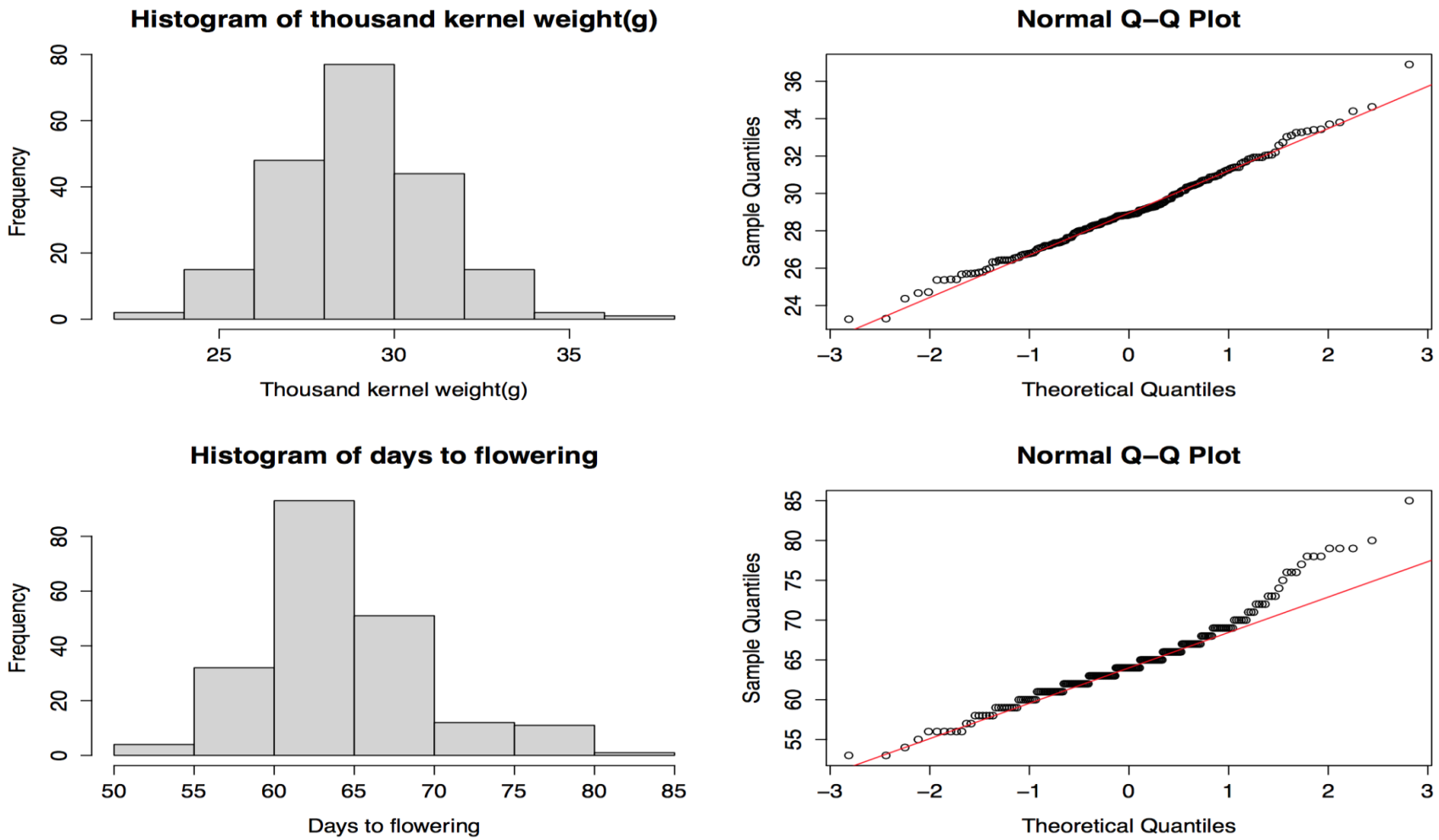
Note: . indicates that the data were not collected due to the missing data.



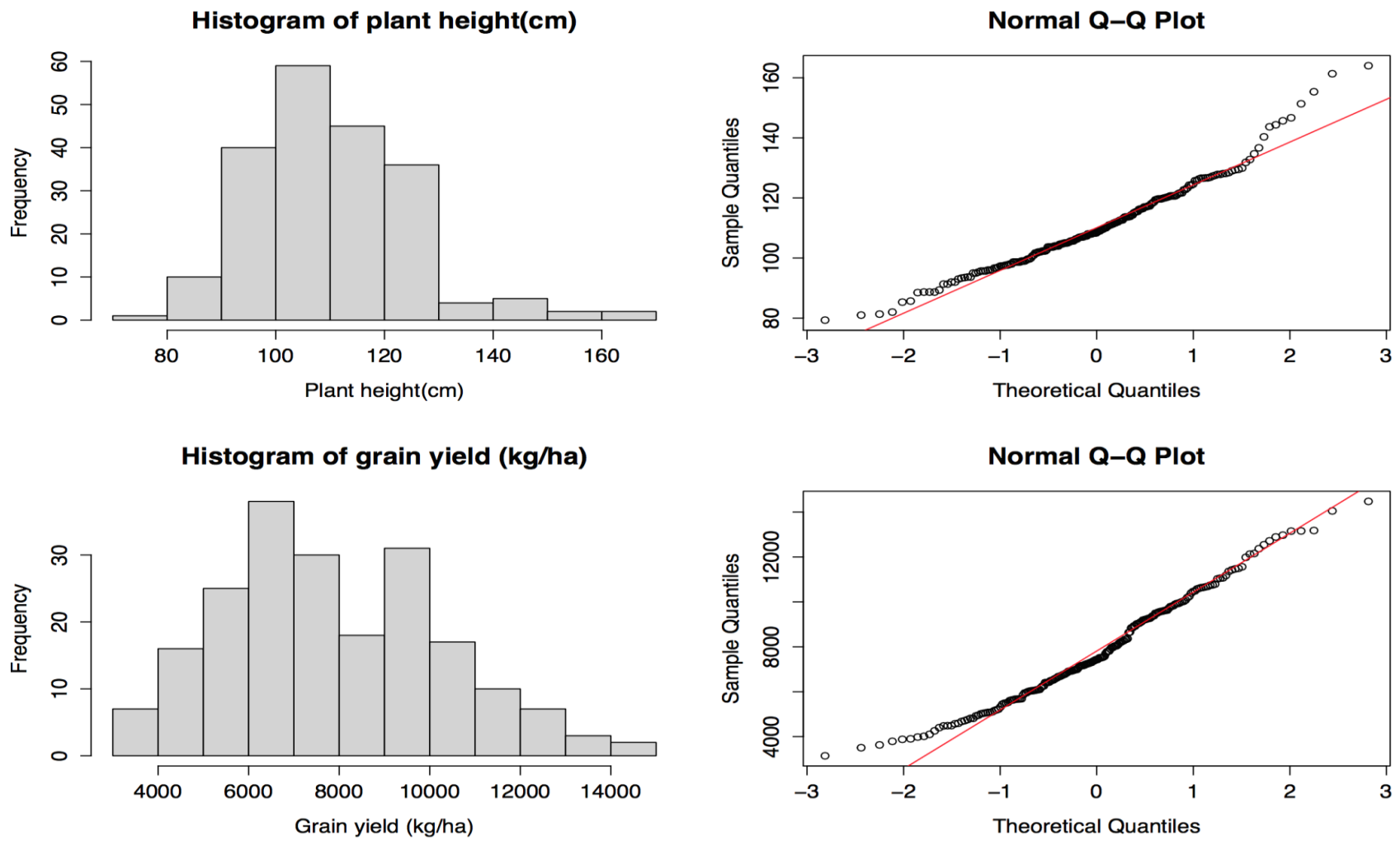
Appendix W - Histograms of panicle length (cm) and panicle weight (g) with their respective Q-Q plots.



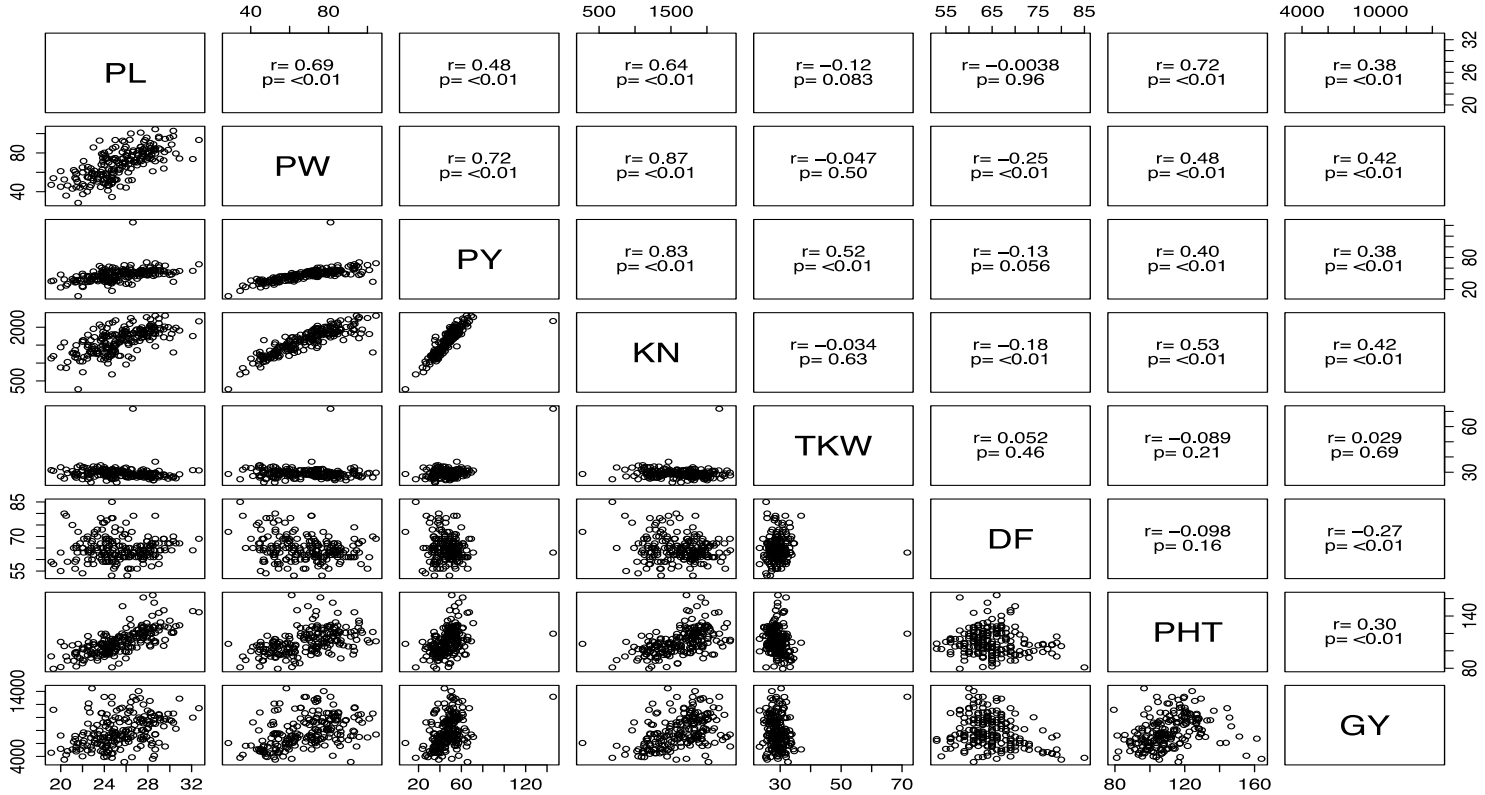
Appendix X - Histograms of kernel weight (g) and number of kernels per panicle with their respective Q-Q plots.



Appendix Y - Histograms of thousand kernel weight (g) and days to flowering with their respective Q-Q plots.



Appendix Z - Histograms of plant height (cm) and grain yield (kg ha^{-1}) with their respective Q-Q plots.



Appendix AA - The Pearson's correlations between all the eight traits. PL=Panicle length (cm), PW = Panicle weight (g); PY=Panicle yield (g), KN= Number of kernels per panicle, TKW = Thousand kernel weight (g), DF = Days to 50% flowering; PH = Plant height (cm); GY= Grain yield (kg ha^{-1}).