ASSESSING THE GENETIC VARIATION, PERFORMANCE, AND GENOMIC PREDICTION OF GRAIN SORGHUM HYBRIDS DEVELOPED FROM U.S. PUBLIC BREEDING PROGRAMS

A Dissertation

by

JALES MENDES OLIVEIRA FONSECA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	William L. Rooney
Committee Members,	Ramasamy Perumal
	Patricia E. Klein
	Alan R. Dabney
Head of Department,	David D. Baltensperger

August 2021

Major Subject: Plant Breeding

ABSTRACT

Many independent events led by public institutions significantly increased grain sorghum yield in the US, including the development of grain sorghum hybrids and the introgression of exotic germplasm. Nonetheless, the rates of genetic gain have been modest, and hence new strategies should be explored. This study assesses alternatives to maximize genetic gain via elite germplasm enhancement based on exchanging inbred lines derived from distinct sorghum breeding programs. Additionally, it presents classical and genomic prediction models that explore combining abilities and their interaction with environments to predict the performance of grain sorghum hybrids. Lastly, it evaluates the adaptability and stability of grain sorghum hybrids within mega-environments. For that, ten elite A- and R-lines from Texas A&M and Kansas State sorghum breeding programs were crossed in a factorial design to generate 100 hybrids. Hybrid combinations were grouped to represent hybrids within and across programs. Grain yield, plant height, and days to anthesis were measured in ten environments over two years. Results indicate that crosses between elite inbred lines developed from distinct sorghum breeding programs can increase selectable variation and generate top performers. General and specific combining abilities and their interaction with the environment can effectively predict hybrid performance, and the inclusion of genomic information further increased the prediction accuracy of models. The mega-environment analysis identified the established subtropical and temperate sorghum production regions and suggested that hybrid combinations between Texas A&M and Kansas State sorghum breeding programs can generate high-performing and stable grain sorghum hybrids across target regions. Plant breeders, growers, and seed companies are encouraged to explore such crosses to produce promising new products. Finally, the opportunity the exchange elite germplasm across breeding programs could foster collaborative efforts between public institutions and enhance the rate of genetic gain in crops with limited public and private resources.

DEDICATION

To the memory of my mother, Eloisa Fraga de Oliveira Fonseca, for her incommensurable love and care; and to my father, Marcio Fonseca da Silva, who always supported me throughout my academic carrier.

ACKNOWLEDGEMENTS

First and foremost, thanks be to God. He gives us the victory through our Lord Jesus Christ. Without him, none of this would have been possible.

Mon and Dad, your love and support are the most precious thing I could ever ask for. Thank you so much for encouraging and assisting me throughout my academic career and life. To my sisters, I am thankful to have you uplifting me when difficult times arrive (and many have arrived during this degree). To my beautiful girlfriend, I cannot quantify how important your presence was during this journey. You were the first and the last person I spoke with every single day since I arrived in the US. Thank you for your patience, love, and encouragement. To the rest of my family, thank you for always believing in me and praying for me. Without each one of you, I would not have made it this far.

Dr. William L. Rooney, I am extremely grateful for the opportunity you have given me. Beyond the guidance related to my academic development and completion of my degree, you have inspired me to become a better scientist, writer, and leader. You have been an extraordinary mentor and a role model to me; thank you very much. To all my committee members, Dr. Ramasamy Perumal, Dr. Patricia E. Klein, and Alan R. Dabney thank you for your guidance, insight, and support throughout this research.

Thanks to my friends from the Texas A&M AgriLife Research Sorghum Breeding Lab: Delroy Collins, Vickie Pineda, Stephen Labar, Dr. Leo Hoffmann Jr., Dr. Geraldo Carvalho, Dr. Francisco Gomes, Dr. Bryan Pfeiffer, Dr. Ace Pugh, Dr. David

Horne, Dr. Lorin Harvey, Ammani Kyanam, Daniel Crozier, Mitchell Kent, Fabian Leon, Noah Winans, Kayla Beechinor, and all student workers for the support, especially during the summer season.

Thanks to the Soil and Crop Science Graduate Student Organization and the Texas A&M Plant Breeding Symposium for allowing me to develop leadership skills and connecting with the university and the academic community around the world.

To all my friends in Brazil who invested some of their time to enjoy a cup of coffee with and listen to my project, thank you for these relaxing yet very important moments. Also, thanks to the "cool plant breeders," Dr. Nolan Bentley, Karina Morales, and Ammani Kyanam for creating a joyful atmosphere during my graduate degree.

To the CIMMYT scientists Dr. Jose Crossa, Dr. Paulino Pérez-Rodríguez, and Dr. Angela Pacheco for teaching me genomic prediction and selection. Finally, an especial thanks to my former advisors Dr. José Airton for teaching me statistics the best way possible and Dr. João Andrade for introducing me to this outstanding field of science called plant breeding.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Dr. William L. Rooney (advisor) of the Department of Soil and Crop Sciences and Professors Dr. Patricia E. Klein of the Department of Horticultural Sciences, Dr. Alan R. Dabney of the Department of Statistics, Texas A&M University, and Dr. Ramasamy Perumal of the Department of Agronomy, Kansas State University.

Dr. Ramasamy Perumal of Kansas State University conducted the experiments at Kansas locations and assisted with data collection.

Dr. Robert Klein, Dr. Patricia Klein, and Julie McCollum of Texas A&M University contributed to extract DNA and develop SNP markers for the sorghum inbred lines used to generate hybrids.

Dr. Jose Crossa, Dr. Paulino Pérez-Rodrigues, and Dr. Ángela Pacheco of CYMMIT contributed to developing genomic prediction models applied in chapter 3 of this dissertation. All other work conducted for the dissertation was completed independently by Jales Fonseca.

The graduate study was supported by a research assistantship funded by a grant from Monsanto, Texas A&M AgriLife Research, and the Texas A&M sorghum breeding program. Additionally, the Dudley Smith Family Travel Scholarship provided an enhancement fund, allowing the collaboration with CIMMYT scientists.

NOMENCLATURE

AMMI	Additive Main Effects and Multiplicative Interaction
AEC	Average Environment Coordinate
AES	Agricultural Experimental Stations
BFW	Bayesian Finlay-Wilkinson regression
BLUE	Best Linear Unbiased Estimates
BLUP	Best Linear Unbiased Prediction
BRR	Bayesian Ridge Regression model
BV	Breeding Value
CI	Coincidence Index
COI	Crossover Interaction
CV	Cross-validation
CVe	Coefficient of Variation
DA	Days to Anthesis
GB	Genomic Bayesian model
GBS	Genotyping-by-Sequencing
GCA	General Combining Ability
GGB	Genotype plus Genotype × Block of environments
GGE	Genotype plus Genotype × Environment
GRM	Genomic Relationship Matrix
GxE	Genotype-by-Environment Interaction

GY	Grain Yield
LRT	Likelihood Ratio Test
MAF	Minor Allele Frequency
ME	Mega-environments
MET	Multi-environment Trials
MM	Multi-environment Main genetic effect model
MMGE	Multi-environment Main genetic effect model plus GxE
MxE GBLUP	Marker-by-environment Genomic BLUP model
PC	Principal Components
PCoA	Principal Coordinate Analysis
PHT	Plant Height
RCBD	Randomized Complete Block Design
REML	Restricted Maximum Likelihood
SCA	Specific Combining Ability
SI	Selection Intensity
SM	Single-environment Main genetic effect model
SNP	Single Nucleotide Polymorphism

TABLE OF CONTENTS

Page

3. ASSESSING COMBINING ABILITIES, GENOMIC DATA, AND GXE INTERACTIONS TO PREDICT HYBRID GRAIN SORGHUM PERFORMANCE ...47

4. MEGA-ENVIRONMENT ANALYSIS TO ASSESS ADAPTABILITY, STABILITY, AND GENOMIC PREDICTIONS IN GRAIN SORGHUM HYBRIDS .96

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Long before plant breeding branched into its “business” aspect (Bernard, 2012), plant breeding was defined as “the art and science” of improving plant genetics to meet human needs (Fehr, 1987; Sleper and Poehlman, 1995). In the early days of scientific crop improvement, public plant breeding programs played a crucial role in developing most of the breeding methodologies that successfully enabled the development of new cultivars and their subsequent hybrids.

Although the ultimate goal of a breeding program is the release of a new product, the process for achieving it is equally important. Public plant breeding programs pioneered the application of quantitative genetics, population genetics, and statistical analysis to select genotypes expressing desirable phenotypic traits. Additionally, these programs elucidated many fundamental concepts involved in crop improvements, including genotype-by-environment interaction effect, heterotic groups, plant genome sequencing techniques, marker assisted selection, and genomic selection. These concepts transformed crop production by developing enhanced methods that increased yield potential more efficiently. In essence, public plant breeding programs significantly impacted the rates of genetic gain and continue to develop new technologies to optimize the breeding pipeline (Cobb et al., 2019).

Public sorghum breeding programs have been responsible for developing much of the elite grain sorghum germplasm worldwide. In the US, the Texas A&M and Kansas State sorghum breeding programs have supported the national sorghum production for over a century. Both programs have released high yielding lines with biotic and abiotic stress resistance and improved quality for the sorghum industry.

1.1. Sorghum improvement

Sorghum [*Sorghum bicolor* (L.) Moench] is the third most-produced grain in the US and the fifth most important grain crop across the globe (FAO, 2019). Although national sorghum production hectarage has been declining, the US continues to be a global producer of the crop with a planted area of 2.30 million hectares, which corresponds to 16% of the total sorghum production worldwide (FAOSTAT, 2019).

Grain sorghum is a grass native to and domesticated in East Africa. Sorghum arrived in the US as “guinea corn” via the West African slave trade, and although widely distributed, eventually disappeared from cultivation. Grain sorghum, as known today, reached the semiarid regions of the Southern Great Plains, where it found a home due to its highly drought-and-heat-tolerant trait. According to Vinall et al. (1936), White and Brown Durra arrived in 1874 from Egypt and became the first grain sorghums to be successfully used in expanding the crop in the country. Afterward, the Red and White Kafirs, Giant Standard Yellow Milo, Blackhull Kafir, Shallu, Pink Kafir, Feterita, and Hegari entered the Great Plains during the late 19th and early 20th centuries. These introductions originated the genetic basis that contributed to the development of the early grain sorghum cultivars in the region (Smith and Frederiksen, 2001).

By the 1880s, farmers already recognized the potential of sorghum as a feed crop (Haney, 1989). As these genotypes were not adapted to the regions, sorghums were easily lodged, often goose-necked, late, and tall (Smith and Frederiksen, 2001; Rooney, 2004). Farmers performed the first selections of genetic variants and randomly outcrossed progeny which led to the fixation of the short-statured, straight peduncle, and early maturing cultivars (Smith and Frederiksen, 2001; Rooney, 2004). However, the development of more adapted cultivars further increased the need for a systematic accumulation of information about sorghum. The establishment of experimental research centers managed such needs.

Selections initially made by farmers benefited sorghum breeders, who started to develop cultivars via hybridization. Desirable characteristics fixed in different plants were combined into distinct cultivars through controlled crosses. Several combine cultivars expressing early maturity were developed and grain sorghum acreage increased. Due to a lack of uniformity, these improved cultivars could not fully express their yield potential. The discovery of cytoplasmic male sterility led to the development of hybrids, which significantly increased grain sorghum yields (Smith and Frederiksen, 2001; Rooney, 2004). Concomitantly, the need for genetic variation enhancement motivated sorghum breeders to develop new strategies to incorporate divergent germplasm to high-performing parents. The introduction of exotic materials to the breeding population positively affected modern parental lines and hybrids agronomic performance (Stephens et al., 1967; Klein et al., 2016). However, the rates of genetic gain in sorghum have been modest, and new methods to leverage yield potential and genetic variation should be developed.

1.2. Sorghum breeding at Texas A&M and Kansas State

Texas A&M and Kansas State sorghum breeding programs have long been important to sorghum crop improvement. It is not a coincidence that most of the national sorghum production is in or adjacent to these states (Smith and Frederiksen, 2001). In 2018, more than 75% of the total US sorghum production was harvested at Kansas and Texas farmlands (FAO, 2019). Both programs began sorghum investigations and breeding in the early twentieth century. They both have origins in the land-grant system established by the Hatch Act, which authorized the establishment of the Agricultural Experimental Stations (AES) in each state (Shelton, 1888; Haney, 1989; Wilson, 1963; Phillips, 2001). The Texas AES was established in 1887 (Haney, 1989), while the Kansas AES was founded in 1901 (Wilson, 1963; Phillips, 2001).

Early sorghum research in College Station (Texas) demonstrated the need for multiple environment testing; additional sorghum research began in Chillicothe, Weslaco, and eventually Lubbock, Amarillo, and Corpus Christi (Haney, 1989). Similar results in Hays (Kansas) led to sorghum research activities in Colby, Garden City, and Manhattan (Phillips, 2001). Both programs initially benefited farmers by developing and diffusing standard agricultural practices to improve the production and processing of grain sorghum. In the most important form, this involved identifying the best cultivars available to each specific production site (Dickson, 1917; Karper, 1917; Jackson 1917; Cunningham and Kenney, 1917; Conner and Karper, 1918; Laude and Swanson 1933; Swanson and Laude 1934).

The rediscovery of Mendel's law impacted crop improvement considerably. Sorghum breeders began to apply the principals of heredity by crossing desirable plants with the expectation to produce selectable progenies. Vinall and Cron, for instance, released Chiltex and Premo as the first sorghum cultivars deliberately developed from crossing Feteritas and Kafirs (Smith and Frederiksen, 2001; Rooney, 2004). Crosses between Milos and Kafirs resulted in short-statured and early-maturing cultivars, such as Beaver and Wheatland (Quinby, 1974). Many other breeding-crosses led to the development of better performing, biotic-stress resistant, and abiotic-stress tolerant cultivars. These cultivars were used as the final product until hybrid production started.

Sorghum hybrids were known to perform better then pure-line cultivars (Stephens and Quinby, 1952), but their production was not economically feasible. Hybrid seeds were obtained from either hand emasculation, or pollen development disruption, using hot water (Stephens and Quinby, 1933). The limited number of seeds led the first hybrids to become a source for population development and genetic studies (Quinby, 1967; Quinby and Karper, 1954). Advanced methods based on male genetic sterility increased the number of seeds, but the inability to produce uniform hybrids terminated their use in commercial production (Rooney, 2004). Hybrid sorghum production became economically feasible only after the identification of cytoplasmic-male-sterility (CMS) at Texas A&M AgriLife Research (Stephens and Holland, 1954). From discovery to development, CMS required approximately ten years, but the effects on the sorghum industry were immediate; hybrids were produced and widely adopted with a concomitant yield increase (Maunder, 1972; Miller and Kebede, 1984).

The CMS facilitated the development of the hybrid sorghum seed industry. Public sector breeding programs began to focus on germplasm development and genetic research. This led to developments of great value to the sorghum industry; the sorghum conversion program (Stephens et al., 1967; Klein et al., 2016); disease and insect resistance (Rooney, 2004); grain and forage quality; and yield enhancement. While many of the sorghum breeding programs were dissolved over time, the programs in Texas and Kansas maintained their identity and programs continuously.

However, the rates of yield gain have been modest since commercial hybrid sorghum production began. A long-term hybrid sorghum selection in either private and public breeding programs provided an increase of 0.005 to 0.015 t $ha^{-1}y^{-1}$ (Manson et al., 2008; Assefa and Staggenborg, 2010; Gizzi and Gambin, 2016; Pfeiffer et al., 2018). Such improvement rates are marginal to keep sorghum as a major crop, and hence new strategies to maximize yield gain must be developed.

The combination of divergent elite inbred lines from distinct sorghum breeding programs can elucidate the relative benefits and limitations regarding the variation and performance existing within and across these programs. Such combinations might maximize yield potential across a broader range of environments by exploiting its dominance variation. It might also indicate how to better explore selectable variance for the future development of superior parents. Additionally, the implementation of genetic prediction models for grain sorghum could aid breeders to appropriately allocate resources to identify outperforming hybrid combinations using genetic information derived from the inbred parents. Thus, this dissertation assessed alternatives to maximize genetic gain via elite germplasm enhancement based on exchanging inbred lines derived from the Texas A&M and Kansas State breeding programs and demonstrate how a collaborative effort between these two programs might benefit sorghum producers. Further, it aims to implement genomic selection into a sorghum breeding pipeline to assist breeders to increase the rates of genetic gain of the crop.

1.3. References

Bernardo, R. 2002. Breeding for quantitative traits in plants. Vol. 1, p. 369. Woodbury: Stemma Press.

Conner, A. B. and Karper, R. E. 1918. Grain Sorghum Improvement. Texas Agriculture Experiment Station. Bulletin 236.

Cunningham, C. C., and Kenney, R. 1917. Growing Sorghum in Kansas. Kansas Agricultural Experiment Station. Bulletin 218.

Dickson, R. E. 1917. Progress report, substation no. 7, Spur, Texas, 1909-1914. Texas Agricultural Experiment Station. Bulletin 218.

Fehr, W.R. 1991. Principles of cultivar development. Vol. 1. Theory and technique. Iowa State Univ. Press, Ames, IA.

Haney, R.L. 1989. Milestones marking ten decades of research. Texas Agricultural Experimental Station. 276p.

Jackson, J. W. 1917 Progress report, substation no. 9, Pecos, Texas, 1910-1914. Texas Agricultural Experiment Station. Bulletin 221.

Karper, R. E. 1917 Progress report, substation no. 8, Lubbock, Texas, 1909-1914. Texas Agricultural Experiment Station. Bulletin 219.

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 Regist. 10(1): 57–61. doi: 10.3198/jpr2015.05.0034crg.

Laude, H. H. and Swanson, A. F. 1933. Sorghum production in Kansas. Kansas Agricultural Experiment Station. Bulletin 265.

Maunder, A.B. 1972. Objectives and approaches to grain and for- age sorghum improvement in the Americas. N.G.P. Rao and L.R. House, Sorghum in seventies. Oxford and IBH Publishing Co., New Delhi. p. 60–100.

Miller, F.R. and Y. Kebede. 1984. Genetic Contributions to Yield Gains in Sorghum, 1950 to 1980. In: W.R. Fehr, editor, Genetic contributions to yield gains of five major crop plants. ASA, CSSA, Madison, WI. p. 1–14.

Pfeiffer, B.K., Pietsch, D., Schnell, R.W. and Rooney, W.L. 2019. Long-Term Selection in Hybrid Sorghum Breeding Programs. Crop Science, 59(1), pp.150-164.

Philips, W. M. 2001. A history of the agricultural research center-hays: the first 100 years. Bulletin 663.

Poehlman, J.M. and Sleper, D.A. 1995. Breeding Field Crops. 4th Iowa State University Press. Ames, USA P, 494.

Quinby, J.R. and Karper, R.E. 1954. Inheritance of Height in Sorghum 1. Agronomy Journal, 46(5), pp.211-216.

Quinby, J.R. 1967. The maturity genes of sorghum. In Advances in agronomy (Vol. 19, pp. 267-305). Academic Press.

Rooney, W.L. 2004. Sorghum improvement: Integrating traditional and new technology to produce improved genotypes. Adv. Agron. 83:37–109. doi:10.1016/S0065-2113(04)83002-5

Rooney, W.L., F.R. Miller, and L.W. Rooney. 2003. Registration of RTx437 sorghum parental line 43:445–446. doi:10.2135/ cropsci2003.4450.

Shalton, E. 1888. Kansas State Agricultural College Bulletin No. 1.

Smith, C.W. and Frederiksen, R.A. 2000. Sorghum: Origin, history, technology, and production (Vol. 2). John Wiley & Sons.

Stephens, J.C. and Holland, R.F. 1954. Cytoplasmic male sterility for hybrid sorghum seed production. Agron. J, 46(1), pp.20-23.

Stephens, J.C., F.R. Miller, and D.T. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Sci. 7:396. doi:10.2135/cropsci19 67.0011183X000700040036x

Swanson, A. F., Laude, H. H. 1934. Varieties of sorghum in Kansas. Kansas Agricultural Experiment Station. Bulletin 266.

Vinall, H. N., Stephens, J. C., and Martin, J. H. 1936. Identification, history, and distribution of common sorghum varieties. US Department of Agriculture, Washington, DC, Bulletin 506.

Wilson, C. P. 1963. A history of the Fort Hays Kansas. Agricultural Experiment Station, Manhattan.

2. COMBINING ABILITIES AND ELITE GERMPLASM ENHANCEMENT ACROSS US PUBLIC SORGHUM BREEDING PROGRAMS

2.1. Synopsis

For mature breeding programs, maintaining genetic variation in elite germplasm requires a continual assessment of the most efficient methods to maximize functional genetic variation while improving productivity. This research assessed the relative value (defined as population means and variances) derived from elite germplasm exchange between distinct public breeding programs. Ten elite A- and R-lines from Texas A&M and Kansas State sorghum breeding programs were crossed in a factorial design to generate 100 hybrids. Hybrid combinations were grouped to represent hybrids within and across programs. Grain yield, plant height, and days to anthesis were measured in ten environments over two years. Combining abilities and their interactions with the environment were assessed. Combined analysis detected significant effects for all traits, but genetic effects for grain yield were not consistently significant within each group of hybrid combinations. Hybrids derived from only Texas inbreds had limited genetic variation for grain yield but the highest mean of all four groups; hybrids derived from only Kansas inbreds produced moderate genetic variation but lower grain yield potential. Maximum genetic variation for grain yield and plant height occurred when Kansas A-lines were crossed to Texas R-lines, whereas hybrids between Texas A-lines and Kansas R-lines maximized variation for days to antheses. Results demonstrated the potential benefit from crossing elite inbred parents derived from distinct breeding programs to increase genetic variation and enhance agronomic performance. Further, these crosses

could foster collaborative efforts between public institutions to enhance the rate of genetic gain in crops with limited public and private resources.

2.2. Introduction

In any plant breeding program, assembling germplasm with genetic variation is essential for long-term genetic improvement of yield and other complex traits in hybrid crops (Dudley and Moll, 1969; Fehr, 1991; Byrne et al., 2018; Bernardo, 2020). While genetic variation is crucial for the long-term success of improvement programs, its presence must be balanced with the need for high performance for agronomic traits such as yield and quality (Rooney, 2004; Smith, 2007). Population improvement approaches have been the standard practice to simultaneously increase genetic variation and agronomic performance (Lynch and Walsh, 1998; Hallauer et al., 2010). Although effective, population improvement involves progeny testing, which may hinder the rates of genetic gain over time (Comstock et al., 1949; Doggett, 1972; Hallauer and Darrah, 1985). Further, population improvement approaches are not readily adaptable to inbred line and hybrid development.

To address these issues, new breeding approaches to balance the mean and variation have been developed. Rapid cycling genomic selection, for instance, addresses some limitations intrinsic to the conventional population improvement scheme (Gaynor et al., 2017; Zhang et al., 2017; Gorjanc et al., 2018). Alternatively, plant breeders often search for exotic germplasm to introgress new sources of variation to the breeding population (Pollak, 2003; Fan et al., 2016). In sorghum [*Sorghum bicolor* (L.) Moench], the conversion program (Stephens et al., 1967) and the reinstated conversion program

(Klein et al., 2016) released many inbred lines adapted to subtropical and temperate environments, and their pedigrees continue to integrate current elite lines and commercial hybrids (Duncan et al., 1991; Pfeiffer et al., 2019; Horne et al., 2020).

Another promising method commonly used in commercial maize breeding programs to increase selectable variation involves exchanging elite germplasm across different sites (Cooper et al., 2014). Rather than centralizing the crosses within a specific program, breeders working in the private sector treat each program as an interconnected network of smaller programs to generate breeding crosses and solve local challenges (Podlich and Cooper, 1998; Cooper et al., 2014; Technow et al., 2020). This method increases genetic gain by introducing divergent yet high-performing inbred lines that allows for efficient selection and accumulation of favorable alleles for critical agronomic traits. Simulation studies demonstrate the benefits of such a method for the long-term genetic improvement of yield and other complex traits in hybrid crops (Podlich and Cooper, 1998; Technow et al., 2020).

Additional results involving exchange of elite germplasm between distinct breeding programs also increases collaborative efforts between the plant breeding communities. Such synergy is particularly advantageous for public institutions, which have downsized significantly over the years (Shelton and Tracy, 2017; Coe et al., 2020). For instance, Adhikari et al. (2020) assessed hybrid combinations of elite wheat lines derived from US public winter wheat breeding programs. The authors highlighted the importance of useful genetic variation within and across programs to make breeding crosses and develop improved lines with the potential to meet local needs. The combination of exchanging elite germplasm across public institutions with the application of new technologies, including modern sequencing techniques and robust statistical models, can enhance the rates of genetic gain in public sector breeding programs (Xu et al., 2017; Cobb et al., 2019; Fonseca et al., 2021).

The introgression of elite germplasm across breeding programs impacts the effective size of breeding populations, causing an increase in the rates of genetic drift, a decrease in genetic variation, and a potential restriction on the response to selection (Yu and Bernardo, 2004; Fu, 2006; Charlesworth, 2009). However, empirical and simulation studies argue that reducing the number of selected individuals to produce breeding populations has a limited effect on genetic variation for yield (Guzman and Lamkey, 2000; Bernardo et al., 2006). Moreover, long-term selection demonstrates that positive response to selection for quantitative traits can continue even in small populations (Dudley and Lambert, 2004; Duvick et al., 2004). Additionally, changing selection environments expose genotypes to diverse environmental effects, implying changes in selection targets, which might reduce the selection pressure on particular alleles or complexes of alleles and, thus, slow or prevent fixation (Technow et al., 2020). Based on this literature, the value of exchanging elite germplasm between breeding programs must be empirically determined and results will depend on the crop and the specific germplasm that is evaluated.

Genetic variation consists of additive and non-additive components (Falconer and Mackay, 1996), and the proportion of additive effects typically account for most of the total genetic variation (Falconer and Mackay, 1996; Lynch and Walsh, 1998; Hill et al., 2008; Fischer et al., 2008). Since additive effects reflect heritable variation, they predict breeding crosses and hybrid performance (Bernardo, 1994; Piepho et al., 2008; Technow et al., 2014; Basnet et al., 2019; Fonseca et al., 2021). In the absence of epistasis, general and specific combining abilities reflect the magnitudes of additive and dominance effects, respectively (Falconer and Mackay, 1996). General combining ability (GCA) is the average performance of an inbred line in hybrid combinations, while the specific combining ability (SCA) describes deviations from those parental contributions with respect to a particular hybrid cross (Sprague and Tatum, 1942). Several mating schemes are frequently applied to generate GCA and SCA estimates (Comstock and Robinson, 1952; Griffing, 1956). Currently, combining ability estimates can be obtained in an array of ways using more flexible models (Möhring et al., 2011; Alves et al., 2019; Fonseca et al., 2021) and modern statistical tools (Pérez and de los Campos, 2014; Bates et al., 2015; Covarrubias-Pazaran, 2016; Onofri et al., 2020).

GCA and SCA are often generated to assist breeders in making breeding decisions from producing breeding crosses and advancing selected hybrid combinations (Fasahat et al., 2016). Typically, crops commercialized as varieties focus on estimating GCA effects (Isleib and Pattee, 2007; Hinze et al., 2011; Zhang et al., 2015; Teodoro et al., 2019; Adhikari et al., 2020), while those commercialized as hybrids also evaluate SCA effects to maximize heterosis (de la Vega and Chapman, 2006; Bagheri and Jelodar, 2010; Fan et al., 2014; Larièpe et al., 2017; Yu et al., 2020). While sorghum is a self-pollinated species, it has been grown as a hybrid crop in the US since the development of the A_1 cytoplasmic male sterility system (Stephens and Holland, 1954).

As a hybrid crop, both GCA and SCA are important in sorghums [(Grain sorghum: Prasad and Biradar, 2017; Crozier et al., 2020; Fonseca et al., 2021), (Biomass sorghum: Sheunda et al., 2019; Oliveira et al., 2019; Wagaw and Tadesse, 2020), (Sweet sorghum: Lombardi et al., 2018; Rocha et al., 2018), and (Forage sorghum: Gorz et al., 1987; Mohammed, 2009; Aruna et al., 2012)]. In countries where the hybrid sorghum seed industry is in its infancy, GCA and SCA estimates can be used to develop heterotic groups which are essential for hybrid sorghum improvement programs (Kenga et al., 2004; Akata et al., 2017; Kante et al., 2019).

Throughout the development of grain sorghum hybrids, the US breeding programs have improved critical agronomic traits for sorghum production. Much of this progress was due to the successful introduction of unadapted elite germplasm through the sorghum conversion program. (Stephens et al., 1967; Laosuwan and Atkins, 1977; Lothrop et al., 1985; Miller and Kebede, 1984; Klein et al., 2016; Horne et al., 2020). While improvement has been consistent over recent decades, the rates of genetic gain in sorghum have been modest (Assefa and Staggenborg, 2010; Gizzi and Gambin, 2016; Pfeiffer et al., 2019). The chance to exchange divergent elite inbred lines across distinct sorghum breeding programs might open an opportunity to maximize the rates of genetic gain and indicate best practices to deploy selectable variation for developing superior parents. Additionally, this could foster collaborative efforts among public institutions to address their local challenges.

The present study aims to assess the relative benefits of developing hybrids using elite lines derived from Texas A&M and Kansas State sorghum breeding programs and evaluate the suitability of exchanging elite germplasm across programs to generate breeding populations. Both of these public programs have an extended history of continual sorghum improvement (Ball and Leidigh, 1908; Cunningham and Kenney, 1918; Laude and Swanson, 1933), and germplasm developed by each program represents elite inbreds adapted to their respective production environments. The specific objectives of this study are to estimate the general and specific combining ability within and across programs using a single-step linear mixed model and also assess the genetic distance between elite lines using molecular markers.

2.3. Material and Methods

2.3.1. Genetic material

Grain sorghum hybrids previously detailed by Fonseca et al. (2021) were used in this research. These 100 hybrids were generated from crossing ten female (A-lines) with ten male (R-lines) in a factorial mating scheme (Comstock and Robinson, 1952). The inbreds selected for this study represent elite germplasm from the sorghum breeding programs at Texas A&M AgriLife Research (College Station, Texas) and Kansas State University, Agricultural Research Center (Hays, Kansas). Each program provided five females and five males that are adapted to their respective target environments and had produced agronomically acceptable hybrids previously (Table 1). The 100 hybrids were subdivided into four groups to represent hybrid combinations derived from elite lines within a breeding programs and also between breeding programs (Figure 1).

Table 1. List of A- and B-lines from Texas A&M (TAM) and Kansas State (KS) breeding programs used in the factorial mating scheme to develop hybrids. The information includes pedigree, seed color, plant color, and publications describing the material, if applicable.

Parents	Pedigree	Seed	Plant	Reference
KS A-lines				
ARCH11051A	ms3tan/ms3/3/TX623B/PI 550610B//01714B	White	Tan	Experimental line
ARCH11129A	KP8BS2#16/OPB/3/Tx623 B/PI550610B//01714B	White	Tan	Experimental line
ARCH11136A	RRSWht-tanB0641701-167B	White	Tan	Experimental line
ARCH11146A	ms3tan/ms3/3/Tx623B/PI 550610B//01714B	White	Tan	Experimental line
KS118A	ms3tan//RTx436/PI550610	White	Tan	Kofoid and Harvey (2005)
KS R-lines				
ARCH11001R	87BH8606-6//ms3/PI 550610/3/SC 414-12ER/PI 550610/4/N310B	White	Tan	Perumal et al. (2019)
ARCH11002R	87BH8606-6//ms3/PI 550610/3/SC 414-12E R/PI 550610/4/N310B	White	Tan	Perumal et al. (2019)
ARCH11028R	2006ISO11(Wht-tan RRS)R	White	Tan	Perumal et al. (2019)
ARCH11055R	9614B/5/ms3tan/BHF14-(B1*Bvar)B/4/ms3//ms3/PI 550610	Red	Tan	Perumal et al. (2019)
ARCH11056R	9614B/5/ms3 tan/BHF14-(B1*Bvar)B/4/ms3//ms3/PI 550610	Red	Tan	Perumal et al. (2019)
TAM A-lines				
A05071	BDLO357/BHF88	White	Pigmented	Experimental line
ATx645	BTx3042//BTx625/BTx642	Red	Pigmented	Rosenow et al., Release Notice
ATx3408	BTx631/08PR047	White	Tan	Mbulwe et al., (2015)
ATx3447	BTx643/BTx635	White	Tan	Rosenow et al., (2021)
A08140	BTx642//B9411/BTx645	Yellow	Tan	Experimental line
TAM R-lines				
R08304	Tx430/00CS765	Red	Pigmented	Experimental line
RTx437	77CS4/RTx430	White	Tan	Rooney et al., (2003)
R07178	91BE7414/R01125	Red	Tan	Experimental line
RTx436	SC120-6-sel/2//Tx7000	White	Tan	Miller et al., (1992)
EON361	R5646/SC326-6	White	Tan	Experimental line

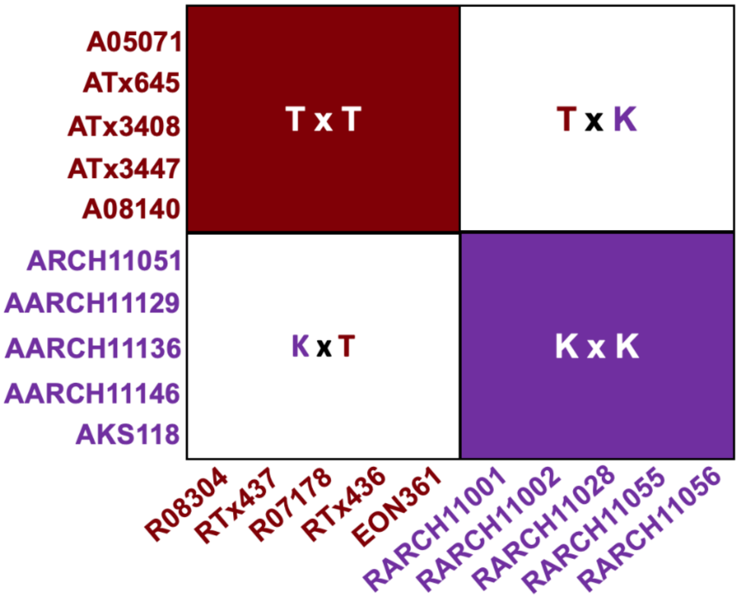

Figure 1. A- and R-lines from Texas (maroon) and Kansas (purple) sorghum breeding programs used in a factorial mating scheme to generate four sets of hybrid combinations. Maroon and purple squares represent hybrid combinations derived from Texas and Kansas programs, respectively. White squares represent hybrids generated across programs.

2.3.2. Experimental data

Hybrids were planted in a randomized complete block design (RCBD) with sets in rep adjustment; each set was composed of one of the four groups of 25 hybrids (Figure 1), with randomization occurring within and between sets. In 2018 trials, each location had three replicates, while in 2019 each location had two replicates. A plot consisted of two adjacent rows, approximately 5.3 m in length, with row spacing that ranged from 0.76 to 1.0 m, depending on the production practices in each environment. In 2018, trials were grown in five environments: Monte Alto (18RF), Victoria (18VC), and College Station (18CS), Texas, and Garden City (18GC) and Colby (18COL), Kansas. In 2019, the trials were in Taft (19TA), Victoria (19VC) and College Station

(19CS), Texas, and Hays (19HAY) and Colby (19COL), Kansas. These Texas and Kansas locations represent distinct adaptation zones and, in each test, agronomic practices standard to the location were followed (Table 2) (Fonseca et al., 2021). GPS coordinates of each environment were used to collect weather data from the NASA POWER database (NASA, 2021).

Table 2. Classification of the environments at which hybrid trials were evaluated. The information includes growing season, average temperature, precipitation, adaptation zone, and GPS coordinates. Designations for environments are as follows: Monte Alto 2018 (18RF), Taft 2019 (19TA), Victoria 2018 (18VC), Victoria 2019 (19VC), College Station 2018 (18CS), College Station 2019 (19CS), Garden City 2018 (18GC), Garden City 2019 (19GC), Colby 2018 (18COL), Colby 2019 (19COL), and Hays 2019 (19HAY).

Location	Temperature (°C)†	Precipitation (mm)‡	Growing season	Adaptation zone	GPS coordinates
18RF	21 - 30	306	Feb - Jun	Subtropical	26°21'06.1"N 97°53'50.3"W
19TA	21 - 30	335	Mar - Jul	Subtropical	28°00'05.4"N 97°15'12.4"W
18VC	20 - 29	373	Mar - Jul	Subtropical	28°47'24.4"N 96°50'22.6"W
19VC	17 - 30	371			
18CS	18 - 32	323	Mar - Jul	Subtropical	30°32'56.6"N 96°26'11.5"W
19CS	13 - 30	534			
18GC	27 - 12	555	Jun - Oct	Temperate	37°59'21.4"N 100°48'52.5"W
19GC	24 - 11	241			
18COL	26 - 10	359	Jun – Oct	Temperate	39°22'56.6"N 101°04'45.0"W
19COL	22 - 9	345			
19HAY	22 - 10	405	Jun – Oct	Temperate	38°51'10.8"N 99°20'24.6"W

† Average soil surface temperature at the beginning and end of the growing season

‡ Cumulative precipitation across growing season

2.3.3. Statistical analysis

For each environment, data were analyzed following a linear mixed model (Henderson, 1984):

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Z_1 h} + \boldsymbol{Z_2 r} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$ is the vector of phenotypes; $\boldsymbol{\mu}$ is an intercept; $\boldsymbol{h}$ is a random effect of hybrid, $\boldsymbol{h} \sim N(\mathbf{0}, \sigma^2_{hyb}\boldsymbol{I})$, $\boldsymbol{r}$ is a random effect of replicates, $\boldsymbol{r} \sim N(\mathbf{0}, \sigma^2_{r}\boldsymbol{I})$; $\boldsymbol{e}$ is the vector of residuals, $\boldsymbol{e} \sim N(\mathbf{0}, \sigma^2_{e}\boldsymbol{I})$; $\mathbf{1}$ is a vector of ones; $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are incidence matrixes; σ^2_{hyb}, σ^2_{r} and σ^2_{e} are variance components for hybrids, replicates and residuals, respectively. Model (1) was extended to incorporate environment and GxE effects for performing a combined analysis. Hybrid effects were partitioned into GCA and SCA effects, environments were defined as a year-location combination, and GxE effects were represented by the interaction between GCA and SCA with the environment as follows:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Z_1 f} + \boldsymbol{Z_2 m} + \boldsymbol{Z_3 h} + \boldsymbol{Z_4 s} + \boldsymbol{Z_5 fs} + \boldsymbol{Z_6 ms} + \boldsymbol{Z_7 fms} + \boldsymbol{Z_8 r(s)} + \boldsymbol{e}, \tag{2}$$

where $\boldsymbol{f}$ is a vector of GCA effects of females, $\boldsymbol{f} \sim N(\mathbf{0}, \sigma^2_{f}\boldsymbol{I})$; $\boldsymbol{m}$ is a vector of GCA effects of male, $\boldsymbol{m} \sim N(\mathbf{0}, \sigma^2_{m}\boldsymbol{I})$; $\boldsymbol{fm}$ is a vector of SCA effects of hybrid combinations, $\boldsymbol{fm} \sim N(\mathbf{0}, \sigma^2_{fm}\boldsymbol{I})$; $\mathbf{s}$ is a vector of environmental effects, $\mathbf{s} \sim N(\mathbf{0}, \sigma^2_{s}\boldsymbol{I})$, $\boldsymbol{fs}$ is a vector of the interaction effect between GCA of female and environment, $\mathbf{fs} \sim N(\mathbf{0}, \sigma^2_{fs}\boldsymbol{I})$; $\boldsymbol{ms}$ is a vector of the interaction effect between GCA of male and environment, $\mathbf{ms} \sim N(\mathbf{0}, \sigma^2_{ms}\boldsymbol{I})$, $\boldsymbol{fms}$ is a vector of the interaction effect between SCA of hybrid combinations and environment, $\boldsymbol{fms} \sim N(\mathbf{0}, \sigma^2_{fms}\boldsymbol{I})$; $\mathbf{r(s)}$ is the vector of replicate effect nested within environment, $\boldsymbol{r(s)} \sim N(\mathbf{0}, \sigma^2_{r(s)}\boldsymbol{I})$; $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Z_3}$, $\boldsymbol{Z_4}$, $\boldsymbol{Z_5}$, $\boldsymbol{Z_6}$, $\boldsymbol{Z_7}$, and $\boldsymbol{Z_8}$ are incidence matrixes for effects. To assess the relative value of crossing elite inbred lines from distinct breeding programs, model 2 was also applied for each set of 25 hybrids

(Figure 1) to generate estimates of GCA and SCA components within and across breeding programs.

Variance components were estimated via restricted maximum likelihood (REML) method (Patterson and Thompson, 1971) using the *lmer* function of the *lme4* R package (Bates et al., 2015), and its significance assessed by the likelihood ratio test (LRT) using the *ranova* function of the *lmerTest* R package (Kuznetsova et al., 2017). To estimate the standard errors of each variance component, the *mmer* function of the *sommer* R package was applied (Covarrubias-Pazaran, 2016). All analyses were conducted in R software (R Core Team 2020). From the variance component estimates, broad-sense heritability at each environment was calculated as $R = \frac{\sigma^2_{hyb}}{\sigma^2_{hyb} + \frac{\sigma^2_e}{r}}$, whereas, for combined environments, $H^2 = \frac{\sigma^{2*}_{hyb}}{\sigma^{2*}_{hyb} + \frac{\sigma^{2*}_{hybs}}{t} + \frac{\sigma^2_e}{tr}}$, where t is the number of environments; *r* is the number of replicates; σ^{2*}_{hyb}: $\sigma^2_f + \sigma^2_m + \sigma^2_{fm}$; and σ^{2*}_{hybs}: $\sigma^2_{fs} + \sigma^2_{ms} + \sigma^2_{fms}$. Additionally, the narrow-sense heritability for female and male was estimated as $h^2_f = \frac{\sigma^2_f}{\sigma^{2*}_{hyb} + \frac{\sigma^{2*}_{hybs}}{t} + \frac{\sigma^2_e}{tr}}$ and $h^2_m = \frac{\sigma^2_m}{\sigma^{2*}_{hyb} + \frac{\sigma^{2*}_{hybs}}{t} + \frac{\sigma^2_e}{tr}}$, respectively. Finally, the coefficient of variation (CV), genotypic coefficient of variation (CVg), and CVg / CV ratio (CVr) was calculated as $CV_e = \frac{\sqrt{\sigma^2_e}}{\bar{x}} \times 100$, $CV_g = \frac{\sqrt{\sigma^2_g}}{\bar{x}} \times 100$, and $CV_r = \frac{CV_g}{CV_e}$, respectively (Resende & Duarte, 2007).

2.3.4. Genetic distances

DNA extraction, genotyping-by-sequencing (GBS), and principal coordinate analysis (PCoA) of the genetic distance were described in Fonseca et al. (2021).

2.4. Results and discussion

2.4.1. Analysis by location

Phenotypic performance across environments varied markedly for all traits (Table 3). The environment averages for GY ranged from 4.31 to 8.34 ton ha^{-1}, 121 to 145 cm for PHT, and 59 to 80 days for DA. While the lowest GY (4.31 ton ha^{-1}) occurred in 18RF, the highest (8.34 ton ha^{-1}) occurred 19COL. According to The National Agricultural Statistics Service, grain sorghum average yields reported in 2020 for Texas and Kansas were 4.23 ton ha^{-1} and 5.71 ton ha^{-1}, respectively (NASS, 2021). This indicates the hybrids presented herein generate yields similar to those of commercial hybrids, thus meeting the expectation of high yielding germplasm.

Genetic variation (σ^2_{hyb}) within each environment was significant for all traits and high H^2 estimates for all three traits confirm the consistency of the experimental data (Table 3). For GY, CVg/CV ratio (CVr) was close to or greater than 1.0 in all environments except 18COL and 19HAY. According to Resende and Duarte (2007), CVr values near 1.0 are considered acceptable in maize experiments. Galli et al. (2020) applied CVr to calculate the experimental accuracy and optimize UAS-based high-throughput phenotyping experiments in grain sorghum trials. Herein, CVr was applied as a measurement of potential response to selection. CVr values for all traits indicated that selection for them would be successful.

Table 3. Phenotypic performance of hybrids developed from Texas A&M and Kansas State elite lines under a factorial II mating design across ten environments. Estimates include genetic variance (σ^2_{hyb}), residual variance (σ^2_{e}), coefficient of variation (CV), genotypic coefficient of variation (CVg), coefficient of variation ratio (CVr), and broad-sense heritability (H^2). Designations for environments are as follows: Monte Alto 2018 (18RF), Victoria 2018 (18VC), College Station 2018 (18CS), Garden City 2018 (18GC), Colby 2018 (18COL), Taft 2019 (19TA), Victoria 2019 (19VC), College Station 2019 (19CS), Hays 2019 (19HAY), and Colby 2019 (19COL).

Trait	Parameter	Environments									
		18RF	18VC	18CS	18GC	18COL	19TA	19VC	19CS	19HAY	19COL
GY	Mean (t/ha)	4.31	5.18	6.26	5.58	5.18	6.65	6.46	7.78	6.51	8.34
	σ_{hyb}^2	0.70*	0.28*	0.66*	0.67*	0.57*	0.47*	0.49*	0.77*	0.47*	1.36*
	σ_e^2	0.86	0.15	0.66	0.95	1.24	0.18	0.29	1.36	1.55	0.84
	CVg (%)	19.43	10.19	12.94	14.7	14.56	10.34	10.83	11.29	10.51	13.97
	CV (%)	21.49	7.35	12.95	17.47	21.51	6.37	8.29	14.97	19.14	10.96
	CVr	0.9	1.39	1	0.84	0.68	1.62	1.31	0.75	0.55	1.27
	H^2	0.71	0.85	0.75	0.68	0.58	0.84	0.77	0.53	0.38	0.76
PHT	Mean (cm)	121	127	126	145	137	138	140	140	122	129
	σ_{hyb}^2	64.6*	91.0*	155.7*	111.7*	94.8*	80.3*	71.2*	137.4*	89.3*	78.1*
	σ_e^2	26.31	19.33	48.76	27.15	25.32	30.14	14.32	53.02	28.05	67.18
	CVg (%)	6.6	7.48	9.83	7.27	7.1	6.46	6.02	8.37	7.73	6.81
	CV (%)	4.22	3.45	5.5	3.58	3.67	3.96	2.7	5.2	4.33	6.32
	CVr	1.57	2.17	1.79	2.03	1.93	1.63	2.23	1.61	1.78	1.08
	H^2	0.88	0.93	0.91	0.93	0.92	0.84	0.91	0.84	0.86	0.70
DA	Mean (day)	67.9	67.2	78.7	70.6	72.4	69.6	59.5	80.4	64.2	69.1
	σ_{hyb}^2	11.1*	4.7*	10.9*	-	-	1.6*	3.5*	4.2*	-	-
	σ_e^2	4.77	0.99	4.36	-	-	0.54	0.78	6.88	-	-
	CVg (%)	4.9	3.24	4.2	-	-	1.85	3.14	2.54	-	-
	CV (%)	3.22	1.48	2.65	-	-	1.06	1.48	3.26	-	-
	CVr	1.52	2.19	1.58	-	-	1.75	2.11	0.78	-	-
	H^2	0.87	0.93	0.88	-	-	0.86	0.90	0.55	-	-

* Significant at 0.05 probability level

2.4.2. Variance components for combined analysis

In the combined analysis, the likelihood ratio test (LRT) indicated that all genetic effects and their interactions with the environment were significant for GY (Table 4), PHT (Table 5), and DA (Table 6). The proportion of the total variation explained by genetic components was highest for PHT (42.47%), followed by DA (11.94%) and GY (10.49%). For DA and GY, the environment effect accounted for the majority of the total variation, and given the wide range of environments in this study, this result was expected. The combined analysis also had high repeatability and low CV for all traits (Tables 4, 5, and 6).

In the combined data set, the GCA of males explained the majority of the total genetic variation for all traits, while the SCA component accounted for the smallest portion (Tables 4, 5, and 6). Although the SCA contribution for the expression of traits is proportionally small, it is significant, and is an important explanation of the value of hybrid grain sorghum. Relative to the males, the smaller contribution of females to each trait likely reflects the narrower genetic variance of elite A-lines necessitated by specific traits of importance in seed parents (i.e., sterility, seed yield) and the additional time required for their development (Menz et al., 2004; Crozier et al., 2020). Specifically, after developing a new B-line, the line has to be male-sterilized by the introgression of sterility-inducing cytoplasm via backcrossing. This process of developing A- and B-line pairs increases the time to develop a new females and therefore hinders the deployment of genetic variation from this heterotic pool (Rooney, 2004).

Table 4. Variance components, mean, coefficient of variation (CV), broad-sense heritability (H^2), and female (h^2_f) and male (h^2_m) narrow sense heritability estimates from Texas A&M and Kansas State elite lines under a factorial II mating design for grain yield (GY). Standard error (SE) and the total variation in percentage for each variance component is shown. Estimates are presented for the analysis of whole data set (combined), for a set of Texas × Texas hybrids (T×T), Texas × Kansas hybrids (T×K), Kansas × Texas hybrids (K×T), and Kansas × Kansas hybrids (K×K). Numbers within parenthesis report the amount of data points used in each analysis. Values underlined represent the sum of correspondent effects.

Variance Components	Combined (2458)			T×T (623)			T×K (615)			K×T (617)			K×K (603)		
	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%
Hybrid	0.315		10.49	0.053		1.91	0.121		3.84	0.210		7.69	0.077		3.04
GCA$_f$	0.128^{***}	0.07	4.27	0.001	0.02	0.03	0.007	0.02	0.21	0.090	0.09	3.30	0	0.03	0.00
GCA$_m$	0.150^{***}	0.08	4.99	0.034	0.05	1.21	0.085^{*}	0.08	2.70	0.039	0.05	1.43	0.026	0.05	1.01
SCA	0.037^{***}	0.01	1.23	0.019	0.02	0.67	0.029	0.02	0.93	0.082^{***}	0.04	2.95	0.051^{**}	0.03	2.03
Environment	1.430^{**}	0.71	47.54	1.564^{***}	0.78	55.99	1.821^{***}	0.96	61.26	1.203^{***}	0.67	43.99	0.985^{***}	0.53	42.05
Hybrid × Env	0.382		12.73	0.349		14.10	0.372		11.85	0.280		10.53	0.421		16.61
GCA$_f$ × Env	0.175^{***}	0.03	5.83	0.122^{***}	0.05	4.37	0.131^{***}	0.05	4.18	0.169^{***}	0.05	6.10	0.216^{***}	0.07	8.54
GCA$_m$ × Env	0.157^{***}	0.03	5.23	0.165^{***}	0.05	5.90	0.154^{*}	0.05	4.90	0.112^{***}	0.04	4.03	0.204^{***}	0.07	8.06
SCA × Env	0.050^{*}	0.02	1.67	0.107^{*}	0.05	3.83	0.087^{***}	0.04	2.77	0	0.04	0.39	0	0.05	0.00
Rep(Env)	0.082^{***}	0.03	2.74	0.035^{*}	0.03	1.27	0.064	0.04	3.11	0.369^{***}	0.15	13.49	0.079^{***}	0.05	4.03
Residual	0.796	0.03	26.51	0.746	0.06	26.72	0.627	0.05	19.94	0.673	0.05	24.31	0.868	0.07	34.27
Mean (t/ha)	6.220			6.797			6.281			6.237			5.540		
CV (%)	14.4			12.7			12.7			13.1			16.8		
H^2	0.829			0.454			0.676			0.807			0.522		
h^2_f	0.338			0.006			0.030			0.346			0.000		
h^2_m	0.394			0.288			0.476			0.149			0.174		

*, **, *** Significant at 0.05, 0.01, 0.001 probability level, respectively

Table 5. Variance components, mean, coefficient of variation (CV), broad-sense heritability (H^2), and female (h^2_f) and male (h^2_m) narrow sense heritability estimates from Texas A&M and Kansas State breeding programs under a factorial II mating design for plant height (PHT). Standard error (SE) and the total variation in percentage for each variance component is shown. Estimates are presented for the analysis of whole data set (combined), for a set of Texas × Texas hybrids (T×T), Texas × Kansas hybrids (T×K), Kansas × Texas hybrids (K×T), and Kansas × Kansas hybrids (K×K). Numbers within parenthesis report the amount of data points used in each analysis. Values underlined represent the sum of correspondent effects.

Variance Components	Combined (2481)			TxT (625)			TxK (625)			KxT (624)			KxK (607)		
	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%
(Hybrid)	89.6		42.5	63.9		35.7	58.7		33.6	103.7		43.6	99.5		45.1
GCA$_f$	34.2^{***}	16.8	16.2	34.9^{***}	25.9	19.5	30.1^{***}	22.9	17.3	38.3^{***}	28.9	16.1	53.7^{***}	39.8	24.3
GCA$_m$	47.1^{***}	23.1	22.3	22.6^{***}	17.7	12.6	24.5^{***}	18.6	14.1	58.3^{***}	43.5	24.5	42.2^{***}	31.4	19.1
SCA	8.3^{***}	1.6	3.9	6.5^{***}	2.7	3.6	4.1^{***}	1.9	2.3	7.1^{***}	3.1	3.0	3.5^{***}	1.8	1.60
Environment	67.9^{***}	33.1	31.9	65.9^{***}	33.8	36.9	69.9^{***}	35.0	40.3	66.5^{***}	35.3	27.9	62.9^{***}	32.6	28.5
(Hybrid × Env)	18.3		8.7	11.8		6.6	9.2		5.3	26.1		10.9	30.6		13.9
GCA$_f$ × Env	5.2^{***}	1.1	2.5	2.8^{***}	1.3	1.6	2.4^{**}	1.2	1.4	9.1^{***}	2.9	3.8	15.9^{***}	4.4	7.2
GCA$_m$ × Env	9.3^{***}	1.7	4.4	9.0^{***}	2.7	5.1	6.8^{***}	2.2	3.9	15.8^{***}	4.5	6.7	12.1^{***}	3.5	5.5
SCA × Env	3.7^{***}	0.9	1.7	0	1.7	0	0	1.7	0	1.1	2.0	0.4	2.5	1.8	1.1
Rep(Env)	3.3^{***}	1.3	1.6	6.7^{***}	3.1	3.7	4.6^{***}	2.3	2.6	6.7^{***}	3.2	2.8	0.2	0.5	0.1
Residual	32.5	1.2	15.4	30.6	2.3	17.1	31.6	2.3	18.0	34.8	2.6	14.6	27.6	2.1	12.5
Mean (cm)	132.9			136.6			128.2			136.9			129.8		
CV (%)	4.3			4.0			4.4			4.3			4.0		
H^2	0.969			0.967			0.968			0.965			0.962		
h^2_f	0.370			0.529			0.497			0.357			0.519		
h^2_m	0.510			0.342			0.405			0.543			0.408		

*, **, *** Significant at 0.05, 0.01, 0.001 probability level, respectively

Table 6. Variance components, mean, coefficient of variation (CV), broad-sense heritability (H^2), and female (h^2_f) and male (h^2_m) narrow sense heritability estimates from Texas A&M and Kansas State breeding programs under a factorial II mating design for days to anthesis (DA). Standard error (SE) and the total variation in percentage for each variance component is shown. Estimates are presented for the analysis of whole data set (combined), for a set of Texas × Texas hybrids (T×T), Texas × Kansas hybrids (T×K), Kansas × Texas hybrids (K×T), and Kansas × Kansas hybrids (K×K). Numbers within parenthesis report the amount of data points used in each analysis. Values underlined represent the sum of correspondent effects.

Variance Components	Combined (1882)			TxT (474)			TxK (474)			KxT (473)			KxK (461)		
	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%
Hybrid	6.0		11.9	4.9		10.2	8.2		16.2	4.9		10.6	8.1		14.5
GCA_f	2.5^{***}	1.3	5.0	2.9^{***}	2.1	5.9	2.6^{***}	1.9	5.2	2.7^{***}	2.0	5.7	2.5^{***}	2.0	4.5
GCA_m	3.2^{***}	1.6	6.4	2.0^{***}	1.5	4.1	5.3^{***}	3.9	10.5	2.1^{***}	1.6	4.4	4.9^{***}	3.7	8.8
SCA	0.3^{***}	0.1	0.5	0.03	0.08	0.1	0.9^{**}	0.8	0.6	0.2^{*}	0.1	0.5	0.7^{***}	0.3	1.2
Environment	37.8^{***}	18.1	75.0	38.6^{***}	18.5	79.2	36.7^{***}	17.5	72.1	36.8^{***}	17.7	77.9	40.3^{***}	19.5	72.9
Hybrid × Env	2.9		5.6	1.9		3.9	2.0		3.9	2.4		5.1	3.8		6.9
GCA_f × Env	1.3^{***}	0.2	2.6	0.7^{***}	0.3	1.4	0.8^{***}	0.3	1.5	1.1^{***}	0.3	2.2	1.8^{***}	0.5	3.3
GCA_m × Env	1.3^{***}	0.2	2.5	1.0^{***}	0.3	2.1	1.2^{***}	0.4	2.4	1.2^{***}	0.4	2.6	1.7^{***}	0.5	3.1
SCA × Env	0.2^{*}	0.1	0.4	0.2	0.2	0.4	0	0.3	0	0.1	0.2	0.3	0.3	0.3	0.6
Rep(Env)	0.6^{***}	0.3	1.1	0.2	0.2	0.4	0.3^{**}	0.2	0.6	0.1	0.1	0.3	0.3^{**}	0.2	0.6
Residual	3.2	0.1	6.3	3.0	0.3	6.2	3.6	0.3	7.0	2.9	0.3	6.1	2.8	0.3	5.0
Mean (Days)	69.94			69.5			69.4			70.5			70.2		
CV (%)	2.5			2.5			2.7			2.4			2.3		
H^2	0.939			0.944			0.962			0.936			0.944		
h^2_f	0.397			0.554			0.306			0. 502			0.289		
h^2_m	0.502			0.383			0.622			0.394			0.574		

*, **, *** Significant at 0.05, 0.01, 0.001 probability level, respectively

Another important aspect of the combined analysis is the proportion of the total variation explained by the G×E effect for GY (Table 4). This demonstrates the potential of developing specific high-yielding hybrid combinations for target environments (Fonseca et al., 2021). Also of interest, the magnitude of female GCA×environment interaction effect is greater than the male GCA×environment interaction, which implies that female parents are either more adapted to specific production regions or that male parents are more stable across all environments (or a combination of these two hypothesis). Regardless, significant genetic variation exists across the two breeding programs, which can be explored by crossing divergent elite lines adapted to different production environments. Further studies should address the adaptability and stability of these hybrids and assess specific target environments to further explore the G×E effect.

2.4.3. Variance components within and across programs

While most of the genetic components and their interactions with the environment for PHT (Table 5) and DA (Table 6) were significant, results varied for GY (Table 4) depending on the set of hybrid combinations. For PHT (Table 5) and DA (Table 6), the SCA×environment interaction was not significant. Additionally, the SCA effect for DA was not significant for the TxT set of hybrids (Table 6). This implies an absence of dominance effects that control flowering time within the TxT pool and indicate that phenotyping the inbred lines should be sufficient to predict flowering time of hybrids derived therefrom. Different from the combined analysis, the GCA of females explained most of the variation for PHT, except for the KxT group (Table 5), and most of the variation for DA in TxT and KxT groups (Table 6). The data indicate that the

GCA and SCA variation present is sufficient to allow sorghum breeders to select hybrids with acceptable height and maturity for both target regions (Tables 5 and 6).

The variance component analysis for GY within and across sets of hybrid combinations demonstrates the challenge of finding useful genetic variation to improve highly complex traits in elite germplasm (Table 4). Only the GCA of males in T×K and SCA in K×T and K×K hybrid combinations were significant but most of the G×E interactions were significant. Among the four groups, the hybrid combination that maximized genetic variation for GY was K×T hybrids (Table 4). However, the highest yielding group was the T×T hybrids (Table 4). These results confirm the continual tradeoff that occurs in breeding programs; higher yields cause a concomitant reduction in genetic variation. While the SCA effect was significant only in the K×K group, the group average yield was the lowest of the four groups. The results indicate that dominance effects are very important in this group, and the Kansas program could benefit from the introduction of elite parental seed stock from other programs.

2.4.4. General and specific combining abilities for combined, within and across programs

Estimates of GCA for each genotype varied substantially within and across programs ranging from -0.629 to 0.614 ton ha^{-1} (Table 7). Although GCA values were narrower across sets of hybrid combinations, most of the lines maintained their positive or negative contribution to GY when compared to the combined set. The only exception was the Kansas line RARCH11001, which performed better in the TxK group (Table 7).

Table 7. General combining ability (GCA) estimates of sorghum inbred lines derived from Texas A&M and Kansas State breeding programs under a factorial II mating design for grain yield (GY). GCA estimates are presented for the analysis of whole data set (combined), for a set of Texas × Texas hybrids (T×T), Texas × Kansas hybrids (T×K), Kansas × Texas hybrids (K×T), and Kansas × Kansas hybrids (K×K).

Parents		Combined	TxT	TxK	KxT	KxK
Female	A05071	0.250	0.002	-0.019	-	-
	A08140	0.355	0.004	0.019	-	-
	ATx3447	0.223	-0.002	-0.006	-	-
	ATx3408	0.069	-0.008	-0.045	-	-
	ATx645	0.431	0.004	0.050	-	-
	AARCH11051	-0.244	-	-	0.076	0.000
	AARCH11129	-0.629	-	-	-0.432	0.000
	AARCH11136	-0.191	-	-	0.057	0.000
	AARCH11146	-0.058	-	-	0.229	0.000
	AKS118	-0.206	-	-	0.070	0.000
Male	EON361	0.258	0.030	-	-0.042	-
	R07178	0.292	0.013	-	0.015	-
	R08304	0.614	0.190	-	0.239	-
	RTx436	-0.002	-0.184	-	-0.149	-
	RTx437	0.168	-0.049	-	-0.062	-
	RARCH11001	0.109	-	0.358	-	0.156
	RARCH11002	-0.179	-	0.081	-	0.035
	RARCH11028	-0.556	-	-0.293	-	-0.112
	RARCH11055	-0.285	-	0.038	-	-0.038
	RARCH11056	-0.419	-	-0.184	-	-0.041

Blue and red colors identify positive and negative CGA values, respectively. Shades indicate the magnitude of values.

The highest SCA estimates occurred when crossing TxK lines. The hybrid AARCH11051/R08304 (KxT) had an SCA of 0.403 ton ha^{-1} and was the top performer in the trial (Table 8). Therefore, the development of hybrid combinations using elite inbred lines from distinct breeding programs has the potential to increase the yield of grain sorghum. Moreover, as mentioned previously, the chance to incorporate elite germplasm from different production regions might open an opportunity to develop breeding populations where useful variation can assist breeders to improve the rates of genetic gain of breeding programs.

Table 8. Specific combining ability (SCA) estimates of sorghum inbred lines derived from Texas A&M and Kansas State breeding programs under a factorial II mating design for grain yield (GY). SCA estimates are presented for the analysis of whole data set.

SCA combined		Texas Males				
		EON361	R07178	R08304	RTx436	RTx437
Texas Females	A05071	0.083	-0.142	0.021	0.156	-0.027
	A08140	0.047	0.130	-0.134	0.002	-0.059
	ATx3447	-0.009	-0.108	0.141	0.028	-0.115
	ATx3408	-0.088	0.004	0.052	-0.175	0.105
	ATx645	0.074	0.047	-0.158	-0.151	0.058
Kansas Females	AARCH11051	-0.071	-0.110	0.403	-0.134	0.221
	AARCH11129	0.102	-0.017	-0.159	-0.193	-0.104
	AARCH11136	0.004	0.156	-0.027	0.101	-0.101
	AARCH11146	0.081	0.073	-0.067	0.153	0.101
	AKS118	-0.159	0.040	0.081	0.212	-0.037

		Kansas Males				
		RARCH11001	RARCH11002	RARCH11028	RARCH11055	RARCH11056
Texas Females	A05071	0.018	-0.124	-0.198	0.164	0.121
	A08140	-0.064	0.050	0.007	0.114	0.011
	ATx3447	0.109	0.053	0.106	-0.140	0.000
	ATx3408	0.179	0.175	-0.002	-0.043	-0.188
	ATx645	0.060	0.011	0.014	0.195	-0.025
Kansas Females	AARCH11051	-0.077	-0.044	-0.102	-0.157	0.000
	AARCH11129	-0.048	-0.078	0.370	-0.216	0.161
	AARCH11136	-0.128	-0.056	0.063	0.072	-0.139
	AARCH11146	-0.085	-0.054	-0.087	-0.113	-0.018
	AKS118	0.063	0.023	-0.309	0.052	-0.026

Blue and red colors identify positive and negative SCA values, respectively. Shades indicate the magnitude of values.

2.4.5. Genetic distance between inbred parents

The genetic distances of Texas and Kansas inbred lines confirm to some extent the results reported for the agronomic traits. The Principal Coordinates Analysis (PCoA) explained almost 42% of the total genetic variation among inbred parents in this study (Figure 2). For the most part, the A-line parents were a distinct group, although some Kansas R-lines clustered within this group due to a B-line lineage in their pedigree. Additionally, one Kansas breeding program female clustered within the male group, but there is no evidence of any pollinator lineage within its pedigree. As in the combined analysis, there was greater genetic variation in the R-lines (Tables 4, 5, and 6); these

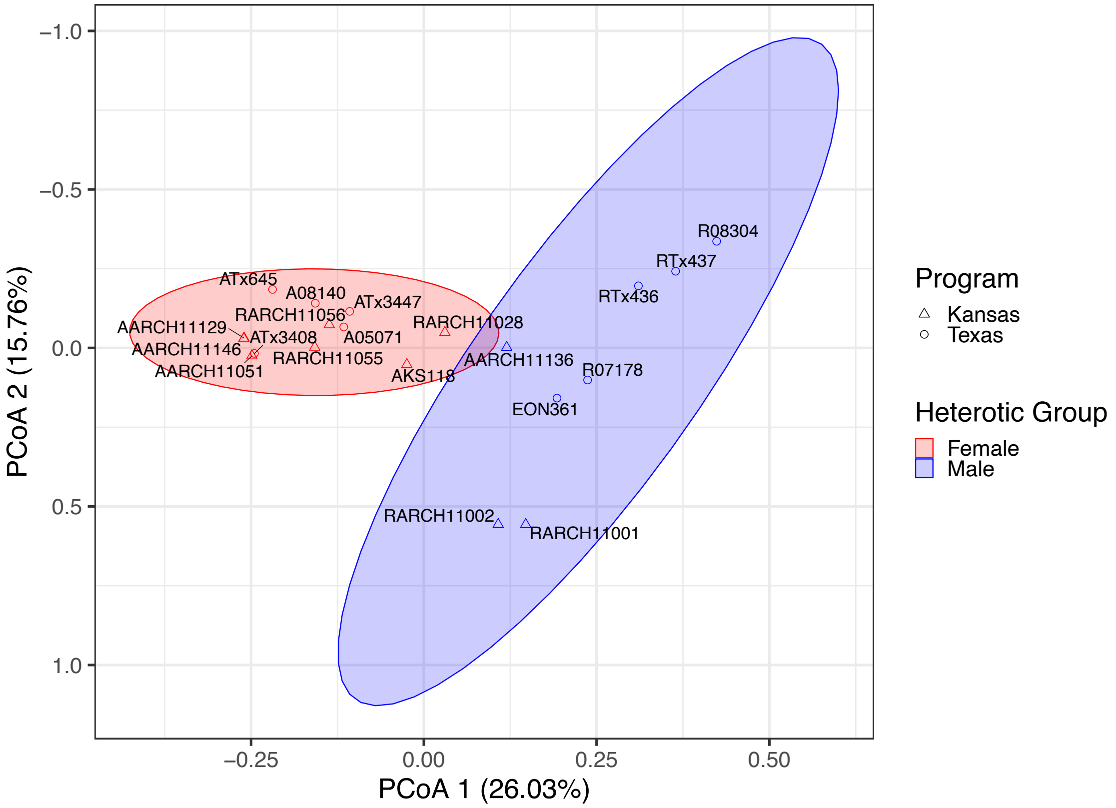

Figure 2. Bidimensional representation of the genetic distance of Texas and Kansas inbred parents used in the factorial mating scheme. Principal Coordinate Analysis (PCoA) displays the first and the second axis with its correspondent % of genetic variance explained by each Principal Coordinate. Clusters represent each heterotic group based on SNP markers.

findings are consistent with previous studies (Menz et al., 2004; Crozier et al., 2020). However, the genetic distance within Texas R-lines was greater than within Kansas R-lines (Figure 2); this contradicts the genetic variation within programs for PHT and DA. One possible explanation for this observation is selection pressure. Within each program, effective selection leads to fixation of loci with favorable effects. Given the different target regions, each program may be selecting different alleles. As loci are fixed, epistatic effects may transition to additive effects and these differences might be difficult to find based on phenotypic analysis (Bernardo, 2020; Technow et al., 2020). The application of genomic-enabled predictions may overcome such limitations since selection is based on genomic sequence data (Fonseca et al., 2021).

2.5. Conclusions

Maintaining selectable genetic variation remains one of the most challenging factors for the established plant breeding programs. In sorghum, hybrid development significantly increases yield and the improvement of each heterotic group is further contributing to develop superior parents and hybrids. While the introduction of exotic but elite germplasm into the US sorghum breeding programs has successfully increased or at least maintained genetic variation within elite pools, new strategies to maintain variation and increase yield potential remain important. The exchange of elite germplasm could become a modern practice for sorghum breeding programs to maximize variation and improve agronomic performance of sorghum grain hybrids. Additional studies that include greater numbers of inbred lines and hybrid combinations from a more inclusive list of public breeding programs are needed to confirm and enhance the benefits presented herein.

2.6. References

Adhikari, A., A.M.H. Ibrahim, J.C. Rudd, P.S. Baenziger, and J.B. Sarazin. 2020. Estimation of heterosis and combining abilities of U.S. winter wheat germplasm for hybrid development in Texas. Crop Sci. 60(2): 788–803. doi: 10.1002/csc2.20020.

Akata, E.A., C. Diatta, J.M. Faye, A. Diop, F. Maina, et al. 2017. Combining ability and heterotic pattern in West African sorghum landraces. African Crop Sci. J. 25(4):

491. doi: 10.4314/acsj.v25i4.7.

Alves, F.C., Í.S.C. Granato, G. Galli, D.H. Lyra, R. Fritsche-Neto, et al. 2019. Bayesian analysis and prediction of hybrid performance. Plant Methods 15(1): 14. doi: 10.1186/s13007-019-0388-x.

Aruna, C., P.K. Shrotria, S.K. Pahuja, A. V. Umakanth, B.V. Bhat, et al. 2012. Fodder yield and quality in forage sorghum: Scope for improvement through diverse male sterile cytoplasms. Crop Pasture Sci. 63(11–12): 1114–1123. doi: 10.1071/CP12215.

Assefa, Y., and S.A. Staggenborg. 2010. Grain sorghum yield with hybrid advancement and changes in agronomic practices from 1957 through 2008. Agron. J. 102(2): 703–706. doi: 10.2134/agronj2009.0314.

Bagheri, N., and N.B. Jelodar. 2010. Heterosis and Combining Ability Analysis for Yield and Related-Yield Traits in Hybrid Rice. Int. J. Biol. 2(2): 222–231. doi: 10.5539/ijb.v2n2p222.

Ball, C.R., and A.H. Leidigh. 1908. Milo as a dry-land grain crop. https://naldc.nal.usda.gov/download/ORC00000188/PDF (accessed 24 October 2019).

Basnet, B.R., J. Crossa, S. Dreisigacker, P. Pérez-Rodríguez, Y. Manes, et al. 2019. Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. Plant Genome 12(1): 1–13. doi: 10.3835/plantgenome2018.07.0051.

Bates, D., M. Mächler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1). doi: 10.18637/jss.v067.i01.

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. 34(1): 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x.

Bernardo, R.N. 2020. Breeding for quantitative traits in plants. Third Edit. Stemma Press.

Bernardo, R., L. Moreau, and A. Charcosset. 2006. Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection. Crop Sci. 46(5): 1972–1980. doi: 10.2135/cropsci2006.01-0057.

Byrne, P.F., G.M. Volk, C. Gardner, M.A. Gore, P.W. Simon, et al. 2018. Sustaining the Future of Plant Breeding: The Critical Role of the USDA-ARS National Plant Germplasm System. Crop Sci. 58(2): 451–468. doi: 10.2135/cropsci2017.05.0303.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10(3): 195–205. doi: 10.1038/nrg2526.

Cobb, J.N., R.U. Juma, P.S. Biswas, J.D. Arbelaez, J. Rutkoski, et al. 2019. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theor. Appl. Genet. 132(3): 627–645. doi: 10.1007/s00122-019-03317-0.

Coe, M.T., K.M. Evans, K. Gasic, and D. Main. 2020. Plant breeding capacity in U.S. public institutions. Crop Sci. 60(5): 2373–2385. doi: 10.1002/csc2.20227.

Comstock, R.E., and H.F. Robinson. 1952. Estimation of average dominance of genes. Estim. Aver. Domin. genes.

Comstock, R.E., H.F. Robinson, and P.H. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability 1. Agron. J. 41(8): 360–367.

Cooper, M., C.D. Messina, D. Podlich, L.R. Totir, A. Baumgarten, et al. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65(4): 311–336. doi: 10.1071/CP14007.

Covarrubias-Pazaran, G. 2016. Genome-Assisted prediction of quantitative traits using the r package sommer. PLoS One 11(6). doi: 10.1371/journal.pone.0156744.

Crozier, D., L. Hoffmann, P.E. Klein, R.R. Klein, and W.L. Rooney. 2020. Predicting heterosis in grain sorghum hybrids using sequence-based genetic similarity estimates. J. Crop Improv. doi: 10.1080/15427528.2020.1748152.

Cunningham, C.C., and R. Kenney. 1918. Growing sorghum in Kansas. https://www.ksre.k-state.edu/historicpublications/pubs/SB218.pdf (accessed 10 December 2019).

Doggett, H. 1972. Recurrent Selection in Sorghum Populations. Heredity (Edinb). 28(1): 9–29. doi: 10.1038/hdy.1972.2.

Dudley, J.W., and R.J. Lambert. 2010. 100 Generations of Selection for Oil and Protein in Corn. Plant Breeding Reviews. John Wiley & Sons, Inc., Oxford, UK. p. 79–110

Dudley, J.W., and R.H. Moll. 1969. Interpretation and Use of Estimates of Heritability and Genetic Variances in Plant Breeding 1 . Crop Sci. 9(3): 257–262. doi: 10.2135/cropsci1969.0011183x000900030001x.

Duncan, R.R., P.J. Bramel-Cox, and F.R. Miller. 1991. Contributions of introduced sorghum germplasm to hybrid development in the USA. Use plant Introd. Cultiv. Dev. Part 1 17: 69–102.

Duvick, D.N., J.S.C. Smith, and M. Cooper. 2010. Long-Term Selection in a Commercial Hybrid Maize Breeding Program. Plant Breeding Reviews. John Wiley & Sons, Inc., Oxford, UK. p. 109–151.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Essex, England, 464p.

Fan, X.M., X.F. Yin, Y.D. Zhang, Y.Q. Bi, L. Liu, et al. 2016. Combining ability estimation for grain yield of maize exotic germplasm using testers from three heterotic groups. Crop Sci. 56(5): 2527–2535. doi: 10.2135/cropsci2016.01.0032.

Fan, X.M., Y.D. Zhang, W.H. Yao, Y.Q. Bi, L. Liu, et al. 2014. Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. Crop Sci. 54(1): 89–97. doi: 10.2135/cropsci2013.06.0393.

Fasahat, P., A. Rajabi, J.M. Rad, and J. Derera. 2016. Principles and utilization of combining ability in plant breeding. Biometrics Biostat. Int. J. 4(1): 1–24.

Fehr, W.R. 1991. Principles of Cultivar Development, vol. 1, Theory and Technique.

Fischer, S., J. Möhring, C.C. Schön, H.-P. Piepho, D. Klein, et al. 2008. Trends in genetic variance components during 30 years of hybrid maize breeding at the University of Hohenheim. Plant Breed. 127(5): 446–451. doi: 10.1111/j.1439-0523.2007.01475.x.

Fonseca, J.M.O., Klein, P.E., Crossa, J., Pacheco, A., Pérez-Rodríguez, P., Perumal, R., Klein, R.R., and Rooney, W.L. 2021. Assessing combining abilities, genomic data, and GxE interactions to predict hybrid grain sorghum performance. The Plant Genome (Submitted).

Fu, Y.-B. 2006. Impact of plant breeding on genetic diversity of agricultural crops: searching for molecular evidence. Plant Genet. Resour. 4(1): 71–78. doi: 10.1079/PGR2006116.

Gaynor, R.C., G. Gorjanc, A.R. Bentley, E.S. Ober, P. Howell, et al. 2017. A two-part strategy for using genomic selection to develop inbred lines. Crop Sci. 57(5): 2372–2386. doi: 10.2135/cropsci2016.09.0742.

Gizzi, G., and B.L. Gambin. 2016. Eco-physiological changes in sorghum hybrids released in Argentina over the last 30 years. F. Crop. Res. 188: 41–49. doi: 10.1016/j.fcr.2016.01.010.

Gorjanc, G., R.C. Gaynor, and J.M. Hickey. 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor. Appl. Genet. 131(9): 1953–1966. doi: 10.1007/s00122-018-3125-3.

Gorz, H.J., F.A. Haskins, J.F. Pedersen, and W.M. Ross. 1987. Combining Ability Effects for Mineral Elements in Forage Sorghum Hybrids 1 . Crop Sci. 27(2): 216–219. doi: 10.2135/cropsci1987.0011183x002700020017x.

Griffing, B. 1956. Concept of General and Specific Combining Ability in Relation to Diallel Crossing Systems. Aust. J. Biol. Sci. 9(4): 463. doi: 10.1071/bi9560463.

Guzman, P.S., and K.R. Lamkey. 2000. Effective population size and genetic variability in the BS11 maize population. Crop Sci. 40(2): 338–346. doi: 10.2135/cropsci2000.402338x.

Hallauer, A.R., M.J. Carena, and J.B. de Miranda Filho. 2010. Quantitative genetics in maize breeding. Springer Science & Business Media.

Hallauer, A.R., and L.L. Darrah. 1985. Compendium of recurrent selection methods and their application. CRC. Crit. Rev. Plant Sci. 3(1): 1–33. doi: 10.1080/07352688509382202.

Henderson, C.R. 1984. Applications of linear models in animal breeding.

Hill, W.G., M.E. Goddard, and P.M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4(2): 1000008. doi: 10.1371/journal.pgen.1000008.

Hinze, L.L., B.T. Campbell, and R.J. Kohel. 2011. Performance and combining ability in cotton (*Gossypium hirsutum* L.) populations with diverse parents. Euphytica 181(1): 115–125. doi: 10.1007/s10681-011-0442-x.

Horne, D.W., N.Y. Patil, R.R. Klein, F.R. Miller, L. Hoffmann, et al. 2020. Registration of 11 diverse sorghum germplasm lines for grain and silage hybrid production. J. Plant Regist. 14(2): 179–188. doi: 10.1002/plr2.20001.

Isleib, T.G., and H.E. Pattee. 2007. A Note on Combining Ability for Sensory Quality of Peanut. Peanut Sci. 34(2): 122–125. doi: 10.3146/0095-3679(2007)34[122:anocaf]2.0.co;2.

Kante, M., F. Rattunde, B. Nébié, I. Sissoko, B. Diallo, et al. 2019. Sorghum hybrids for low-input farming systems in west africa: Quantitative genetic parameters to guide hybrid breeding. Crop Sci. 59(6): 2544–2561. doi: 10.2135/cropsci2019.03.0172.

Kenga, R., S.O. Alabi, and S.C. Gupta. 2004. Combining ability studies in tropical sorghum (Sorghum bicolor (L.) Moench). F. Crop. Res. 88(2–3): 251–260. doi: 10.1016/j.fcr.2004.01.002.

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 Regist. 10(1): 57–61. doi: 10.3198/jpr2015.05.0034crg.

Kofoid, K.D., and T.L. Harvey. 2005. Registration of Greenbug Resistant Sorghum Germplasm Lines KS 116 A/B through KS 120 A/B. Crop Sci. 45(2): 802. doi: 10.2135/cropsci2005.0802.

Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82(13). doi: 10.18637/jss.v082.i13.

De La Vega, A.J., and S.C. Chapman. 2006. Multivariate analyses to display interactions between environment and general or specific combining ability in hybrid crops. Crop Sci. 46(2): 957–967. doi: 10.2135/cropsci2005.08-0287.

Laosuwan, P., and R.E. Atkins. 1977. Estimates of Combining Ability and Heterosis in Converted Exotic Sorghums 1. Crop Sci. 17(1): 47–50. doi: 10.2135/cropsci1977.0011183x001700010014x.

Larièpe, A., L. Moreau, J. Laborde, C. Bauland, S. Mezmouk, et al. 2017. General and specific combining abilities in a maize (Zea mays L.) test-cross hybrid panel: relative importance of population structure and genetic divergence between parents. Theor. Appl. Genet. 130(2): 403–417. doi: 10.1007/s00122-016-2822-z.

Laude, H.H., and A.F. Swanson. 1933. Sorghum production in Kansas /. https://www.ksre.k-state.edu/historicpublications/pubs/SB265.pdf (accessed 24

October 2019).

Lombardi, G.M.R., P.C.A. Navegantes, C.H. Pereira, J.M.O. Fonseca, R.A. da C. Parrella, et al. 2018. Heterosis in sweet sorghum. Pesqui. Agropecu. Bras. 53(5): 593–601. doi: 10.1590/S0100-204X2018000500008.

Lothrop, J.E., R.E. Atkins, and O.S. Smith. 1985. Variability for Yield and Yield Components in IAP1R Grain Sorghum Random-Mating Population. I. Means, Variance Components, and Heritabilities1. Crop Sci. 25(2): 235. doi: 10.2135/cropsci1985.0011183x002500020009x.

Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer.

Mbulwe, L., G.C. Peterson, J. Scott-Armstrong, and W.L. Rooney. 2015. Registration of sorghum germplasm Tx3408 and Tx3409 with tolerance to sugarcane aphid [*Melanaphis sacchari* (Zehntner)]. J. Plant Regist. 10(1): 51–56. doi: 10.3198/jpr2015.04.0025crg.

Menz, M.A., R.R. Klein, N.C. Unruh, W.L. Rooney, P.E. Klein, et al. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44(4): 1236–1244. doi: 10.2135/cropsci2004.1236.

Miller, F.R., T.F. Dusek, K.L. Prihoda, and L.W. Rooney. 1992. Registration of RTx436 sorghum parental line. Crop Sci. 32(6): 1518–1518. doi: 10.2135/cropsci1992.0011183x003200060059x.

Miller, F.R., and Y. Kebede. 1984. Genetic Contributions to Yield Gains in Sorghum, 1950 to 1980. In: Fehr, W.R., editor, Genetic Contributions to Yield Gains of Five Major Crop Plants. Volume 7. John Wiley & Sons, Ltd. p. 1–14

Mohammed, M.I. 2009. Line x tester analysis across locations and years in Sudanese x exotic lines of forage sorghum. J. Plant Breed. Crop Sci. 1(9): 311–319. http://www.academicjournals.org/jpbcs (accessed 18 December 2020).

Möhring, J., A.E. Melchinger, and H.P. Piepho. 2011. Reml-based diallel analysis. Crop Sci. 51(2): 470–478. doi: 10.2135/cropsci2010.05.0272.

NASS (U. S. Department of Agriculture National Agricultural Statistics Service). 2021. *2020 STATE AGRICULTURE OVERVIEW Texas and Kansas*. Available online at: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=TEXAS and https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=KANSAS (accessed Mar 25, 2021).

NASA. 2021. Prediction Of Worldwide Energy Resources (POWER). Available online at: https://power.larc.nasa.gov/data-access-viewer/ (accessed April 16, 2021)

Oliveira, I.C.M., T.D.S. Marçal, K.D.C. Bernardino, P.C.D.O. Ribeiro, R.A.D.C. Parrella, et al. 2019. Combining ability of biomass sorghum lines for agroindustrial characters and multitrait selection of photosensitive hybrids for energy cogeneration. Crop Sci. 59(4): 1554–1566. doi: 10.2135/cropsci2018.11.0693.

Onofri, A., N. Terzaroli, and L. Russi. 2020. Linear models for diallel crosses: a review with R functions. Theor. Appl. Genet. (Mahgoub 2011). doi: 10.1007/s00122-020-03716-8.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58(3): 545–554. doi: 10.1093/biomet/58.3.545.

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198(2): 483–495. doi: 10.1534/genetics.114.164442.

Perumal, R., T. Tesso, K.D. Kofoid, R.M. Aiken, P.V. V. Prasad, et al. 2019. Registration of six grain sorghum pollinator (R) lines. J. Plant Regist. 13(1): 113–117. doi: 10.3198/jpr2017.12.0087crp.

Pfeiffer, B.K., D. Pietsch, R.W. Schnell, and W.L. Rooney. 2019. Long-term selection in hybrid sorghum breeding programs. Crop Sci. 59(1): 150–164. doi: 10.2135/cropsci2018.05.0345.

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1–2): 209–228. doi: 10.1007/s10681-007-9449-8.

Podlich, D.W., and M. Cooper. 1998. Modelling plant breeding programs as search strategies on a complex response surface. Asia-Pacific Conference on Simulated Evolution and Learning. Springer. p. 171–178

Pollak, L.M. 2003. The history and success of the public-private project on germplasm enhancement of maize (GEM). Adv. Agron. 78: 46–89.

Prasad, B.H. V., and B. Biradar. 2017. Combining ability studies in grain sorghum using diallel analysis. J. FARM Sci. 30(2): 164–167. http://14.139.155.167/test5/index.php/kjas/article/viewFile/8387/8640 (accessed 18 December 2020).

R Core Team. 2020. R: A language and environment for statistical computing.

https://www.r-project.org/.

Rocha, M.J.D., J.A.R. Nunes, R.A.D.C. Parrella, P.S.D.S. Leite, G.M.R. Lombardi, et al. 2018. General and specific combining ability in sweet sorghum General and specific combining ability in sweet sorghum. Crop Breed. Appl. Biotechnol. 18: 365–372. doi: 10.1590/1984.

Rooney, W.. 2004. Sorghum improvement—integrating traditional and new technology to produce improved genotypes. Advances in agronomy. New York: Academic Press, 1949-. p. 37–109

Rooney, W.L., F.R. Miller, and L.W. Rooney. 2003. Registration of RTx437 sorghum parental line. Crop Sci. 43(1): 445. doi: 10.2135/cropsci2003.4450.

Rosenow, D.T., L.E. Clark, G.C. Peterson, G.N. Odvody, and W.L. Rooney. 2021. Registration of Tx3440 through Tx3482 sorghum germplasm. J. Plant Regist. (1–9): plr2.20082. doi: 10.1002/plr2.20082.

Shelton, A.C., and W.F. Tracy. 2017. Cultivar development in the U.S. public sector. Crop Sci. 57(4): 1823–1835. doi: 10.2135/cropsci2016.11.0961.

Sheunda, P., F.M. Nzuve, E.O. Manyasa, and G.N. Chemining'wa. 2019. Heterosis and combining ability for grain and biomass yield in sorghum hybrids for the semi-arid lowlands of Eastern Kenya. Acad. Res. J. Agric. Sci. Res. 7(4): 176–189. doi: 10.14662/ARJASR2019.035.

Smith, S. 2007. Pedigree background changes in U.S. hybrid maize between 1980 and 2004. Crop Sci. 47(5): 1914–1926. doi: 10.2135/cropsci2006.12.0763.

Sprague, G.F., and L.A. Tatum. 1942. General vs. Specific Combining Ability in Single

Crosses of Corn1. Agron. J. 34(10): 923. doi: 10.2134/agronj1942.00021962003400100008x.

Stephens, J.C., and R.F. Holland. 1954. Cytoplasmic Male-Sterility For Hybrid Sorghum Seed Production. Agron. J. 46(1): 20–23. doi: 10.2134/agronj1954.00021962004600010006x.

Stephens, J.C., F.R. Miller, and D.T. Rosenow. 1967. Conversion of Alien Sorghums to Early Combine Genotypes 1. Crop Sci. 7(4): 396–396. doi: 10.2135/cropsci1967.0011183X000700040036x.

Technow, F., D. Podlich, and M. Cooper. 2020. Back to the future: Implications of genetic complexity for hybrid breeding strategies. bioRxiv: 2020.10.21.349332. doi: 10.1101/2020.10.21.349332.

Technow, F., T.A. Schrag, W. Schipprack, E. Bauer, H. Simianer, et al. 2014. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197(4): 1343–1355. doi: 10.1534/genetics.114.165860.

Teodoro, L.P.R., L.L. Bhering, B.E.L. Gomes, C.N.S. Campos, F.H.R. Baio, et al. 2019. Understanding the combining ability for physiological traits in soybean. PLoS One 14(12). doi: 10.1371/journal.pone.0226523.

Wagaw, K., and T. Tadesse. 2020. Combining Ability and Heterosis of Sorghum (*Sorghum bicolor* L. Moench) Hybrids for Grain and Biomass Yield. Am. J. Plant Sci. 11(12): 2155–2171. doi: 10.4236/ajps.2020.1112151.

Xu, Y., P. Li, C. Zou, Y. Lu, C. Xie, et al. 2017. Enhancing genetic gain in the era of molecular breeding. J. Exp. Bot. 68(11): 2641–2666. doi: 10.1093/jxb/erx135.

Yu, J., and R. Bernardo. 2004. Changes in Genetic Variance during Advanced Cycle Breeding in Maize. Crop Sci. 44(2): 405–410. doi: 10.2135/cropsci2004.4050.

Yu, K., H. Wang, X. Liu, C. Xu, Z. Li, et al. 2020. Large-Scale Analysis of Combining Ability and Heterosis for Development of Hybrid Maize Breeding Strategies Using Diverse Germplasm Resources. Front. Plant Sci. 11(June): 1–16. doi: 10.3389/fpls.2020.00660.

Zhang, X., L. Lv, C. Lv, B. Guo, and R. Xu. 2015. Combining ability of different agronomic traits and yield components in hybrid barley. PLoS One 10(6). doi: 10.1371/journal.pone.0126828.

Zhang, X., P. Pérez-Rodríguez, J. Burgueño, M. Olsen, E. Buckler, et al. 2017. Rapid cycling genomic selection in a multiparental tropical maize population. G3 Genes, Genomes, Genet. 7(7): 2315–2326. doi: 10.1534/g3.117.043141.

3. ASSESSING COMBINING ABILITIES, GENOMIC DATA, AND GXE INTERACTIONS TO PREDICT HYBRID GRAIN SORGHUM PERFORMANCE*

3.1. Synopsis

Genomic selection in maize has been one factor that has increased the rate of genetic gain when compared to other cereals. However, the technological foundations in maize also exist in other cereal crops that would allow prediction of hybrid performance based on general (GCA) and specific (SCA) combining abilities applied through genomic-enabled prediction models. Further, the incorporation of genotype-by-environment (GxE) interaction effects presents an opportunity to deploy hybrids to targeted environments. To test these concepts, a factorial mating design of elite yet divergent grain sorghum lines generated hybrids for evaluation. Inbred parents were genotyped, and markers were used to assess population structure and develop the genomic relationship matrix (GRM). Grain yield, height, and days to anthesis were collected for hybrids in replicated trials, and best linear unbiased estimates were used to train classical GCA-SCA-based and genomic (GB) models under a hierarchical Bayesian framework. To incorporate population structure, GB was fitted using the GRM of both parents and hybrids. For GB models, GxE interaction effects were included by the Hadamard product between GRM and environments. A leave-one-out cross-validation scheme was used to study the prediction capacity of models. Classical and genomic

---

*Reprinted with permission from "Assessing combining abilities, genomic data, and GxE interactions to predict hybrid grain sorghum performance" by Fonseca, J.M.O., Klein, P.E., Crossa, J., Pacheco, A., Pérez-Rodríguez, P., Perumal, R., Klein, R.R., and Rooney, W.L., 2021. *The Plant Genome*, DOI 10.1002/tpg2.20127, Copyright [2021] by John Wiley & Sons, Inc.

models effectively predicted hybrid performance and prediction accuracy increased by including genomic data. GB models effectively partitioned the variation due to GCA, SCA, and their interaction with the environment. A strategy to implement genomic selection for hybrid sorghum breeding is presented herein.

3.2. Introduction

Genotypic data in breeding programs evolved from mapping simple inherited characteristics to the use of genetic markers for detailed genome-wide association studies of quantitatively-inherited traits (Nadeem et al., 2018). Despite these advancements, the successful application of marker-based genomic technologies for crop improvement is limited (Jannink et al., 2010). However, with the advent of genomic selection (GS), a growing wealth of information indicates that genome sequencing technologies are a valuable resource to augment traditional phenotypic based crop improvement efforts. A method for improving quantitative traits, GS relies on genomic-enabled prediction models to predict genetic values using all available molecular markers across the genome. A seminal report by Meuwissen *et al.* (2001) demonstrated the principles of GS by predicting breeding values (BV) via simulation studies, and the authors proposed that the implementation of GS in animal and plant breeding programs could increase the rate of genetic gain.

In crop improvement, genomic prediction studies in several crops have shown the potential to increase rates of genetic gain over classical phenotypic-based methodologies (Crossa et al., 2014; He et al., 2016; Bernal-Vasquez et al., 2017; Tan et al., 2017; Dias et al., 2018; Hunt et al., 2018; Xu et al., 2018; Islam et al., 2020).

Additionally, more cost-effective genotyping techniques have encouraged plant breeders to develop suitable models to introduce GS into the breeding pipeline (Eathington et al., 2007; Heffner et al., 2009, 2010; Bhat et al., 2016). While these GS methods have the potential to revolutionize the improvement in several crops, most of these methods have primarily benefitted a few crops that cover vast areas or have notably high value, including maize (*Zea mays* L.) and soybeans (*Glycine max* (L.) Merr.) (Technow *et al.*, 2012, 2014; Jarquín *et al.*, 2014; Stewart-Brown *et al.*, 2019). While other crops are beginning to experiment with GS in the breeding pipeline, few studies have addressed the issues related to genotype x environment (GxE) interactions, population structure in elite germplasm, and their effects on effectively predicting hybrid performance. Consequently, more complex genomic prediction models may be required to account for GxE effects and include alternative prediction objectives, i.e., general combining ability (GCA) or hybrid performance (Piepho et al., 2008).

Early genomic prediction models included statistical assumptions that were not applicable to hybrid crops (Technow et al., 2012; de los Campos et al., 2015). For instance, standard genomic best linear unbiased prediction (GBLUP) models assume equal variance across all loci and population homogeneity (Clark et al., 2011). By their very nature, hybrid crops are bred with population structure (i.e., heterotic groups) and, consequently, less restrictive genomic prediction models for hybrid crops are needed (de los Campos et al., 2015). To circumvent these limitations, extensions of standard GBLUP models have been developed under a hierarchical Bayesian framework (Meuwissen et al., 2001; Clark et al., 2011; Heslot et al., 2012; Technow et al., 2012;

Zhao et al., 2013; de los Campos et al., 2015; Ramstein and Casler, 2019). This approach models the genetic variation using prior information to define hyperparameters from which variances can be sampled to adjust the effect of each locus separately (Meuwissen *et al.*, 2001; de los Campos *et al.*, 2009; Gianola *et al.*, 2009; Habier *et al.*, 2011). Alves et al. (2019) showed some advantages of Bayesian models to predict the performance of maize hybrids.

Parametric Bayesian models can incorporate population structure through kernel methods that represent genetic similarities between inbred lines via a genomic (co)variance matrix (VanRaden, 2008; de los Campos *et al.*, 2009). Essentially, kernels include dense whole-genome marker information in genomic-enabled prediction models via a genomic relationship matrix (GRM) whose dimensions are equal to the number of inbred lines. This drastically reduces problems related to data dimensionality (i.e., big p small n) (Gianola and de los Campos, 2008). Non-linear kernels can also be used in semi-parametric models, and several studies report superior accuracy on genomic predictions (Gianola and de los Campos, 2008; Crossa et al., 2011; Morota and Gianola, 2014; Costa-Neto et al., 2020). The application of kernel methods in genomic prediction models substantially increased the opportunities for plant breeders. For example, kernels can be used with almost any information set (e.g., covariates, strings, images, and graphs) and, due to their flexibility, kernels can accommodate multiple information within a genomic prediction model (de los Campos *et al.*, 2009; de los Campos *et al.*, 2010; Crossa *et al.*, 2011). Moreover, GxE effects can be designed with a kernel based on the interaction effect between markers (i.e., G-matrix) and environments (Burgueño et al., 2012; Lopez-Cruz et al., 2015; Cuevas et al., 2016, 2017; Acosta-Pech et al., 2017; Sukumaran et al., 2018).

The GS methodologies developed for maize may be applicable to other hybrid grain crops such as sorghum *(Sorghum bicolor (L.) Moench)* (Hunt et al., 2018), rice *(Oryza sativa)* (Xu et al., 2014; Wang et al., 2017), sunflower *sunflower (Helianthus annuus L.)* (Mangin et al., 2017), and wheat (*Triticum aestivum)* (Basnet et al., 2019). Initial studies on genomic prediction models in sorghum indicated that GS could be effective (de Oliveira et al., 2018; Fernandes et al., 2018; Hunt et al., 2018; Velazco et al., 2019; dos Santos et al., 2020; Sapkota et al., 2020). However, GS is not implemented in most sorghum breeding pipelines due to limited model testing and optimization for predicting hybrid performance. For instance, genomic prediction studies in sorghum have focused on predicting BV (de Oliveira et al., 2018; Fernandes et al., 2018; Hunt et al., 2018; dos Santos et al., 2020). As such, these models predict only the GCA of untested sorghum inbred lines. In addition, most genomic prediction studies involving sorghum consist of diverse germplasm panels and/or testcrosses derived therefrom, but these populations are not representative of the heterotic groups used in hybrid seed production (Fristche-Neto et al., 2018). Finally, given the importance of GxE effects in all crop improvement efforts and the extreme variability of sorghum production environments, the inclusion of GxE effects in genomic prediction models seems imperative to improve selection efficiency and hybrid deployment (Hunt et al., 2018).

This study introduces the concept of predicting the performance of hybrid grain crops using sorghum as a model species. We implement models with GCA and SCA genetic main effects in two forms; a classical and a genomic hierarchical Bayesian framework. Through this research, we have addressed critical issues associated with the implementation of GS into an applied hybrid crop improvement program including GxE interactions, population structure, environment-specific training sets, and selection intensity on the coincidence between observed and predicted values. Based on the results, we propose a conceptual framework for implementing GS into the breeding pipeline for those hybrid crops that lack the level of resources employed in crops such as maize, which includes grain sorghum.

3.3. Material and methods

3.3.1. Phenotypic data

A set of 20 grain sorghum inbred lines developed in sorghum breeding programs at Texas A&M AgriLife Research (College Station, Texas) and Kansas State Research Center (Hays, Kansas) were selected and crossed using a factorial mating scheme (Comstock and Robinson, 1952). Each breeding program provided ten lines equally represented by female (A-lines) and male (R-lines) parents which are adapted to their respective target environments and had been evaluated in hybrid combinations prior to this study (Table 1). The 100 hybrids naturally divided into four groups of 25 hybrids: TAM (A)/TAM (R) (TxT); TAM (A)/KSU (R) (TxK); KSU (A)/TAM (R) (KxT); and KSU (A)/KSU (R) (KxK) (Figure 1).

For phenotyping, hybrids were planted in a randomized complete block design (RCBD) with a set (i.e., the four groups of 25 hybrid combinations aforementioned) in replication adjustment. In each replication, randomization occurred within and between sets to keep the set structure; entries were randomized within the set while sets were randomly ordered. In 2018, the tests were in five environments: Monte Alto (RF) (26°21'06.1"N 97°53'50.3"W), Victoria (VC) (28°47'24.4"N 96°50'22.6"W), and College Station (CS) (30°32'56.6"N 96°26'11.5"W) in Texas, and Garden City (GC) (37°59'21.4"N 100°48'52.5"W), and Colby (COL) (39°22'56.6"N 101°04'45.0"W) in Kansas. In 2019, the environments were similar except that the Monte Alto location was replaced by Taft (TA) (28°00'05.4"N 97°15'12.4"W), Texas, and Garden City was replaced by Hays (HAY) (38°51'10.8"N 99°20'24.6"W), Kansas. These locations represent distinct adaptation zones and, in each test, agronomic practices standard to the location were implemented (Table 2). In 2018 trials, each location had three replications while two replications were planted in each site in 2019. A plot consisted of two rows, approximately 5.3 m in length, with row spacing between 0.76 and 1.0 m, consistent with production practices in each environment.

In this study, three agronomic traits were measured. Days to anthesis (DA) were recorded as the number of days from planting to when 50% of the plants in the plot were at mid-anthesis; at Kansas locations, DA was collected only on the first replication. Just prior to harvest, plant height (PHT) was measured in cm using a representative plant in each plot at the length from the soil to the tip of the panicle. Plots were harvested using plot combines fitted with a Harvest Master GrainGage System (Juniper Systems), to measure grain weight and moisture content. After adjusting yield to a 14% moisture, grain yield (GY) was converted to ton ha^{-1}.

The data were analyzed in each environment using the following linear mixed model:

$$\boldsymbol{y} = \boldsymbol{1}\boldsymbol{\mu} + \boldsymbol{Z_1}\boldsymbol{h} + \boldsymbol{Z_2}\boldsymbol{r} + \boldsymbol{e}, \quad (2)$$

where $\boldsymbol{y}$ is the vector of phenotypes; $\boldsymbol{\mu}$ is an intercept; $\boldsymbol{h}$ is a random effect of hybrid, $\boldsymbol{h} \sim N(\boldsymbol{0}, \sigma_h^2\boldsymbol{I})$, $\boldsymbol{r}$ is a random effect of replicates, $\boldsymbol{r} \sim N(\boldsymbol{0}, \sigma_r^2\boldsymbol{I})$; $\boldsymbol{e}$ is the vector of residuals, $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma_e^2\boldsymbol{I})$; $\boldsymbol{1}$ is a vector of ones; $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are incidence matrixes; σ_h^2, σ_r^2 and σ_e^2 are variance components for hybrids, replicates and residuals, respectively. Model (1) was extended to incorporate environment and GxE interaction effects for performing a combined analysis. Environments were defined as a year and location combination, hybrid effects were partitioned into GCA and SCA, and GxE effects were represented by the interaction between GCA and SCA with the environment as follows:

$$\boldsymbol{y} = \boldsymbol{1}\boldsymbol{\mu} + \boldsymbol{Z_1}\boldsymbol{f} + \boldsymbol{Z_2}\boldsymbol{m} + \boldsymbol{Z_3}\boldsymbol{fm} + \boldsymbol{Z_4}\boldsymbol{s} + \boldsymbol{Z_5}\boldsymbol{fs} + \boldsymbol{Z_6}\boldsymbol{ms} + \boldsymbol{Z_7}\boldsymbol{fms} + \boldsymbol{Z_8}\boldsymbol{r}(\boldsymbol{s}) + \boldsymbol{e}, \quad (2)$$

where $\boldsymbol{f}$ is a vector of GCA effects of females, $\boldsymbol{f} \sim MVN(\boldsymbol{0}, \sigma_f^2\boldsymbol{I})$; m is a vector of GCA effects of male, $\boldsymbol{m} \sim N(\boldsymbol{0}, \sigma_m^2\boldsymbol{I})$; $\boldsymbol{fm}$ is a vector of SCA effects of hybrid combinations, $\boldsymbol{fm} \sim N(\boldsymbol{0}, \sigma_{fm}^2\boldsymbol{I})$; **s** is a vector of environmental effects, $\mathbf{s} \sim N(\boldsymbol{0}, \sigma_s^2\boldsymbol{I})$, $\boldsymbol{fs}$ is a vector of the interaction effect between GCA of female and environment, $\mathbf{fs} \sim N(\boldsymbol{0}, \sigma_{fs}^2\boldsymbol{I})$; $\boldsymbol{ms}$ is a vector of the interaction effect between GCA of male and environment, $\mathbf{ms} \sim N(\boldsymbol{0}, \sigma_{ms}^2\boldsymbol{I})$, $\boldsymbol{fms}$ is a vector of the interaction effect between SCA of hybrid combinations and environment, $\boldsymbol{fms} \sim N(\boldsymbol{0}, \sigma_{fms}^2\boldsymbol{I})$; **r(s)** is the vector of replication effect nested within environment, $\boldsymbol{r}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \sigma_{r(s)}^2\boldsymbol{I})$; $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Z_3}$, $\boldsymbol{Z_4}$, $\boldsymbol{Z_5}$, $\boldsymbol{Z_6}$, $\boldsymbol{Z_7}$, and $\boldsymbol{Z_8}$ are incidence matrixes for effects.

Variance components were estimated via restricted maximum likelihood (REML) method (Patterson and Thompson, 1971), and significance assessed by the likelihood ratio test (LRT) at a 5% significance level using the *ranova* function of the *lmerTest* R package (Kuznetsova et al., 2017). To assess phenotypic data, the *lmer* function of the *lme4* R package was applied (Bates et al., 2015). From the variance component estimates, repeatability (i.e., broad sense heritability) at each environment was calculated as $R = \frac{\sigma_h^2}{\sigma_h^2 + \frac{\sigma_e^2}{r}}$, whereas, for combined environments, $R = \frac{\sigma_h^{2*}}{\sigma_h^{2*} + \frac{\sigma_{hs}^{2*}}{t} + \frac{\sigma_e^2}{tr}}$, where *t* is the number of environments; *r* is the number of replications; $\sigma_h^{2*} = \sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2$; and $\sigma_{hs}^{2*} = \sigma_{fs}^2 + \sigma_{ms}^2 + \sigma_{fms}^2$. To assess the quality of the experiment, coefficient of variation (CV$_e$) was calculated as $CV_e = \frac{\sqrt{\sigma_e^2}}{\bar{x}}$, where $\bar{x}$ is the mean of a given trait. Finally, the best linear unbiased estimates (BLUEs) of hybrids were calculated for each traits using model (1) via the solution of Henderson's mixed equations (Henderson, 1984), assuming hybrids as fixed. BLUEs were later employed to develop genomic prediction models and calculate the Pearson's correlation between traits.

3.3.2. Genotypic data

DNA extraction and genotyping-by-sequencing (GBS) were conducted according to Morishige *et al.* (2013). In brief, seed from the parental lines were germinated in a greenhouse and DNA extracted from ~14-day-old leaf tissue. Illumina template libraries were prepared by digesting each DNA sample with the restriction enzyme *Ngo*MIV. Fragments from each DNA sample were ligated to unique identifier barcodes; samples were pooled together for PCR-amplification, and templates sequenced on an Illumina

HiSeq 2500 sequencer. Sequences were mapped to the *Sorghum bicolor* BTx623 reference genome (Sbicolor v3.1.1) downloaded from Phytozome (https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Phytozome), and single nucleotide polymorphisms (SNPs) detected using the CLC Genomics Workbench v20 (CLC Bio, Aarhus, Denmark). Genomic polymorphisms scored in at least 25% of the parental lines were called SNPs, and a total of 117,582 SNPs were identified. Markers with more than 5% missing values were removed. For the remaining markers, missing values were imputed using FastPHASE (Scheet and Stephens, 2006), and SNPs with minor allele frequency (MAF) < 5% were removed. After quality control, 35,546 SNP were available for further analysis.

Genomic relationship matrixes for female and male parents were computed using markers, the matrices were computed as follows: $G_k = W_k W_k'/r$, where W_k is a matrix of markers centered and standardized, r is the number of markers, $k \in \{f, m\}$ (Technow et al., 2014; Lopez-Cruz et al., 2015). Alternatively, the genetic distance between parents was calculated on a pairwise basis using the Prevosti's absolute genetic distance (Prevosti et al., 1975) implemented in *poppr* R package (Kamvar and Grünwald, 2020). Distances were computed based on the differences in the frequencies of SNPs as $D = \frac{1}{2r}\sum_{j=1}^{r}\sum_{k=1}^{s(j)}\left|p_{1jk} - p_{2jk}\right|$, where r is the number of SNPs considered, p_{1jk} the frequency of the SNP k in the chromosome j in the first parent, and p_{2jk} the corresponding value in the second parent. A principal coordinate analysis (PCoA) of the genetic distance was plotted to represent the genetic relatedness of parents. Clustering analysis was performed using the function *CLARA* (Struyf et al., 1997) from R packages *ggplot2* (Wickham, 2016) and *ggfortify* (Tang et al., 2016).

3.3.3. Statistical models

Kernel-based Bayesian GBLUP models (GB) and classical GCA-SCA-based models under a Bayesian Ridge Regression (BRR) framework were fitted to predict yield, height, and days to anthesis of grain sorghum hybrids. These models differed on the (co)variance matrix related to GCA and SCA effects; genomic models included genomic information to account for heterogeneity within parents and hybrids (i.e., accommodate population structure), whereas classical GCA-SCA-based models replaced genomic information with the identity matrix (i.e., without the genomic kernel). GB and BRR models were developed based on genetic main effects of GCA (seed and males) and SCA at three distinct levels of complexity: a) for a single environment, b) multi-environment , and c) multi-environment plus the GxE interaction effect

3.3.3.1. Single-environment main genetic effect model (SM)

The linear model adjusted the BLUE of hybrids derived from model (1) separately for each environment and included GCA and SCA genetic effects as follows:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_1 f} + \boldsymbol{Z_2 m} + \boldsymbol{Z_3 h} + \boldsymbol{e}, \tag{3}$$

where $\boldsymbol{y} = [\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}]'$ is the response vector, and $\boldsymbol{y_n}$ represents the observation in the i^{th} hybrid ($i = 1, \cdots, \text{n}$) in each environment; **μ** is an intercept, $\boldsymbol{f}$ is the vector of GCA effects for females, $\boldsymbol{f} \sim N(\mathbf{0}, \sigma^2_{f_j}\boldsymbol{G}_f)$, where $\boldsymbol{j}$ represents environment; $\boldsymbol{m}$ is a vector of GCA effects for males, $\boldsymbol{m} \sim N(\mathbf{0}, \sigma^2_{m_j}\boldsymbol{G}_m)$; $\boldsymbol{h}$ is a vector of SCA effects for hybrid combinations, $\boldsymbol{h} \sim N(0, \sigma^2_{h_j}\boldsymbol{H})$; $\boldsymbol{e}$ is a vector of residuals, $\boldsymbol{e} \sim N(\mathbf{0}, \sigma^2_{e_j}\boldsymbol{I})$, **1** is a vector of ones; $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, and $\boldsymbol{Z}_3$ are incidence matrixes for females, males, and hybrids, respectively; and $\boldsymbol{G}_f$, $\boldsymbol{G}_m$, and $\boldsymbol{H}$ are the GRM for females, males, and hybrids, respectively. The elements of matrix $\boldsymbol{H}$ were calculated *in silico* by the Kronecker product of $\boldsymbol{G}_f$ and $\boldsymbol{G}_m$ as $\boldsymbol{H} = \boldsymbol{G}_f \otimes \boldsymbol{G}_m$ (Technow *et al.*, 2014; Acosta-Pech *et al.*, 2017). This model was also used to build prediction models using the grand mean of hybrids across environments (i.e., mean of the BLUEs). For BRR, model (3) was fitted without incorporating genomic kernels, thus the GRM of parents and hybrids were replaced by an identity matrix.

3.3.3.2. Multi-environment main genetic effect model (MM)

One multi-environment model that includes the fixed effects of environments and the random genetic effects across environments is presented by Bandeira e Sousa et al., (2017) as follows

$$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{Z_s\beta} + \boldsymbol{Z_1 f} + \boldsymbol{Z_2 m} + \boldsymbol{Z_3 h} + \boldsymbol{e}, \tag{4}$$

where $\boldsymbol{y} = [\boldsymbol{y_1}, \cdots, \boldsymbol{y_j}, \cdots, \boldsymbol{y_s}]'$ is the vector of observations of hybrids ($i = 1, \cdots, \mathrm{n}_j$) in the jth environment ($j = 1, \cdots, \mathrm{s}$); $\boldsymbol{\beta}$ is a vector of the environmental effects, $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I})$; and $\boldsymbol{Z_s}$ is the incidence matrix that connects hybrids to phenotypes for each environment. Random effects are described in model (3) and are assumed to be equal across environments (see Lopez-Cruz et al., 2015). For BRR, model (4) was fitted without incorporating genomic kernels to genetic main effects.

3.3.3.3. Multi-environment main genetic effect model (MM)

The following model (5) is an extended version of model (4) to include GxE effects as proposed by Acosta-Pech *et al.* (2017). These authors considered the incorporation of GxE effects (u) by applying the Hadamard product between genetic main effects (i.e., $\boldsymbol{G}_f$, $\boldsymbol{G}_m$, and $\boldsymbol{H}$) and the environmental effect. The model is given as follows:

$$\boldsymbol{y} = \boldsymbol{Z}_s\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{f} + \boldsymbol{Z}_2\boldsymbol{m} + \boldsymbol{Z}_3\boldsymbol{h} + \boldsymbol{u}_f + \boldsymbol{u}_m + \boldsymbol{u}_h + \boldsymbol{e}, \tag{5}$$

where $\boldsymbol{u}_f$ is a random vector to include the interaction between GCA of females and environment, $\boldsymbol{u}_f \sim N(\boldsymbol{0}, \sigma^2_{fs}\boldsymbol{V}_f)$; $\boldsymbol{u}_m$ is a random vector that models the interaction between males and environment, $\boldsymbol{u}_m \sim N(\boldsymbol{0}, \sigma^2_{ms}\boldsymbol{V}_m)$; u_h is a vector that takes into account the interaction between hybrids and environment, $\boldsymbol{u}_h \sim N(\boldsymbol{0}, \sigma^2_{hs}\boldsymbol{V}_h)$; $\boldsymbol{V}_f$, $\boldsymbol{V}_m$, and $\boldsymbol{V}_H$ are variance-covariance matrices that associate each respective genomic-main-effect-by-environment variance component, i.e., σ^2_{fs}, σ^2_{ms}, and σ^2_{hs}. The variance-covariance matrix was calculated as $V_f = \boldsymbol{Z}_1\boldsymbol{G}_f\boldsymbol{Z}'_1\#\boldsymbol{Z}_s\boldsymbol{Z}'_s$, $\boldsymbol{V}_m = \boldsymbol{Z}_2\boldsymbol{G}_m\boldsymbol{Z}'_2\#\boldsymbol{Z}_s\boldsymbol{Z}'_s$, and $\boldsymbol{V}_h = \boldsymbol{Z}_3\boldsymbol{H}\boldsymbol{Z}'_3\#\boldsymbol{Z}_s\boldsymbol{Z}'_s$ where # stands for the Hadamard product (see Acosta-Pech *et al.* (2017) for further details about the derivations). For BRR, model (5) was fitted without incorporating genomic kernels to genetic main effects and GxE was calculated as $\boldsymbol{V}_{f\prime} = \boldsymbol{Z}_1\boldsymbol{Z}'_s$, $\boldsymbol{V}_{m\prime} = \boldsymbol{Z}_2\boldsymbol{Z}'_s$, and $\boldsymbol{V}_{h\prime} = \boldsymbol{Z}_1\boldsymbol{Z}'_2\#\boldsymbol{Z}_s\boldsymbol{Z}'_s$ for the interaction of GCA$_\text{f}$, GCA$_\text{m}$, and SCA with the environment.

3.3.4. Software and cross-validation Scheme

GB and BRR models were fit using Bayesian methods implemented in the *BGLR* statistical package (Pérez and de los Campos, 2014) available for R (R Core Team 2020). Inferences were based on 10,000 Gibbs sampler iterations with a *burn-in* of 5,000, and a *thin* of 5. For BRR models (i.e., classical GCA-SCA-based models 3, 4, and

5), all effects assumed Gaussian prior distribution and included an identity (I) matrix to structure the data set, whereas, for GB models, only environmental effect followed this assumption. The remaining effects assumed a Gaussian prior distribution and applied the kernel-based method to create the appropriate variance-covariance structures (Pérez and de los Campos, 2014).

To validate models and estimate the predictability of parents, a simple leave-one-out cross-validation (CV) scheme was applied following Basnet *et al.* (2019). For that, seed parents in hybrid combinations were removed, and models were then trained with the remaining records where that seed parent was absent, i.e., predicting hybrids of unobserved females (T1F). The same method was applied to pollinator parent (T1M). Further, seed and pollinator parents were removed simultaneously to predict hybrids (T0MF). Finally, T1F and T1M were applied to train models considering just TxT and KxK hybrids under subtropical and temperate environments (Table 2). Subtropical and temperate environments were assumed to be the target environments for the respective TxT and KxK hybrids because each set of hybrids derived from parents adapted to each respective environment. The prediction accuracy (r) was estimated as the Pearson's correlation between observed and predicted values of hybrids.

3.3.5. Coincidence index

The coincidence index (CI) originally proposed by Hamblin and Zimmermann (1986) aimed to assess the proportion of coincident genotypes selected under two different planting systems assuming a specific selection intensity (SI). Herein, a modified version of such CI is proposed to calculate the optimum SI that maximizes the

CI between observed and predicted hybrids utilizing a genomic-enabled prediction model. Calculations of CI was performed using GB-MMGE and BRR-MMGE under T0MF. For that, simulated SI values ranging from 1-50% were considered and CI computed as:

$$CI = \frac{C-R}{T-R},$$

where C is the number of coincident hybrids between observed and predicted values; T is the number of observed hybrids selected according to the SI; and R: the number of expected hybrids selected by chance, i.e., a fraction of T that also varies depending on a SI. For instance, considering a population of 100 hybrids, if SI = 10; then T = 10 (100*.10), R = 1(10*.10), and C will depend on the number of coincident hybrids between observed and predicted values. If C = 2, therefore CI = 1/9 = 11%.

3.4. Results

3.4.1. Phenotypic analysis

The average of the best linear unbiased estimate (BLUE) of hybrids across environments varied significantly for all traits (Figure 3). Values for hybrid GY, ranged from 1.83 to 10.98 ton ha^{-1}, from 100 to 177 cm for PHT, and from 54 to 88 days for DA. The lowest mean hybrid GY occurred in 2018 Monte Alto, TX and the highest occurred in 2019 Colby, KS. The repeatability of PHT and DA were consistently high while the repeatability of GY was lower and varied more across environments (Figure 2). Coefficient of variation (CVe) indicated that data variation within locations were typical for sorghum grain hybrid yield trials (Figure 4).

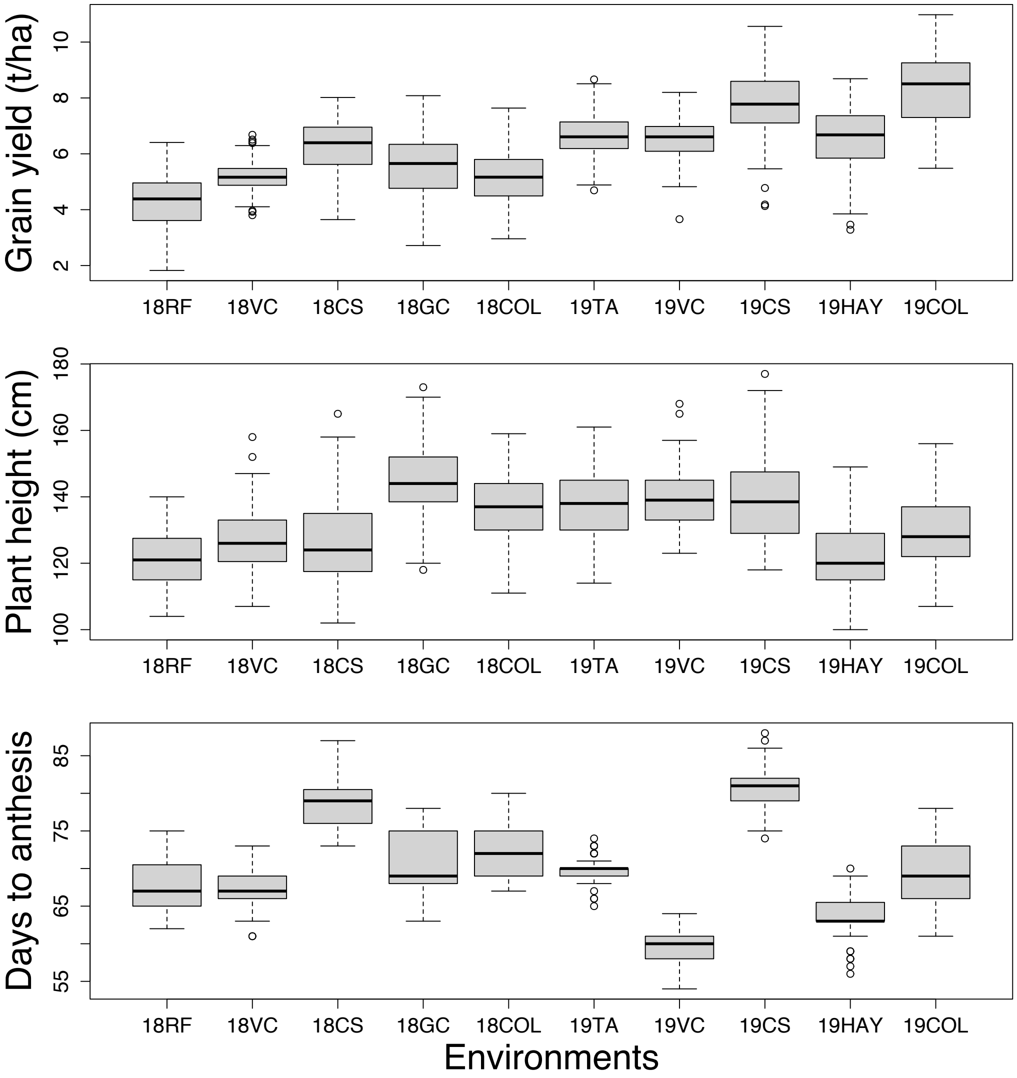

Figure 3. Boxplot of best linear unbiased estimate (BLUE) for grain yield (t ha^{-1}), plant height (cm), and days to anthesis in each environment. Environments are indicated as the combination of year and location where hybrid trials were conducted. Designations for environments are as follows: Monte Alto 2018 (18RF), Victoria 2018 (18VC), College Station 2018 (18CS), Garden City 2018 (18GC), Colby 2018 (18COL), Taft 2019 (19TA), Victoria 2019 (19VC), College Station 2019 (19CS), Hays 2019 (19HAY), and Colby 2019 (19COL).

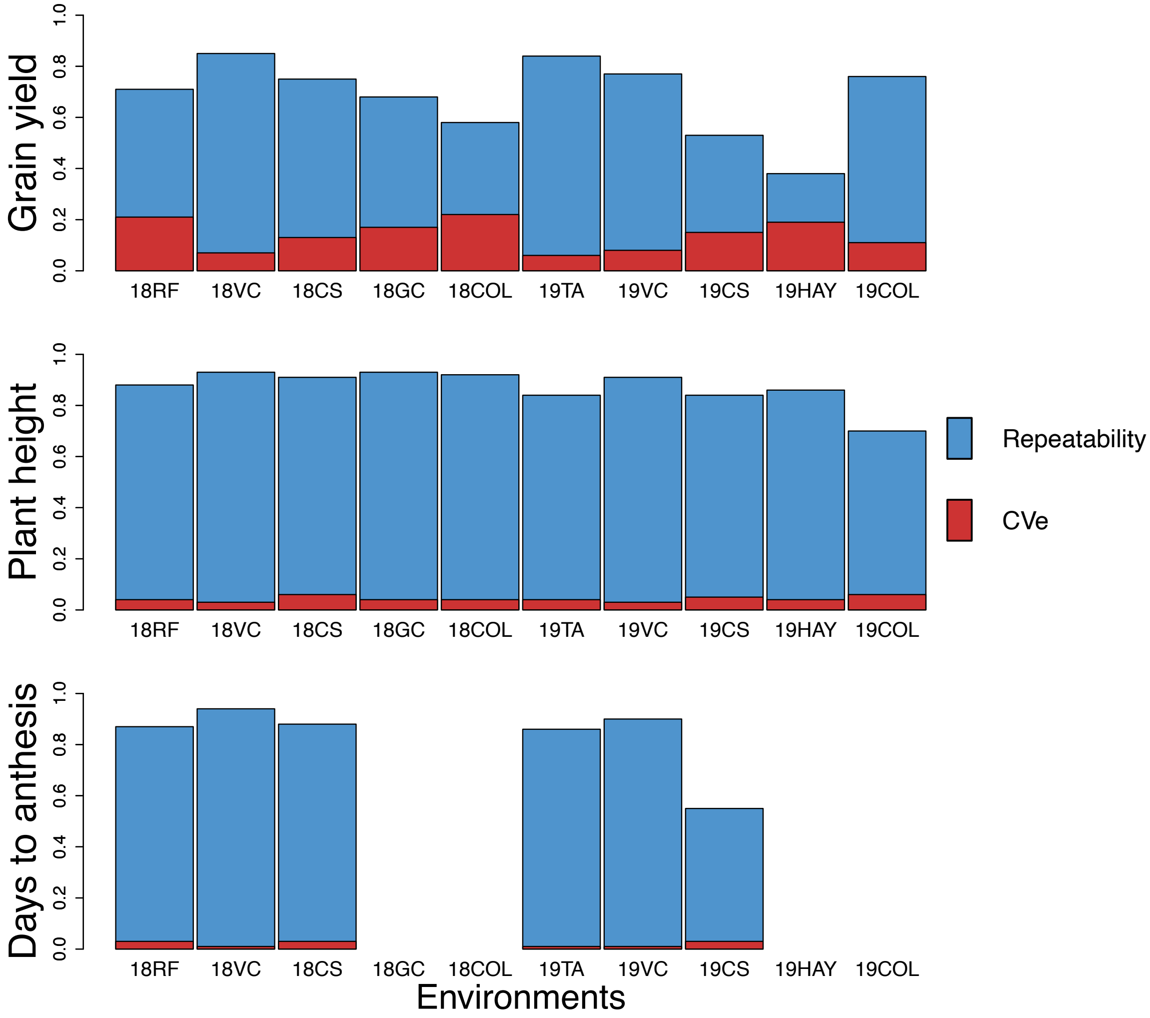

Figure 4. Repeatability and experimental coefficient of variation (CVe) estimates for grain yield, plant height, and days to anthesis in each environment. Environments are indicated as the combination of year and location where hybrid trials were conducted. Designations for environments are as follows: Monte Alto 2018 (18RF), Victoria 2018 (18VC), College Station 2018 (18CS), Garden City 2018 (18GC), Colby 2018 (18COL), Taft 2019 (19TA), Victoria 2019 (19VC), College Station 2019 (19CS), Hays 2019 (19HAY), and Colby 2019 (19COL).

The likelihood ratio test indicated that all genetic effects and their interactions with the environment were significant (Table 9). As expected, the environmental effect was highly significant, accounting for the majority of variation for GY (50%) and DA (76%). While the majority of the genetic variation was associated with GCA, SCA was also significant, indicating the importance of specific hybrid combinations to maximize productivity.

Pearson's correlations between traits were assessed for each environment and across environments assuming a 95% confidence interval (data not shown). While correlations between traits were somewhat inconsistent across individual environments, correlations across all environments were significant and positive between GY and PHT (0.17 ±0.06), and PHT and DY (0.15 ±0.06). Overall correlations across all environments indicate that taller hybrid grain sorghums are generally higher in yield; however this trend is mitigated by environmental stresses in selected environments (Rooney et al., 2004).

Table 9. Variance components, repeatability, and experimental coefficient of variation (CVe) estimates derived from the combined analysis for grain yield (GY), plant height (PHT), and days to anthesis (DA). Likelihood ration test (LRT) indicate the significance of effects. The percent of the total variation (%) for each variance component is shown with its respective estimate.

Variance components	GY		PHT		DA	
	Estimate	%	Estimate	%	Estimate	%
(Hybrid)	(0.315)	(10.4)	(89.6)	(42.5)	(6.0)	(11.9)
GCA_f	0.128^{***}	4.2	34.2^{***}	16.2	2.5^{***}	5.0
GCA_m	0.150^{***}	4.9	47.1^{***}	22.3	3.2^{***}	6.4
SCA	0.037^{***}	1.2	8.3^{***}	3.9	0.3^{***}	0.5
Environment	1.430^{**}	47.5	67.9^{***}	31.9	37.8^{***}	75.0
(Hybrid x Env)	(0.382)	(12.7)	(18.3)	(8.7)	(2.9)	(5.6)
GCA_f x Env	0.175^{***}	5.8	5.2^{***}	2.5	1.3^{***}	2.6
GCA_m x Env	0.157^{***}	5.2	9.3^{***}	4.4	1.3^{***}	2.5
SCA x Env	0.050^{*}	1.6	3.7^{***}	1.7	0.2^{*}	0.4
Rep(Env)	0.082^{***}	2.7	3.3^{***}	1.6	0.6^{***}	1.1
Residual	0.796	26.5	32.5	15.4	3.2	6.3
Repeatability	0.819		0.966		0.934	
CVe	0.144		0.043		0.024	

*, **, *** Significant at 0.05, 0.01, 0.001 probability level, respectively

3.4.2. Genetic distance between inbred parents

The plotted Principal Coordinates (PCoA) explained almost 42% of the total genetic variation among inbred parents in this study (Figure 3). For the most part, the A-line parents were a distinct group although some R-lines clustered within this group. The R-lines within the A-line cluster originated from the Kansas breeding program, and their pedigrees reveal the presence of B-line lineage. Similarly, one A-line from Kansas clustered within the male group, indicating some R-line lineage within its pedigree. Overall there was greater genetic variation in the R-lines, which has been well documented in previous studies (Menz et al., 2004; Crozier et al., 2020). In general, the genetic distance within Texas R-lines was greater than within Kansas R-lines (Figure 2).

3.4.3. Predictive ability across parents, traits, and models to predict hybrids

The predictive ability of parents to predict hybrids varied substantially across traits and models. Across their hybrids, KS118 and R07178 had the highest predictabilities for an A-line and R-line, respectively. As a group, the A-lines had higher prediction ability than the R-lines. Concerning traits, predictions for PHT and DA were better than for GY, regardless of the genetic group (Table 10).

The inclusion of the environmental effect (MM) reduced the accuracy of models to predict hybrids in most cases while the inclusion of the GxE effect (MMGE) considerably increased the prediction ability of models. These increases in prediction ability of models were greatest when parental genomic information was included (GB). For GB, the superiority of models that included the GxE effects was consistent across traits (Table 10).

Table 10. Prediction accuracy of classical GCA-SCA-based Bayesian Ridge Regression (BRR) and kernel-based Bayesian GBLUP (GB) models for single environment (SM), multi-environment (MM), and multi-environment plus GxE effect (MMGE) from leave-one-female-out , and leave-one-male-out cross-validation schemes for grain yield, plant height, and days to anthesis. The average prediction ability across all A- and R-lines represents the overall model predictability. Higher predictabilities within BRR and GB approaches are in black boldface, while the best predictability between approaches are in blue boldface.

	Parents	Grain Yield					
		BRR			GB		
		SM*	MM	MMGE	SM*	MM	MMGE
A-line	A05071	**0.913**	0.496	0.492	**0.923**	0.496	0.652
	A08140	**0.792**	0.455	0.452	**0.785**	0.466	0.588
	AARCH11051	**0.851**	0.666	0.666	**0.696**	0.668	0.673
	AARCH11129	**0.616**	0.300	0.302	**0.648**	0.306	0.373
	AARCH11136	**0.633**	0.455	0.458	0.597	0.463	**0.689**
	AARCH11146	0.446	**0.489**	0.487	0.508	0.484	**0.625**
	AHF14	0.048	**0.468**	0.467	0.040	0.457	**0.597**
	AKS118	**0.732**	0.626	0.619	**0.748**	0.627	0.706
	ATx3408	0.434	**0.528**	0.521	0.388	0.526	**0.681**
	ATx645	0.348	0.490	**0.494**	0.396	0.500	**0.667**
	Mean	0.581	0.497	0.496	0.573	0.499	**0.625**
R-line	EON361	0.208	0.445	**0.518**	0.251	0.443	**0.546**
	R07178	**0.881**	0.487	0.615	**0.904**	0.490	0.640
	R08304	0.370	0.335	**0.481**	0.459	0.347	**0.508**
	RARCH11001	0.136	0.413	**0.550**	0.463	0.422	**0.624**
	RARCH11002	0.231	0.422	**0.664**	0.481	0.438	**0.712**
	RARCH11028	**0.548**	0.251	0.467	**0.559**	0.260	0.473
	RARCH11055	0.503	0.556	**0.623**	0.318	0.554	**0.577**
	RARCH11056	**0.597**	0.381	0.568	0.282	0.387	**0.563**
	RTx436	**0.583**	0.415	0.497	**0.623**	0.407	0.488
	RTx437	**0.640**	0.447	0.614	**0.636**	0.447	0.575
	Mean	0.470	0.415	0.560	0.498	0.420	**0.571**

	Parents	Plant height					
		BRR			GB		
		SM*	MM	MMGE	SM*	MM	MMGE
A-line	A05071	0.730	**0.789**	**0.789**	0.752	0.800	**0.845**
	A08140	0.719	**0.773**	0.772	0.817	0.806	**0.851**
	AARCH11051	**0.874**	0.853	0.854	**0.918**	0.866	0.912
	AARCH11129	**0.817**	0.805	0.802	0.846	0.820	**0.872**
	AARCH11136	**0.910**	0.779	0.777	**0.882**	0.774	0.816
	AARCH11146	**0.921**	0.744	0.745	**0.921**	0.733	0.791
	AHF14	**0.745**	0.732	0.733	0.781	0.750	**0.803**
	AKS118	**0.909**	0.739	0.740	**0.926**	0.748	0.816
	ATx3408	0.829	**0.839**	0.835	0.817	0.841	**0.898**
	ATx645	0.620	0.719	**0.720**	0.694	0.752	**0.825**
	Mean	0.807	0.777	0.777	0.835	0.789	**0.843**

Table 10. Continued

R-line	EON361	0.397	0.607	**0.614**	0.525	0.651	**0.676**
	R07178	0.564	0.765	**0.781**	0.764	0.809	**0.841**
	R08304	0.710	0.805	**0.824**	0.755	**0.825**	**0.825**
	RARCH11001	0.759	0.714	**0.771**	0.715	0.764	**0.830**
	RARCH11002	**0.827**	0.783	0.821	0.836	0.809	**0.858**
	RARCH11028	0.637	0.640	**0.683**	0.632	0.651	**0.693**
	RARCH11055	0.697	0.829	**0.831**	0.685	0.802	**0.809**
	RARCH11056	**0.765**	0.722	0.757	**0.805**	0.735	0.774
	RTx436	0.672	0.690	**0.728**	0.661	0.699	**0.738**
	RTx437	0.656	0.768	**0.798**	0.697	0.776	**0.805**
	Mean	0.668	0.732	0.761	0.707	0.752	**0.785**

Parents		Days to anthesis					
		BRR			GB		
		SM*	MM	MMGE	SM*	MM	MMGE
A-line	A05071	0.694	**0.720**	**0.720**	0.660	0.714	**0.764**
	A08140	0.489	**0.707**	0.705	0.404	**0.701**	0.694
	AARCH11051	0.622	**0.644**	**0.644**	0.600	0.656	**0.760**
	AARCH11129	**0.821**	0.589	0.584	**0.835**	0.596	0.685
	AARCH11136	**0.867**	0.779	0.776	**0.865**	0.777	0.863
	AARCH11146	**0.837**	0.741	0.739	0.781	0.739	**0.808**
	AHF14	0.794	0.802	**0.806**	0.806	0.809	**0.831**
	AKS118	**0.834**	0.776	0.777	**0.864**	0.779	0.839
	ATx3408	0.663	**0.811**	**0.811**	0.704	0.811	**0.849**
	ATx645	0.654	**0.555**	**0.555**	0.562	0.543	**0.617**
	Mean	0.728	0.712	0.712	0.708	0.712	**0.771**
R-line	EON361	**0.760**	0.651	0.730	**0.795**	0.646	0.738
	R07178	0.798	0.813	**0.833**	0.824	0.813	**0.833**
	R08304	0.750	0.769	**0.779**	0.781	0.770	**0.799**
	RARCH11001	0.490	0.679	**0.747**	0.539	0.683	**0.753**
	RARCH11002	0.496	0.690	**0.716**	0.542	0.696	**0.745**
	RARCH11028	0.699	0.611	**0.739**	0.662	0.607	**0.739**
	RARCH11055	0.201	0.459	**0.551**	0.209	0.489	**0.630**
	RARCH11056	**0.919**	0.677	0.754	**0.933**	0.708	0.830
	RTx436	**0.888**	0.616	0.657	**0.877**	0.609	0.673
	RTx437	**0.939**	0.725	0.719	**0.950**	0.727	0.734
	Mean	0.694	0.669	0.723	0.711	0.675	**0.747**

* Predictions for the environment 19CS

As expected, the best predictability of hybrid performance occurred when the grand mean of BLUEs was used to train models (data not shown). This likely happened because the number of observations available to predict the grand average of hybrids increased tenfold compared to predicting hybrid performance for each location. While training models with the grand mean of BLUEs effectively improved predictability, this strategy is problematic when significant GxE effects exist (as in this case) due to changes in the rank of hybrids across environments.

The average predictability of parents for all hybrid traits was assessed within each environment using the SM model (Figure 5). A- and R-lines had similar and generally high predictabilities for PHT and DA across environments (Figure 5). For hybrid GY, predictability varied; A-lines had higher predictability in subtropical environments, whereas the R-lines predicted hybrids more accurately in temperate environments. The exception occurred in 19HAY, where females presented higher prediction accuracy than males.

3.4.4. Predictive ability across parents and target environment using GB-MMGE

The hybrids derived from Texas parents (TxT) were better predicted when genomic prediction models were developed with data from subtropical environments except for PHT in the T2 cross-validation scheme (when males are removed). Similar results occurred when genomic prediction models were developed using phenotypic records derived from temperate environments to predict KxK hybrids. In this case, the accuracy was higher for all traits and cross-validation schemes (Table 11). These observations likely reflect the adaptability and stability of inbred lines bred in the respective target environments. Hybrids derived from lines bred for the target environment benefit from positive covariances existing between the loci and the target environment, which thereby increases the accuracy of genomic prediction models.

Table 11. Genomic-enabled prediction accuracies under kernel-based Bayesian GBLUP models (GB) approach for parents at subtropical and temperate environments using the multi-environment plus GxE effect model (MMGE) for grain yield (GY), plant height (PHT), and days to anthesis (DA). Cross-validation scheme (T1F and T1M) for each set of hybrids (TxT and KxK) were applied for each trait within each subtropical and temperate environment. Higher predictabilities are depicted in boldface.

Hybrids	Parents	Subtropical Env			Temperate Env		
		GY	PHT	DA	GY	PHT	DA
TxT (T1F)	A05071	0.720	0.837	0.852	0.442	0.610	0.625
	A08140	0.387	0.822	0.870	0.489	0.151	0.497
	AHF14	0.589	0.842	0.887	0.276	0.649	0.814
	ATx3408	0.470	0.719	0.831	0.463	0.832	0.913
	ATx645	0.532	0.735	0.831	0.745	0.214	0.533
	Mean	**0.540**	**0.791**	**0.854**	0.483	0.491	0.677
TxT (T1M)	EON361	0.536	0.657	0.859	0.532	0.688	0.742
	R07178	0.451	0.830	0.866	0.165	0.884	0.829
	R08304	0.066	0.798	0.846	0.427	0.836	0.689
	RTx436	0.224	0.899	0.795	-0.076	0.929	0.598
	RTx437	0.294	0.854	0.883	0.374	0.834	0.735
	Mean	**0.314**	0.807	**0.850**	0.285	**0.834**	0.718
KxK (T1F)	AARCH11051	0.388	0.778	0.699	0.451	0.945	0.815
	AARCH11129	0.420	0.847	0.765	0.121	0.840	0.704
	AARCH11136	0.203	0.901	0.913	0.758	0.933	0.933
	AARCH11146	0.326	0.854	0.792	0.363	0.913	0.838
	AKS118	0.521	0.698	0.804	0.394	0.823	0.950
	Mean	0.372	0.816	0.795	**0.417**	**0.891**	**0.848**
KxK (T1M)	RARCH11001	0.588	0.742	0.914	0.779	0.938	0.929
	RARCH11002	0.453	0.887	0.796	0.776	0.940	0.867
	RARCH11028	0.432	0.577	0.630	0.297	0.814	0.824
	RARCH11055	0.241	0.886	0.208	0.465	0.946	0.431
	RARCH11056	0.196	0.836	0.549	0.242	0.883	0.635
	Mean	0.382	0.786	0.619	**0.512**	**0.904**	**0.737**

3.4.5. Impact of genomic and classical GCA-SCA-based models on the coincidence index

Means across environments were used to calculate variance components for hybrid grain yield to compare genomic (GB) and classical GCA-SCA-based (BRR) models (Figure 5). In both models, the majority of the variation is attributed to the environment, but the partition of the variation differed between models. For example, GCA for females and males was lower in GB; alternatively, BRR did not capture GxE (e.g., Env x GCA). Both models showed similar estimates for SCA but the Env x SCA estimate was considerably smaller in BRR than GB. The error component in the GB model was lower than the same in BRR. The generally lower standard deviations in GB are likely the reason for the higher precision of the variance component estimates. Finally, the deviance information criteria (DIC) for BRR was higher than for GB indicating that multi-environment models to predict hybrid performance should include the GRM. The robustness of GB-MMGE increased the CI at any SI so that GB-MMGE was efficient even when SI assumed values less than 10% (Figure 6). The highest CI values (~79%) were obtained when SI assumed values around 9%. After a slight decay in the CI with an increase in SI, CI maintained a value of ~70% for the GB-MMGE model, and ~60% for the BRR-MMGE model.

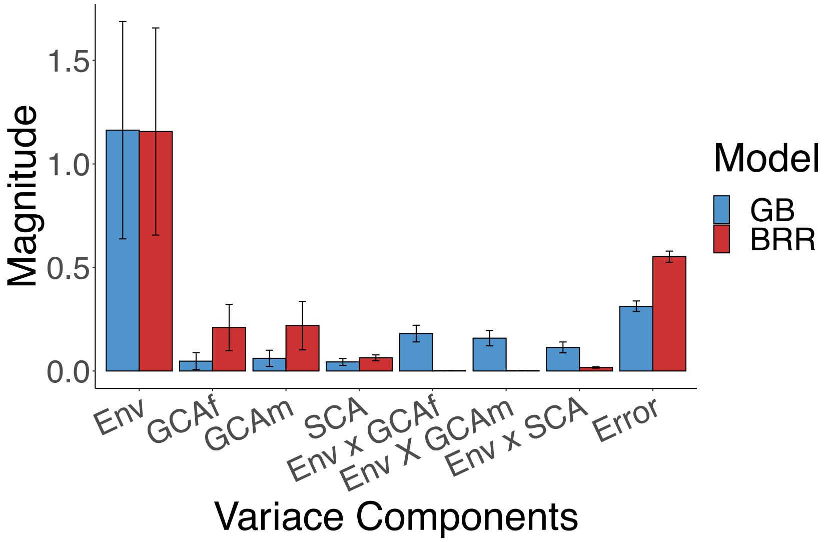

Figure 5. Variance component estimates and standard deviations for grain yield using best linear unbiased estimate (BLUE) under multi-environment genomic (GB) and classical GCA-SCA-based (BRR) models. The graph includes estimates for environmental effect (Env), general combining ability of females (GCAf), general combining ability of males (GCAm), specific combining ability (SCA), interaction effect between environment and combining abilities (Env x GCAf, Env x GCAm, Env x SCA), and error term.

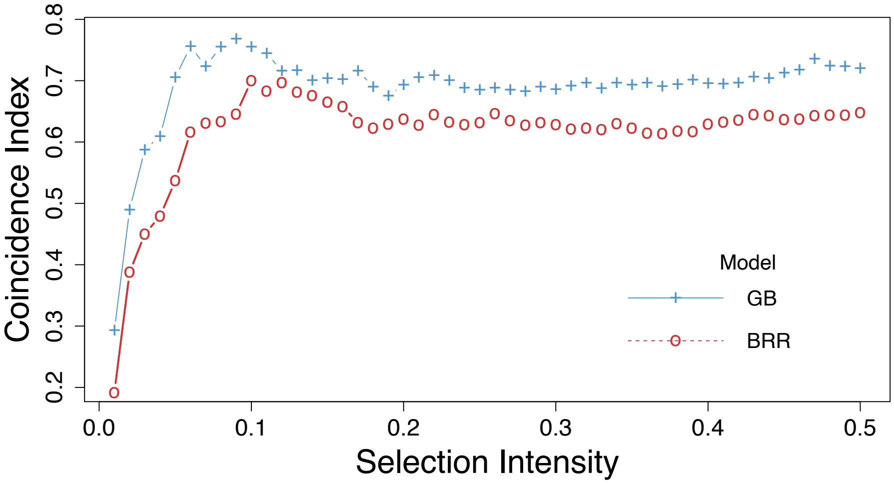

Figure 6. Coincidence index and selection intensity for grain yield using multi-environment genomic (GB) and classical GCA-SCA-based (BRR) models under cross-validation scheme T0MF (both male and female is taken out simultaneously).

3.5. Discussion

3.5.1. Predicting grain sorghum hybrid performance

Bernardo (1994) first described the concept of predicting hybrids with the application of GCA and SCA genetic effects using the Best Linear Unbiased Prediction (BLUP) model in maize. Technow *et al.* (2012) developed a simulation study based on a maize breeding program having complementary heterotic groups to assess genomic prediction of hybrid performance using different models; results supported the application of models based on GCA and SCA to suitably include genetic effects. In the present study, we extend that concept of predicting hybrids using GCA and SCA to a hybrid cereal crop that lacks the extensive resources of maize, and we expanded the prediction models to include genomic information and GxE interaction effects under a kernel-based hierarchical Bayesian framework. Further, the results demonstrate the importance of including GCA and SCA effects to effectively account for the inherent genetic structure that exists in all hybrid breeding populations, including but not limited to sorghum. Both classical GCA-SCA-based and genomic methods successfully predicted grain sorghum hybrid performance and the application of kernel-based Bayesian GBLUP models (GB) increased the prediction ability of the models. While the inclusion of the GxE interaction effects in GB models led to consistently higher prediction accuracies, the same trend was not observed in BRR models.

3.5.2. Advantages of using kernel-based models

The present implementation of kernel-based regression models featured practical advantages relative to the use of standard genomic models in the computational process

(Crossa et al., 2011), which included a significant reduction with issues related to data dimensionality (e.g., when the number of markers considerably exceeds the number of genotypes) (Gianola and de los Campos, 2008). Additionally, kernel-based regression models can be used with nearly every information set (e.g., covariates, strings, images, and graphs) (de los Campos et al., 2010; Crossa et al., 2011). For instance, parametric and semi-parametric kernels accommodate the genomic relationship matrix to represent identical-by-state similarities among parents (de los Campos *et al.*, 2009). This creates a unique opportunity for breeders to account for genetic relatedness among genotypes without tracing genealogy (Morota and Gianola, 2014). Nonetheless, if molecular marker datasets are not available, pedigree information can also be used as a kernel. Models based on marker information manifested higher predictive ability than pedigree-based models in studies involving wheat and maize data (Crossa et al., 2010). Such results exist because pedigree can only explain the expected degree of genetic similarity but it cannot consider segregation distortion. In addition, dense molecular marker datasets can account for the realized degree of genetic similarity, i.e., the mendelian sample. The linear kernel applied herein that included the genetic relationship among divergent elite grain sorghum parents was effective in adjusting the degree of genetic similarity.

The application of the genomic relationship matrix (VanRaden, 2008) via kernel-based framework is also critical in hybrid prediction models because of its subtle ability to include population structure in genomic-enabled prediction models. de los Campos *et al.* (2015) demonstrated that the inclusion of multi-bred or sub-populations is adequate to increase the size of training sets when heterogeneity is modeled. Technow *et al.* (2013) increased the prediction accuracy of models when assessing northern leaf blight resistance in maize by combining data from heterotic groups. The application of kernels appears to be especially relevant for predicting hybrids using GCA and SCA due to the chance of suitably incorporating population structure to each main genetic effect. Herein, kernels representing the genetic (co)variance between inbred parents were applied to each specific genetic effect. Genetic similarities within A- and R-lines were modeled by unique kernels to adjust the GCA of females and males, respectively. Additionally, the SCA effect was generated *in silico* by the Kronecker product between the A- and R-lines relationship matrix (de los Campos et al., 2015). This process adequately accounted for heterogeneity and genetic main effect simultaneously, which increased the prediction accuracy of the models. It is important to note that the predictability greatly benefited from the inclusion of kernels especially when A-lines were involved in the cross-validation scheme (T1F). This is likely due to the higher degree of similarity among A-lines compared to among R-lines. The limited genetic variability present in the A-lines relative to the R-lines reflects the longer timeline and difficulty associated with developing high-performing A-lines. Thus, kernels make it possible to “borrow” information from closely related parents resulting in the higher predictabilities reported herein.

Another advantage of the kernel-based genomic prediction models is the opportunity to include GxE effects using markers. Burgueño *et al.* (2012) introduced a multi-environment version of the standard GBLUP model to account for the genetic correlation between wheat lines and environments, and the authors reported a significant increase in prediction accuracy. The application of reaction norm models within the context of genomic predictions exhibited similar results (Jarquín et al., 2014a). When reaction norm models are applied, kernels usually include GxE effects by the interaction between markers and environmental covariables. Acosta-Pech *et al.* (2017) applied the reaction norm framework using linear kernels to model GCA, SCA, and their interaction with environments; and their results demonstrated the importance of including each of these effects in the genomic prediction of hybrids. Basnet *et al.* (2019) extended a similar model to include environmental covariables to describe sites and results support its practice. Costa-Neto *et al.* (2020) presented nonlinear kernels to incorporate GxE and nonadditive effects in genomic-based prediction in multi-environment trials and suggested its application to predict maize hybrids. In the present study, a linear kernel was applied to model GxE effects using the method presented by Acosta-Pech *et al.* (2017), and higher prediction accuracy was observed. It is important to highlight that the inclusion of GxE effects in genomic prediction models increases prediction accuracy because it can borrow information existing not only among the females and males involved in hybrid combinations, but can also utilize the genetic correlations among environments.

3.5.3. Factors affecting the prediction accuracy in grain sorghum hybrids

In our cross-validation studies, we were able to ascertain additional factors that impact the prediction accuracy of models for hybrid performance traits. Results herein demonstrated that prediction accuracy (often defined as the Pearson correlation between the predicted and observed values) of grain sorghum hybrids is affected by repeatability and genetic architecture of the trait, the degree of genetic similarity among parents, the structure of the training set, the method used to perform predictions (genomic or classical GCA-SCA-based models), and the complexity of models (single or multi-environments). Of these factors, the repeatability of trait measurement is most critical. For instance, the 19HAY environment had lower prediction accuracies for GY and DA; 19HAY also has the lowest repeatabilities for these traits of any environment. Alternatively, in the same environment, PHT had high repeatability and good prediction accuracy. Similar results have been documented in wheat and maize (He et al., 2016; Alves et al., 2019).

Previous studies have also shown the importance of the genetic architecture in genomic-enabled predictions (Hayes et al., 2010; Jia and Jannink, 2012; Daetwyler et al., 2013; Fernandes et al., 2018). Traits such as PHT and DA are controlled by few loci compared to GY (Quinby, 1974; Hilley et al., 2017; Casto et al., 2019). Thus, as expected, the prediction accuracy of these traits was higher than for GY. An additional factor influencing prediction accuracy is the precision of measuring PHT and DY compared to GY (Velazco et al., 2019). Because GY is a highly complex quantitative trait, efforts to increase the precision and repeatability of GY estimates are essential and should be implemented.

As discussed previously, higher prediction ability for the T1F cross-validation scheme (i.e., when females are removed) occurred due to similarity within the female heterotic group. An interesting finding in this study was the positive relation between prediction ability and the structure of the training set – this occurred when the training set involved hybrids tested in environments where the inbred parents were well adapted. Such outcomes are probably due to a natural covariance between the genotypes and their adaptive environmental gradient. The inclusion of environmental covariables to classify sites might capture this covariance and explain such relations (Jarquín et al., 2014a; Costa-Neto et al., 2020). Additionally, classic GCA-SCA-based (BRR) and less complex models (single/multiple environments without GxE) had poorer prediction ability compared to GB-MMGE models due to the aforementioned advantages of kernel-based models. Several studies have reviewed factors influencing the prediction ability of genomic selection in hybrids (Crossa et al., 2014; Zhao et al., 2015; Wang et al., 2018). Among these factors, the size of the training set and its relationship with the validation population is pivotal because genetic relatedness and variation in the elite germplasm is essential. Therefore, a large reference population that accounts for the genetic variation in the breeding program should produce the most accurate genomic prediction models. Finally, high marker density is also important to ensure that most of the genetic variation is being captured (Daetwyler et al., 2013). Although these effects were not directly assessed herein, they should be considered before developing a training set and initiating genomic selection.

3.5.4. Pre-screening hybrids for field evaluation

Genomic selection offers the advantage of predicting hybrid performance of new inbreds without field evaluation; this approach allows for a much larger initial screening of new inbred parents without the costly investment in hybrid seed production and phenotypic evaluation. This combination should improve both the efficiency of crop improvement as well as the rates of genetic gain. Alves *et al.* (2019) reported on the correlation between pre-screening intensity and the proportion of the 5% best maize hybrids for different traits and environments. According to their study, the selection of the top 30% of the predicted hybrids was able to identify approximately 85% of the best evaluated hybrids. Herein with sorghum, predictions using the GB-MMGE model under cross-validation scheme T0MF (both male and female are removed), showed a coincidence index (CI) of almost 80% when less than the top 10% was selected. These results indicate that a significantly smaller subset of all possible hybrid combinations can be pre-screened for subsequent field trials. Thus, breeders would evaluate only the promising hybrids, significantly increasing the efficiency of resource allocation.

A typical application of the CI is to determine the efficiency of models to perform indirect selection, assuming a fixed selection intensity. Fernandes *et al.* (2018) applied the CI to assess the efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum using a diverse panel. Dos Santos *et al.* (2020) used CI to assess the merit of early selection in biomass sorghum and proposed a two-level selection framework to enhance genetic gain per unit of time. Such an approach has proven to be capable of assessing the superiority of models under evaluation. Alternatively, results of this study demonstrate that the CI can identify the optimum selection intensity for a breeding program.

3.5.5. Implementing Genomic Prediction and Selection in sorghum breeding programs

Since the development of sorghum hybrids, sorghum breeding programs have focused on improving the performance of A- and R-lines in a concurrent and complementary mean with the expectation that hybrid combinations between elite inbred lines derived from complementary heterotic groups will maximize heterosis. There is some evidence that this has occurred over the past 60 years (Assefa and Staggenborg, 2010; Gizzi and Gambin, 2016; Pfeiffer et al., 2019).

This strategy has increased the performance of hybrids, but these increments have been modest, and continual modification to the production environments has exacerbated the selection priorities in sorghum breeding programs (Monk et al., 2014; Pfeiffer et al., 2019). Technow *et al.* (2014) proposed a paradigm shift on plant breeding program pipelines designed to develop hybrids using genomic selection to scan the early generation of hybrids and only evaluate promising experimental hybrids at the final stages. In that approach, genomic prediction of hybrid performance allows the program to screen for a larger number of hybrids.

To implement a genomic-enabled prediction model for a hybrid sorghum breeding program, some changes are needed. For instance, programs would need to increase the number of experimental inbreds and the structure of testcrossing to build training models; this testcrossing will be different than the traditional approaches as described by Rooney (2004). The testcross mating scheme would focus more on hybrids *per se* rather than estimating GCA. Fristche-Neto *et al.* (2018) reported that testcross was the least effective mating design to develop training sets to predict maize-single crosses, and recommended using either a full diallel or a factorial, due to the number of hybrid combinations these methods can develop. Herein, a factorial mating design generated sorghum hybrids and the results suggest that such a mating scheme could become a benchmark for producing experimental hybrids within the context of genomic prediction of sorghum hybrid performance.

Gaynor *et al.* (2017) proposed a two-part strategy for the effective use of genomic selection to develop inbred lines in a simulation study; part one focused on identifying new inbreds, and a part two focused on identifying parents for subsequent breeding cycles. In concept, it should be possible to accomplish both objectives simultaneously. Genomic selection can be effective and efficiently implemented in sorghum breeding programs by i) selecting promising hybrids that will participate in future hybrid trials using the models presented previously, while ii) breeding crosses would be designed by combining the information from the genetic relationship derived from the genotyping data plotted in a PCoA, and the GCA of the tested lines (when applicable), or GEBV (using genomic prediction models to estimate breeding values).

3.5.6. Limitations and prospects

The results in this study were based on a full factorial of 20 inbred parents that produced 100 hybrids that were evaluated in ten environments. Further research, including more parents with an incomplete set of hybrid combinations, is needed to confirm the utility of classical GCA-SCA-based and genomic models. Additionally, this study was based on a simple leave-one-out cross-validation scheme, which usually results in higher prediction accuracy due to the relatively larger population size left for the training set. Future studies should consider other cross-validation schemes that remove higher percentages of parents, hybrids and contemplate common challenges faced by plant breeders. Moreover, only linear kernels were used to compute the GRM; the assessment of non-linear kernels must be evaluated for their potential to increase prediction accuracies.

Recently, kernel-based regression models have included high-throughput remote-sensing and phenomics data to perform prediction of hybrids (Cuevas et al., 2019; Lane et al., 2020). Studies on grain sorghum hybrid predictions should incorporate such technique and evaluate their potential benefits. Moreover, multi-trait and reaction norm models using environmental covariables are likely to increase the prediction ability of untested grain sorghum hybrids in untested environments; thus, future research should also include these effects.

3.6. Conclusion

Classical and genomic hierarchical Bayesian models based on GCA, SCA, and their interaction with the environment can effectively predict relevant agronomic traits in untested grain sorghum hybrids; molecular markers can further improve the efficiency of the breeding program pipeline. The inclusion of genomic information in kernel-based GBLUP models (GB) suitably incorporated the natural population structure existent in a hybrid crop breeding scheme and, simultaneously, allocated the genetic similarity to each specific genetic main effect (i.e., general and specific combining abilities). The incorporation of the GxE interaction effect in GB methods allowed models to utilize information existing among the females and males involved in hybrid combinations and to exploit the genetic correlations among environments. The application of the GxE interaction effect also permits breeders to develop strategies to deploy specific hybrids to the targeted environment. These procedures are likely to become the standard method to implement genomic selection in sorghum breeding programs for the reasons presented herein. Finally, the suitability of the GB-MMGE model improves the prediction capacity of grain sorghum hybrids, which permits increasing selection intensity and, ultimately, increasing the rates of genetic gain.

3.7. References

Acosta-Pech, R., J. Crossa, G. de los Campos, S. Teyssèdre, B. Claustres, S. Pérez-Elizalde, and P. Pérez-Rodríguez. 2017. Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. Theor. Appl. Genet. 130:1431–1440. doi:10.1007/s00122-017-2898-0

Alves, F.C., Í.S.C. Granato, G. Galli, D.H. Lyra, R. Fritsche-Neto, and G. de los Campos. 2019. Bayesian analysis and prediction of hybrid performance. Plant Methods 15:14. doi:10.1186/s13007-019-0388-x

Assefa, Y., and S.A. Staggenborg. 2010. Grain sorghum yield with hybrid advancement and changes in agronomic practices from 1957 through 2008. Agron. J. 102:703–706. doi:10.2134/agronj2009.0314

Bandeira e Sousa, M., J. Cuevas, E.G. de O. Couto, P. Pérez-Rodríguez, D. Jarquín, R. Fritsche-Neto, J. Burgueño, and J. Crossa. 2017. Genomic-enabled prediction in maize using kernel models with genotype × environment interaction. G3 Genes, Genomes, Genet. 7:1995–2014. doi:10.1534/g3.117.042341

Basnet, B.R., J. Crossa, S. Dreisigacker, P. Pérez-Rodríguez, Y. Manes, R.P. Singh, U.R. Rosyara, F. Camarillo-Castillo, and M. Murua. 2019. Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. Plant Genome 12:1–13. doi:10.3835/plantgenome2018.07.0051

Bates, D., M. Mächler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67. doi:10.18637/jss.v067.i01

Bernal-Vasquez, A.M., A. Gordillo, M. Schmidt, and H.P. Piepho. 2017. Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. BMC Genet. 18:51. doi:10.1186/s12863-017-0512-8

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. 34:20–25. doi:10.2135/cropsci1994.0011183X003400010003x

Bhat, J.A., S. Ali, R.K. Salgotra, Z.A. Mir, S. Dutta, V. Jadon, A. Tyagi, M. Mushtaq, N. Jain, P.K. Singh, G.P. Singh, and K. V. Prabhu. 2016. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front. Genet. 7:1–11. doi:10.3389/fgene.2016.00221

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52:707–719. doi:10.2135/cropsci2011.06.0299

Casto, A.L., A.J. Mattison, S.N. Olson, M. Thakran, W.L. Rooney, and J.E. Mullet.

2019. Maturity2, a novel regulator of flowering time in Sorghum bicolor, increases expression of SbPRR37 and SbCO in long days delaying flowering. PLoS One 14. doi:10.1371/journal.pone.0212154

Clark, S.A., J.M. Hickey, and J.H. Van Der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol. 43. doi:10.1186/1297-9686-43-18

Comstock, R.E., and H.F. Robinson. 1952. Estimation of average dominance of genes.. Estim. Aver. Domin. genes.

Costa-Neto, G., R. Fritsche-Neto, and J. Crossa. 2020. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity (Edinb).. doi:10.1038/s41437-020-00353-1

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724. doi:10.1534/genetics.110.118521

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett, and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb). 112:48–60. doi:10.1038/hdy.2013.16

Crossa, J., P. Pérez, G. de los Campos, G. Mahuku, S. Dreisigacker, and C. Magorokosho. 2011. Genomic selection and prediction in plant breeding. J. Crop

Improv. 25:239–261. doi:10.1080/15427528.2011.558767

Crozier, D., L. Hoffmann, P.E. Klein, R.R. Klein, and W.L. Rooney. 2020. Predicting heterosis in grain sorghum hybrids using sequence-based genetic similarity estimates. J. Crop Improv.. doi:10.1080/15427528.2020.1748152

Cuevas, J., J. Crossa, O.A. Montesinos-López, J. Burgueño, P. Pérez-Rodríguez, and G. de los Campos. 2017. Bayesian genomic prediction with genotype × environment interaction kernel models. G3 Genes, Genomes, Genet. 7:41–53. doi:10.1534/g3.116.035584

Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez, G. de los Campos, O.A. Montesinos-López, and J. Burgueño. 2016. Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. Plant Genome 9:1–20. doi:10.3835/plantgenome2016.03.0024

Cuevas, J., O. Montesinos-López, P. Juliana, C. Guzmán, P. Pérez-Rodríguez, J. González-Bucio, J. Burgueño, A. Montesinos-López, and J. Crossa. 2019. Deep Kernel for genomic and near infrared predictions in multi-environment breeding trials. G3 Genes, Genomes, Genet. 9:2913–2924. doi:10.1534/g3.119.400493

Daetwyler, H.D., M.P.L. Calus, R. Pong-Wong, G. de los Campos, and J.M. Hickey. 2013. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. Genetics 193:347–365. doi:10.1534/genetics.112.147983

Dias, K.O.D.G., S.A. Gezan, C.T. Guimarães, A. Nazarian, L. Da Costa E Silva, S.N. Parentoni, P.E. De Oliveira Guimarães, C. De Oliveira Anoni, J.M.V. Pádua, M. De

Oliveira Pinto, R.W. Noda, C.A.G. Ribeiro, J.V. De Magalhães, A.A.F. Garcia, J.C. De Souza, L.J.M. Guimarães, and M.M. Pastina. 2018. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity (Edinb). 121:24–37. doi:10.1038/s41437-018-0053-6

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular markers in a commercial breeding program. Crop Sci. 47:S-154-S-163. doi:10.2135/cropsci2007.04.0015IPBS

Fernandes, S.B., K.O.G. Dias, D.F. Ferreira, and P.J. Brown. 2018. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor. Appl. Genet. 131:747–755. doi:10.1007/s00122-017-3033-y

Fristche-Neto, R., D. Akdemir, and J.L. Jannink. 2018. Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. Theor. Appl. Genet. 131:1153–1162. doi:10.1007/s00122-018-3068-8

Gaynor, R.C., G. Gorjanc, A.R. Bentley, E.S. Ober, P. Howell, R. Jackson, I.J. Mackay, and J.M. Hickey. 2017. A two-part strategy for using genomic selection to develop inbred lines. Crop Sci. 57:2372–2386. doi:10.2135/cropsci2016.09.0742

Gianola, D., and G. de los Campos. 2008. Inferring genetic values for quantitative traits non-parametrically. Genet. Res. (Camb). 90:525–540. doi:10.1017/S0016672308009890

Gianola, D., G. De Los Campos, W.G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183:347–363.

doi:10.1534/genetics.109.103952

Gizzi, G., and B.L. Gambin. 2016. Eco-physiological changes in sorghum hybrids released in Argentina over the last 30 years. F. Crop. Res. 188:41–49. doi:10.1016/j.fcr.2016.01.010

Habier, D., R.L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12. doi:10.1186/1471-2105-12-186

Hamblin, J., and M.J. de O. Zimmermann. 1986. Breeding common bean for yield in mixtures. Plant Breed. Rev. 4:245–272. doi:10.1002/9781118061015.ch8

Hayes, B.J., J. Pryce, A.J. Chamberlain, P.J. Bowman, and M.E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet 6:1001139. doi:10.1371/journal.pgen.1001139

He, S., A.W. Schulthess, V. Mirdita, Y. Zhao, V. Korzun, R. Bothe, E. Ebmeyer, J.C. Reif, and Y. Jiang. 2016. Genomic selection in a commercial winter wheat population. Theor. Appl. Genet. 129:641–651. doi:10.1007/s00122-015-2655-1

Heffner, E.L., A.J. Lorenz, J.-L.L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681–1690. doi:10.2135/cropsci2009.11.0662

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Henderson, C.R. 1984. Applications of Linear Models in Animal Breeding

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. doi:10.2135/cropsci2011.06.0297

Hilley, J.L., B.D. Weers, S.K. Truong, R.F. McCormick, A.J. Mattison, B.A. McKinley, D.T. Morishige, and J.E. Mullet. 2017. Sorghum Dw2 encodes a protein kinase regulator of stem internode length. Sci. Rep. 7:1–13. doi:10.1038/s41598-017-04609-5

Hunt, C.H., F.A. van Eeuwijk, E.S. Mace, B.J. Hayes, and D.R. Jordan. 2018. Development of genomic prediction in sorghum. Crop Sci. 58:690. doi:10.2135/cropsci2017.08.0469

Islam, M.S., D.D. Fang, J.N. Jenkins, J. Guo, J.C. McCarty, and D.C. Jones. 2020. Evaluation of genomic selection methods for predicting fiber quality traits in Upland cotton. Mol. Genet. Genomics 295:67–79. doi:10.1007/s00438-019-01599-z

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9:166–177. doi:10.1093/bfgp/elq001

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez , M. Calus, and J. Burgueño. 2014a. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127:595–607. doi:10.1007/s00122-013-2243-1

Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz.

2014b. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15:740. doi:10.1186/1471-2164-15-740

Jia, Y., and J.L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. doi:10.1534/genetics.112.144246

Kamvar, Z.N., and N.J. Grünwald. 2020. Algorithms and equations utilized in poppr version 2.8.6

Kofoid, K.D., and T.L. Harvey. 2005. Registration of Greenbug Resistant Sorghum Germplasm Lines KS 116 A/B through KS 120 A/B. Crop Sci. 45:802. doi:10.2135/cropsci2005.0802

Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82. doi:10.18637/jss.v082.i13

Lane, H.M., S.C. Murray, O.A. Montesinos-López, A. Montesinos-López, J. Crossa, D.K. Rooney, I.D. Barrero-Farfan, G.N. De La Fuente, and C.L.S. Morgan. 2020. Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. Plant Phenome J. 3:1–19. doi:10.1002/ppj2.20002

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.L. Jannink, R.P. Singh, E. Autrique, and G. de los Campos. 2015. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 Genes, Genomes, Genet. 5:569–582. doi:10.1534/g3.114.016097

de Los Campos, G., D. Gianola, and G.J. Rosa. 2009. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation.. J. Anim. Sci. 87:1883–1887. doi:10.2527/jas.2008-1259

de los Campos, G., D. Gianola, G.J.M. Rosa, K.A. Weigel, and J. Crossa. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet. Res. (Camb). 92:295–308. doi:10.1017/S0016672310000285

de los Campos, G., Y. Veturi, A.I. Vazquez, C. Lehermeier, and P. Pérez-Rodríguez. 2015. Incorporating genetic heterogeneity in whole-genome regressions using interactions. J. Agric. Biol. Environ. Stat. 20:467–490. doi:10.1007/s13253-015-0222-5

Mangin, B., F. Bonnafous, N. Blanchet, M.C. Boniface, E. Bret-Mestries, S. Carrère, L. Cottret, L. Legrand, G. Marage, P. Pegot-Espagnet, S. Munos, N. Pouilly, F. Vear, P. Vincourt, and N.B. Langlade. 2017. Genomic prediction of sunflower hybrids oil content. Front. Plant Sci. 8:1633. doi:10.3389/fpls.2017.01633

Mbulwe, L., G.C. Peterson, J. Scott-Armstrong, and W.L. Rooney. 2015. Registration of sorghum germplasm Tx3408 and Tx3409 with tolerance to sugarcane aphid [*Melanaphis sacchari* (Zehntner)]. J. Plant Regist. 10:51–56. doi:10.3198/jpr2015.04.0025crg

Menz, M.A., R.R. Klein, N.C. Unruh, W.L. Rooney, P.E. Klein, and J.E. Mullet. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44:1236–1244. doi:10.2135/cropsci2004.1236

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

Miller, F.R., T.F. Dusek, K.L. Prihoda, and L.W. Rooney. 1992. Registration of RTx436 sorghum parental line. Crop Sci. 32:1518–1518. doi:10.2135/cropsci1992.0011183x003200060059x

Monk, R., C. Franks, and J. Dahlberg. 2014. Sorghum. Wiley Online Library.

Morishige, D.T., P.E. Klein, J.L. Hilley, S.M. Sahraeian, A. Sharma, and J.E. Mullet. 2013. Digital genotyping of sorghum – a diverse plant species with a large repeat-rich genome. BMC Genomics 14:448. doi:10.1186/1471-2164-14-448

Morota, G., and D. Gianola. 2014. Kernel-based whole-genome prediction of complex traits: a review. Front. Genet. 5:1–13. doi:10.3389/fgene.2014.00363

Nadeem, M.A., M.A. Nawaz, M.Q. Shahid, Y. Doğan, G. Comertpay, M. Yıldız, R. Hatipoğlu, F. Ahmad, A. Alsaleh, N. Labhane, H. Özkan, G. Chung, and F.S. Baloch. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol. Biotechnol. Equip. 32:261–285. doi:10.1080/13102818.2017.1400401

de Oliveira, A.A., M.M. Pastina, V.F. de Souza, R.A. da Costa Parrella, R.W. Noda, M.L.F. Simeone, R.E. Schaffert, J.V. de Magalhães, C.M.B. Damasceno, and G.R.A. Margarido. 2018. Genomic prediction applied to high-biomass sorghum for bioenergy production. Mol. Breed. 38:49. doi:10.1007/s11032-018-0802-5

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554. doi:10.1093/biomet/58.3.545

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the

BGLR statistical package. Genetics 198:483–495. doi:10.1534/genetics.114.164442

Perumal, R., T. Tesso, K.D. Kofoid, R.M. Aiken, P.V. V. Prasad, S.R. Bean, J.D. Wilson, T.J. Herald, and C.R. Little. 2019. Registration of six grain sorghum pollinator (R) lines. J. Plant Regist. 13:113–117. doi:10.3198/jpr2017.12.0087crp

Pfeiffer, B.K., D. Pietsch, R.W. Schnell, and W.L. Rooney. 2019. Long-term selection in hybrid sorghum breeding programs. Crop Sci. 59:150–164. doi:10.2135/cropsci2018.05.0345

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. doi:10.1007/s10681-007-9449-8

Prevosti, A., J. Ocaña, and G. Alonso. 1975. Distances between populations of Drosophila subobscura, based on chromosome arrangement frequencies. Theor. Appl. Genet. 45:231–241. doi:10.1007/BF00831894

Quinby, J.R. 1974. Sorghum Improvement and the Genetics of Growth. Texas agricultural experiment station.

R Core, T. 2020. R: A language and environment for statistical computing

Ramstein, G.P., and M.D. Casler. 2019. Extensions of BLUP models for genomic prediction in heterogeneous populations: Application in a diverse switchgrass sample. G3 Genes, Genomes, Genet. 9:789–805. doi:10.1534/g3.118.200969

Rooney, W. 2004. Sorghum improvement – integrating traditional and new technology to produce improved genotypes. New York: Academic Press, 1949-.

Rooney, W.L., F.R. Miller, and L.W. Rooney. 2003. Registration of RTx437 sorghum parental line. Crop Sci. 43:445. doi:10.2135/cropsci2003.4450

dos Santos, J.P.R., S.B. Fernandes, S. McCoy, R. Lozano, P.J. Brown, A.D.B. Leakey, E.S. Buckler, A.A.F. Garcia, and M.A. Gore. 2020. Novel bayesian networks for genomic prediction of developmental traits in biomass sorghum. G3 Genes, Genomes, Genet. 10:769–781. doi:10.1534/g3.119.400759

Sapkota, S., J.L. Boatwright, K. Jordan, R. Boyles, and S. Kresovich. 2020. Multi-trait regressor stacking increased genomic prediction accuracy of sorghum grain composition. bioRxiv. doi:10.1101/2020.04.03.023531

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644. doi:10.1086/502802

Stewart-Brown, B.B., Q. Song, J.N. Vaughn, and Z. Li. 2019. Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3 Genes, Genomes, Genet. 9:2253–2265. doi:10.1534/g3.118.200917

Struyf, A., M. Hubert, and P.J. Rousseeuw. 1997. Clustering in an object-oriented environment. J. Stat. Softw. 1:1–30

Sukumaran, S., D. Jarquin, J. Crossa, and M. Reynolds. 2018. Genomic-enabled prediction accuracies increased by modeling genotype × environment interaction in Durum wheat. Plant Genome 11:170112. doi:10.3835/plantgenome2017.12.0112

Tan, B., D. Grattapaglia, G.S. Martins, K.Z. Ferreira, B. Sundberg, and P.K. Ingvarsson. 2017. Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. BMC Plant Biol. 17.

doi:10.1186/s12870-017-1059-6

Tang, Y., M. Horikoshi, and W. Li. 2016. Ggfortify: Unified interface to visualize statistical results of popular r packages. R J. 8:478–489. doi:10.32614/rj-2016-060

Technow, F., A. Bürger, and A.E. Melchinger. 2013. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 Genes, Genomes, Genet. 3:197–203. doi:10.1534/g3.112.004630

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. doi:10.1007/s00122-012-1905-8

Technow, F., T.A. Schrag, W. Schipprack, E. Bauer, H. Simianer, and A.E. Melchinger. 2014. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343–1355. doi:10.1534/genetics.114.165860

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Velazco, J.G., D.R. Jordan, E.S. Mace, C.H. Hunt, M. Malosetti, and F.A. van Eeuwijk. 2019. Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. Front. Plant Sci. 10:1–12. doi:10.3389/fpls.2019.00997

Wang, X., L. Li, Z. Yang, X. Zheng, S. Yu, C. Xu, and Z. Hu. 2017. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity (Edinb). 118:302–310. doi:10.1038/hdy.2016.87

Wang, X., Y. Xu, Z. Hu, and C. Xu. 2018. Genomic selection methods for crop improvement: Current status and prospects. Crop J. 6:330–340. doi:10.1016/j.cj.2018.03.001

Wickham, H. 2016. Ggplot2: Elegant Graphics for Data Analysis. springer.

Xu, S., D. Zhu, and Q. Zhang. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. U. S. A. 111:12456–12461. doi:10.1073/pnas.1413750111

Xu, Y., X. Wang, X. Ding, X. Zheng, Z. Yang, C. Xu, and Z. Hu. 2018. Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11. doi:10.1186/s12284-018-0223-4

Zhao, Y., M.F. Mette, and J.C. Reif. 2015. Genomic selection in hybrid breeding. Plant Breed. 134:1–10. doi:10.1111/pbr.12231

Zhao, Y., J. Zeng, R. Fernando, and J.C. Reif. 2013. Genomic prediction of hybrid wheat performance. Crop Sci. 53:802–810. doi:10.2135/cropsci2012.08.0463

4. MEGA-ENVIRONMENT ANALYSIS TO ASSESS ADAPTABILITY, STABILITY, AND GENOMIC PREDICTIONS IN GRAIN SORGHUM HYBRIDS

4.1. Synopsis

Multi-environment trials (MET) are fundamental for assessing genotype-by-environment interaction (GxE) effects, adaptability and stability of genotypes and provide valuable information about target regions. As such, a MET involving grain sorghum hybrid combinations derived from elite inbred lines adapted to diverse sorghum production regions was developed to assess agronomic performance, stability, and genomic-enabled prediction accuracies within mega-environments (ME). Ten females and ten males from the Texas A&M and Kansas State sorghum breeding programs were crossed following a factorial mating scheme to generate 100 hybrids. Grain yield, plant height, and days to anthesis were assessed in a MET consisting of ten environments across Texas and Kansas locations over two years. Genotype plus Genotype-by-block-of-environment biplot assessed ME, while the "mean-vs-stability" view of the biplot and the Bayesian Finlay-Wilkinson regression evaluated the adaptability and stability of hybrids. A genomic prediction model considering the GxE effect was applied within ME to assess prediction accuracy. Results suggest that grain sorghum hybrid combinations involving lines adapted to different target regions can produce superior hybrids. ME analysis identified established grain sorghum production regions in the U.S. Further, genomic predictions within ME reported inconsistent results, suggesting that additional effects rather than the correlations between environments are influencing genomic prediction of grain sorghum hybrids.

4.2. Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is the fifth most important cereal crop in the world, following corn (*Zea mays* L.), wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), and barley (*Hordeum vulgare* L.) with 57.8 million tons of global annual production in 2019 (FAOSTAT, 2021). While Africa is responsible for almost 50% of the total grain sorghum production, sorghum yield is significantly higher in developed countries, where production technology is not a limiting factor. For example, the average sorghum yield in 2019 was 4.58- and 1.23-tons ha^{-1} in the U.S. and Nigeria, respectively (FAOSTAT, 2021). Among the U.S. states, Kansas and Texas account for 82% of total U.S. production, with average yields of 5.72- and 5.38-tons ha^{-1}, respectively (NASS, 2020).

Grain sorghum production is predominantly cultivated in agricultural areas where water is limited (Monk et al., 2014). Sorghum is favored in these regions because the species possesses and uses several drought-tolerance mechanisms, including stay-green and pre-flowering drought-tolerance (Rooney, 2016). This drought tolerance results in greater yield stability compared to other crops such as maize (*Zea mays*) that are less able to perform under water-stress conditions.

Because sorghum is produced in more stress environments, developing improved grain sorghum genotypes requires special consideration about traits that affect yield, including adaptability and stability. For instance, sorghum hybrids that are highly responsive to distinct locations may be appropriate for those non-limiting environments, which correspond to an atypical sorghum production region. Hence, assessing the stability of grain sorghum hybrids across target environments before releasing new products becomes particularly important for the effective deployment of sorghum hybrids.

Multi-environment trials (MET) are commonly used to assess genotype-by-environment interaction (GxE) effects. When MET detect GxE that do not involve crossover interaction (COI), GxE may be ignored since significant genotype rank changes are unlikely to occur. However, GxE is often associated with COI, and judicious product placement requires breeders to investigate the adaptability and stability of genotypes (Crossa et al., 2004; Burgueño et al., 2008). Many stability tests contemplate univariate measures of stability (Wricke, 1962; Finlay and Wilkinson, 1963; Shukla, 1972), while others have proposed the use of multivariate measures based on biplot technique (Gabriel, 1971) (i.e., additive main effects and multiplicative interaction, AMMI, Gauch and Zobel, 1997; genotype + GxE, GGE, Yan et al., 2000), and nonparametric measures (Nassar and Huhn, 1987; Huehn, 1990). Depending on the test, stability can be classified as static and dynamic (Becker, 1981; Becker and Leon, 1988).

As MET consist of a continuous series of testing genotypes across target environments, not all genotypes are present in every environment. As such, MET are usually unbalanced, challenging the ability to assess the stability of a genotype. A common practice to circumvent unbalanced data and assess stability is to remove genotypes that are missing in some environments and perform stability tests via classical models that assume genotype and environment effects as fixed (Finlay and Wilkinson, 1963; Gauch, 1988; Yan et al., 2000). A better alternative to manage unbalanced data involves applying linear mixed models that accommodate random effects and include (co)variance structure to model genetic and environmental effects.

Among the models that assume effects as random, factor analytics models are commonly applied to assess MET (Piepho, 1994; Smith et al., 2001, 2005; Piepho et al., 2008). Linear mixed models allow better prediction of those missing data by incorporating the genetic relationship matrix of tested and untested genotypes (Jarquin et al., 2020). Many studies have reported on the benefits of using linear mixed models to assess MET (Piepho, 1997, 1998; Smith et al., 2001, 2019; Resende and Thompson, 2004; Piepho and Möhring, 2005; Burgueño et al., 2008; Balestre et al., 2010; Hu, 2015; Jarquin et al., 2020). Bayesian models have also demonstrated their advantages for analyzing genotypic data (Sorensen and Gianola, 2007; Crossa et al., 2010; Alves et al., 2019). Lian and de los Campos (2016) developed a Finlay and Wilkinson stability test under a Bayesian framework, and studies in grain sorghum hybrids using such a technique are lacking.

In addition to the inconsistency of genotypes in the MET, different evaluation stages can also affect the number of genotypes and environments involved in an experiment. Many genotypes are tested in only a few environments at initial stages, whereas advanced stages evaluate a few genotypes across many environments. Although generally applied, such a breeding pipeline is under adjustments to accommodate predictive breeding technologies (Crossa et al., 2021). For instance, genomic-enabled prediction is likely to replace the phase I hybrid trials in grain sorghum breeding schemes (Fonseca et al., 2021a). Further, MET allow the development of genomic-enabled prediction models to include GxE, and results indicate better prediction accuracy (Burgueño et al., 2012; Lopez-Cruz et al., 2015; Lado et al., 2016; Cuevas et al., 2016; Acosta-Pech et al., 2017; Basnet et al., 2019; Costa-Neto et al., 2020).

Although MET primarily focus on identifying superior genotypes for target environments, mature breeding programs also apply MET to generate relevant information about tested environments (Gauch and Zobel, 1997; Yan et al., 2000; Yan and Kang, 2002; Laffont et al., 2013; Gauch, 2013; Yan, 2015, 2016). This analysis can be useful to identify mega-environments (ME) across a wider production area (Rakshit et al., 2012; Nielsen and Vigil, 2018; Dalló et al., 2019; Sharma et al., 2020; Ansarifard et al., 2020). By definition, ME reduce GxE and allow breeders to extract genotypic information about genotypes more effectively within a ME (González-Barrios et al., 2019). Lado et al. (2016) indicated that wheat line performance was more predictable when ME information is considered. Alves et al. (2021) also showed that the prediction accuracy of hybrids in low correlated environments can be increased when including additional phenotypic information from other positive correlated trials.

Lopez-Cruz et al. (2015) presented a marker-by-environment genomic best linear unbiased prediction (MxE GBLUP) model to account for GxE effects in genomic-enabled prediction models. The MxE model borrows genetic information across environments and increases prediction accuracies for untested genotypes in tested environments. Acosta-Pech et al. (2017) further expanded the MxE GBLUP model to include combining abilities, environment, and combining abilities-by-environment interactions and reported higher prediction accuracies in maize. Fonseca et al. (2021a) then applied the Acosta-Pech approach to sorghum with similar results. Most genomic-enabled prediction models that include GxE effects assume the GxE (co)variance matrix to be positive semi-definite (Cuevas et al., 2016; Crossa et al., 2019). Thus, higher correlations between environments are expected to yield higher prediction accuracies. The combination of the Acosta-Pech model under ME may generate good prediction accuracies as ME are expected to increase the correlation between environments, causing a reduction in the residual variance, and potentially capturing more of the genetic variation.

A common breeding approach for multinational breeding programs is to cross elite germplasm adapted to diverse target environments (Cooper et al., 2014). This strategy is not as common in public institutions (Fonseca et al., 2021b), likely due to restrictions on germplasm exchange and that state-based breeding programs function primarily in state. These programs could benefit from this exchange of germplasm since simulation studies have demonstrated the potential advantages of utilizing this strategy in hybrid crops (Podlich and Cooper, 1998; Technow et al., 2020).

The development of hybrid combinations between elite inbred lines derived from distinct breeding programs fosters a collaborative system between public breeding programs and mitigates the decrease these institutions are experiencing (Shelton and Tracy, 2017; Coe et al., 2020; Fonseca et al., 2021b). This approach could also generate agronomically competitive grain sorghum hybrids. Therefore, this research aims to develop ME for grain sorghum production regions and assess the adaptability and stability of grain sorghum hybrids derived from U.S. public breeding programs within ME. Further, it evaluates genomic-enabled prediction accuracies to predict the performance of grain sorghum hybrids using GxE models within ME as a strategy to explore GxE effects among closely related environments. Comparison between genomic predictions based on ME and without ME are presented.

4.3. Material and methods

4.3.1. Genetic material

Grain sorghum hybrids described by Fonseca et al. (2021a) were used in this research. These 100 hybrids were generated from crossing ten female with ten male lines in a factorial mating scheme (Comstock and Robinson, 1952). The inbreds selected for this study represent elite germplasm from the sorghum breeding programs at Texas A&M AgriLife Research (College Station, Texas) and Kansas State Research Center (Hays, Kansas). Each program provided five females (A-lines) and five males (R-lines) that are adapted to their respective target environments and had produced agronomically acceptable hybrids previously (Table 1). The 100 hybrids were subdivided into four groups to represent hybrid combinations derived from elite lines within a breeding programs and between breeding programs (Figure 1).

4.3.2. Experimental data

Hybrids were planted in a randomized complete block design (RCBD) with sets in replicate adjustment; each set was composed of one of the four groups of 25 hybrids (Figure 1), with randomization occurring within and between sets. In 2018 trials, each location had three replicates, while in 2019 each location had two replicates. A plot consisted of two adjacent rows, approximately 5.3 m in length, with row spacing that ranged from 0.76 to 1.0 m, depending on the production practices in each environment. In 2018, trials were grown in Monte Alto – TX, Victoria – TX, and College Station – TX, and Garden City – KS and Colby – KS. In 2019, the trials were in Taft – TX, Victoria – TX, College Station – TX, Hays – KS, and Colby – Kansas. These Texas and Kansas locations represent distinct adaptation zones and, in each test, agronomic practices standard to the location were followed (Table 2) (Fonseca et al., 2021a). GPS coordinates of each environment were used to collect weather data from the NASA POWER database (NASA, 2021).

Three agronomic traits were measured. Days to anthesis (DA) were recorded as the number of days from planting to when 50% of the plants in the plot were at mid-anthesis. In Kansas locations, DA was collected on one replicate only. Just prior to harvest, plant height (PHT) was measured in cm using a representative plant in each plot at the length from the soil to the tip of the panicle. Plots were harvested using plot combines fitted with a Harvest Master GrainGage System (Juniper Systems), to measure grain weight and moisture content. After adjusting yield to a 14% moisture, grain yield (GY) was converted to tons ha^{-1} (Fonseca et al., 2021a).

4.3.3. Statistical analysis

For each environment, data were analyzed following a linear mixed model (Henderson, 1984):

$$\boldsymbol{y} = \boldsymbol{1\mu} + \boldsymbol{Z_1 h} + \boldsymbol{Z_2 r} + \boldsymbol{e}, \tag{3}$$

where $\boldsymbol{y}$ is the vector of phenotypes; $\boldsymbol{\mu}$ is an intercept; $\boldsymbol{h}$ is a random effect of hybrid, $\boldsymbol{h} \sim N(\mathbf{0}, \sigma_h^2 \boldsymbol{I})$, $\boldsymbol{r}$ is a random effect of replicates, $\boldsymbol{r} \sim N(\mathbf{0}, \sigma_r^2 \boldsymbol{I})$; $\boldsymbol{e}$ is the vector of residuals, $\boldsymbol{e} \sim N(\mathbf{0}, \sigma_e^2 \boldsymbol{I})$; $\mathbf{1}$ is a vector of ones; $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are incidence matrixes; σ_h^2, σ_r^2 and σ_e^2 are variance components for hybrids, replicates and residuals, respectively. Model (1) was extended to incorporate environment and GxE effects for performing a combined analysis as followed:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Z_1 h} + \boldsymbol{Z_2 s} + \boldsymbol{Z_3 hs} + \boldsymbol{Z_4 r(s)} + \boldsymbol{e}, \qquad (2)$$

where **s** is a vector of environmental effects, $\boldsymbol{s} \sim N(\mathbf{0}, \sigma_s^2 \boldsymbol{I})$; $\boldsymbol{hs}$ is a vector of the GxE effect $\boldsymbol{hs} \sim N(\mathbf{0}, \sigma_{hs}^2 \boldsymbol{I})$, **r(s)** is the vector of replicate effect nested within environment, $\boldsymbol{r(s)} \sim N(\mathbf{0}, \sigma_{r(s)}^2 \boldsymbol{I})$; $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Z_3}$, $\boldsymbol{Z_4}$, are incidence matrixes for the corresponding effect.

Variance components were estimated via restricted maximum likelihood (REML) method (Patterson and Thompson, 1971) using the *lmer* function of the *lme4* R package (Bates et al., 2015), and its significance assessed by the likelihood ratio test (LRT) using the *ranova* function of the *lmerTest* R package (Kuznetsova et al., 2017). All analysis were conducted in R software (R Core Team 2020). From the variance component estimates, broad-sense heritability at each environment was calculated as $R = \frac{\sigma_h^2}{\sigma_h^2 + \frac{\sigma_e^2}{r}}$, whereas, for combined environments, $H^2 = \frac{\sigma_h^2}{\sigma_h^2 + \frac{\sigma_{hs}^2}{t} + \frac{\sigma_e^2}{tr}}$, where *t* is the number of environments and *r* is the number of replicates. Best linear unbiased estimates (BLUEs) of each hybrid at each location were obtained using the R package *emmeans* (Lenth 2021) from model (1) assuming hybrids as fixed. BLUEs were used to develop the training set in genomic prediction models.

Mega-environments (ME) were developed using the *gge* function of the *gge* R package (Wright and Laffont, 2020). Genotype plus genotype × environment (GGE) biplot and which-won-where polygon were used to assess grain sorghum hybrid performance, while genotype plus genotype × block of environments (GGB) biplots were used to define ME. Biplots were scaled and environment-centered. For GGE biplot, stability was assessed using loading projections on the average environment coordinate (AEC) at each ME. For Bayesian Finlay-Wilkinson regression (BFW - Lian and De Los Campos, 2016), static and dynamic stability was assessed based on the slope of genotypes across environment. Regression lines presenting slopes close to 0 classified static stability while slopes close to 1 defined dynamic stability. For BFW, genomic information of hybrids was included in the model as (co)variance matrix of the genotypic effect. Posterior distribution was calculated based on 10,000 interactions, burn-in of 5,000 and a thin of 5. Genomic relationship of hybrids were developed *in silico* by the Kronecker product of the parental marker matrix. For details see Fonseca et al. (2021a).

4.3.4. Genotypic data

DNA extraction, genotyping-by-sequencing (GBS), and development of molecular markers were described in Fonseca et al. (2021a).

4.3.5. Genomic prediction model and cross validation scheme

Genomic-enabled prediction model used in this study was first introduced by Acosta-Pech et al. (2017) and applied in grain sorghum hybrids by Fonseca et al. (2021a). Herein, the model was applied for each ME separately and included combining abilities, environment and GxE effects to predict the performance of sorghum hybrids as follows:

$$\boldsymbol{y} = \boldsymbol{Z_s\beta} + \boldsymbol{Z_1 f} + \boldsymbol{Z_2 m} + \boldsymbol{Z_3 h} + \boldsymbol{u_f} + \boldsymbol{u_m} + \boldsymbol{u_h} + \boldsymbol{e}, \tag{3}$$

where $\boldsymbol{y} = [\boldsymbol{y_1}, \cdots, \boldsymbol{y_j}, \cdots, \boldsymbol{y_s}]'$ is the vector of observations of hybrids ($i = 1, \cdots, \mathrm{n}_j$) in the j$^{\text{th}}$ environment ($j = 1, \cdots, \mathrm{s}$); $\boldsymbol{\beta}$ is a vector of the environmental effects, $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I})$, $\boldsymbol{f}$ is the vector of GCA effects for females, $\boldsymbol{f} \sim N(\boldsymbol{0}, \sigma_{f_j}^2 \boldsymbol{G}_f)$, where $\boldsymbol{j}$ represents environment; $\boldsymbol{m}$ is a vector of GCA effects for males, $\boldsymbol{m} \sim N(\boldsymbol{0}, \sigma_{m_j}^2 \boldsymbol{G}_m)$; $\boldsymbol{h}$ is a vector of SCA effects for hybrid combinations, $\boldsymbol{h} \sim N(0, \sigma_{h_j}^2 \boldsymbol{H})$; $\boldsymbol{u}_f$ is a random vector to include the interaction between GCA of females and environment, $\boldsymbol{u}_f \sim N(\boldsymbol{0}, \sigma_{fs}^2 \boldsymbol{V}_f)$; $\boldsymbol{u}_m$ is a random vector that models the interaction between males and environment, $\boldsymbol{u}_m \sim N(\boldsymbol{0}, \sigma_{ms}^2 \boldsymbol{V}_m)$; u_h is a vector that takes into account the interaction between hybrids and environment, $\boldsymbol{u}_h \sim N(\boldsymbol{0}, \sigma_{hs}^2 \boldsymbol{V}_h)$; $\boldsymbol{e}$ is a vector of residuals, $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma_{e_j}^2 \boldsymbol{I})$; $\boldsymbol{Z_s}$, $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, and $\boldsymbol{Z}_3$ are incidence matrixes for environment, female, male, and hybrid, respectively. $\boldsymbol{G}_f$, $\boldsymbol{G}_m$, and $\boldsymbol{H}$ are the genomic relationship of matrix for females, males, and hybrids, respectively. The elements of matrix $\boldsymbol{H}$ were calculated *in silico* by the Kronecker product of $\boldsymbol{G}_f$ and $\boldsymbol{G}_m$ as $\boldsymbol{H} = \boldsymbol{G}_f \otimes \boldsymbol{G}_m$ (Technow *et al.*, 2014; Acosta-Pech *et al.*, 2017). $\boldsymbol{V}_f$, $\boldsymbol{V}_m$, and $\boldsymbol{V}_H$ are variance-covariance matrices that associate each respective genomic-main-effect-by-environment variance component, i.e., σ_{fs}^2, σ_{ms}^2, and σ_{hs}^2. The variance-covariance matrix was calculated as $V_f = \boldsymbol{Z}_1 \boldsymbol{G}_f \boldsymbol{Z'}_1 \# \boldsymbol{Z}_s \boldsymbol{Z'}_s$, $\boldsymbol{V}_m = \boldsymbol{Z}_2 \boldsymbol{G}_m \boldsymbol{Z'}_2 \# \boldsymbol{Z}_s \boldsymbol{Z'}_s$, and $\boldsymbol{V}_h = \boldsymbol{Z}_3 \boldsymbol{H} \boldsymbol{Z'}_3 \# \boldsymbol{Z}_s \boldsymbol{Z'}_s$ where # stands for the Hadamard product (see Acosta-Pech *et al.* (2017) for further details about the derivations). To validate models and estimate the predictability of parents, a simple leave-one-out cross-validation (CV) scheme was applied following Fonseca et al. (2021a). Briefly, female parents in hybrid combinations were removed, and models trained with the remaining records where that seed parent was absent. The same process was applied for males. Thus, the CV scheme led to prediction of hybrids when a female parent or a male parent were removed. Female and male parents were removed individually. The prediction accuracy (r) at each ME was estimated as the Pearson's correlation between observed and predicted values of hybrids.

Genomic prediction model was fit using Bayesian methods implemented in the *BGLR* statistical package (Pérez and de los Campos, 2014) available for R. Inferences were based on 10,000 Gibbs sampler iterations with a burn-in of 5,000, and a thin of 5.

4.4. Results and discussion

4.4.1. Analysis by location

Genetic variation was significant for all traits in all environments (Table 12). Broad-sense heritabilities were generally high, indicating that most of the variation within a location was caused by genetic effects. Grain yield ranged from 4.31 to 8.34 ton ha^{-1}, PHT ranged from 121 to 145 cm, and DA from 59 to 80 days across environments. The lowest GY occurred in 2018 Monte Alto, TX, and the highest occurred in 2019 Colby, KS. In 2020, the National Agricultural Statistics Service reported grain sorghum average yields of 4.23 ton ha^{-1} and 5.71 ton ha^{-1} for Texas and Kansas, respectively (NASS, 2021). Such grain sorghum averages indicate that hybrids presented in this study generate yields similar to those of commercial hybrids, thus meeting the expectation of high-yielding germplasm. Fonseca et al. (2021b) provided a detailed report on the combining abilities of these grain sorghum hybrids and highlighted the benefits concerning genetic variance in exchanging elite germplasm between distinct public plant breeding programs.

Table 12. Phenotypic performance of hybrids developed from Texas A&M and Kansas State elite lines under a factorial II mating design across environments. Estimates include genetic variance (σ^2_{hyb}), residual variance (σ^2_e), coefficient of variation (CV), and broad-sense heritability (H^2). Designations for environments are as follows: Monte Alto 2018 (18RF), Victoria 2018 (18VC), College Station 2018 (18CS), Garden City 2018 (18GC), Colby 2018 (18COL), Taft 2019 (19TA), Victoria 2019 (19VC), College Station 2019 (19CS), Hays 2019 (19HAY), and Colby 2019 (19COL).

Trait	Parameter	Environments									
		18RF	18VC	18CS	18GC	18COL	19TA	19VC	19CS	19HAY	19COL
GY	Mean (t/ha)	4.31	5.18	6.26	5.58	5.18	6.65	6.46	7.78	6.51	8.34
	σ^2_{hyb}	0.70*	0.28*	0.66*	0.67*	0.57*	0.47*	0.49*	0.77*	0.47*	1.36*
	σ^2_e	0.86	0.15	0.66	0.95	1.24	0.18	0.29	1.36	1.55	0.84
	CV (%)	21.49	7.35	12.95	17.47	21.51	6.37	8.29	14.97	19.14	10.96
	H^2	0.71	0.85	0.75	0.68	0.58	0.84	0.77	0.53	0.38	0.76
PHT	Mean (cm)	121	127	126	145	137	138	140	140	122	129
	σ^2_{hyb}	64.6*	91.0*	155.7*	111.7*	94.8*	80.3*	71.2*	137.4*	89.3*	78.1*
	σ^2_e	26.31	19.33	48.76	27.15	25.32	30.14	14.32	53.02	28.05	67.18
	CV (%)	4.22	3.45	5.5	3.58	3.67	3.96	2.7	5.2	4.33	6.32
	H^2	0.88	0.93	0.91	0.93	0.92	0.84	0.91	0.84	0.86	0.7
DA	Mean (day)	67.9	67.2	78.7	70.6	72.4	69.6	59.5	80.4	64.2	69.1
	σ^2_{hyb}	11.1*	4.7*	10.9*	-	-	1.6*	3.5*	4.2*	-	-
	σ^2_e	4.77	0.99	4.36	-	-	0.54	0.78	6.88	-	-
	CV (%)	3.22	1.48	2.65	-	-	1.06	1.48	3.26	-	-
	H^2	0.87	0.93	0.88	-	-	0.86	0.90	0.55	-	-

* Significant at 0.05 probability level

4.4.2. Combined analysis and development of mega-environments

The likelihood ratio test (LRT) for the combined analysis indicated that genetic, environmental, and GxE effects were highly significant for all traits (Table 13). Environment effects explained most of the variation for GY (48.9%) and DA (76.7%), while genetic effects were the largest for PHT (40.5%). For GY, the GxE variance component accounted for almost 12% of the total variation, which is more than the variance associated with the genetic effects (9.8%). Yan (2020) indicates GxE effects that are greater than the genotype effect suggest that distinct mega environments exist. Given these results, the likelihood of finding a single best genotype for all environments is fairly low. Hence, mega-environment 1 (ME1) and mega-environment 2 (ME2) were developed (Table 13).

Table 13. Combined analysis for grain yield (GY), plant height (PHT), and days to anthesis (DA) considering all locations (Combined), Mega-environment 1 (ME1), and Mega-environment 2 (ME2). Estimates include variance components of model effects, mean of trait, coefficient of variation (CVe), and broad-sense heritability (H^2) for each dataset.

Trait	Variance Components	Combined		ME1		ME2	
		Estimate	%	Estimate	%	Estimate	%
GY	Hybrid	0.29**	9.75	0.28**	10.68	0.54**	13.92
	Environment	1.45**	48.92	1.40**	52.73	1.96**	50.90
	Hybrid × Env	0.35**	11.71	0.28**	10.45	0.19*	4.93
	Rep(Env)	0.08**	2.72	0.13**	4.76	0.01	0.36
	Residual	0.80	26.91	0.57	21.37	1.15	29.89
	Mean (t/ha)	6.22		6.10		6.40	
	CVe (%)	14.35		12.36		16.77	
	H^2	0.83		0.78		0.79	
PHT	Hybrid	82.39**	40.49	86.81**	42.94	84.60**	37.14
	Environment	68.50**	33.66	64.30**	31.80	97.06**	42.62
	Hybrid × Env	16.62**	8.17	14.15**	7.00	12.04**	5.28
	Rep(Env)	3.34**	1.64	5.27**	2.61	0.43	0.19
	Residual	32.65	16.04	31.64	15.65	33.63	14.77
	Mean (cm)	132.9		132.4		133.6	
	CVe (%)	4.30		4.25		4.34	
	H^2	0.97		0.95		0.94	
DA	Hybrid	5.33**	10.72	4.27**	6.07	-	-
	Environment	38.11**	76.70	60.41**	85.87	-	-
	Hybrid × Env	2.46**	4.96	2.01**	2.86	-	-
	Rep(Env)	0.58**	1.16	0.58**	0.82	-	-
	Residual	3.22	6.47	3.08	4.38	-	-
	Mean (days)	69.9		70.5		69.07	
	CVe (%)	2.56		2.49		-	
	H^2	0.94		0.89		-	

*, ** Significant at 0.01 and 0.001 respectively

The GGE biplot analysis for grain yield explained more than 50% of the variation with the first two principal components (PC) (Figure 7). The which-won-where polygon suggested that three ME exist. Nonetheless, the relative lack of representativeness of the environment 19HAY combined with empirical knowledge about sorghum production regions led to the establishment of only two ME (Figure 8 and Table 13). These two ME are consistent with previous grain sorghum studies that define production regions as subtropical and temperate environments (Monk et al., 2014). Mega-environment 1 included Texas locations (i.e., Monte Alto, Taft, Victoria, and College Station), while mega-environment 2 described Kansas locations (i.e., Hays, Colby, and Garden City) (Figure 8). See appendix A for details about the codes used in the biplot.

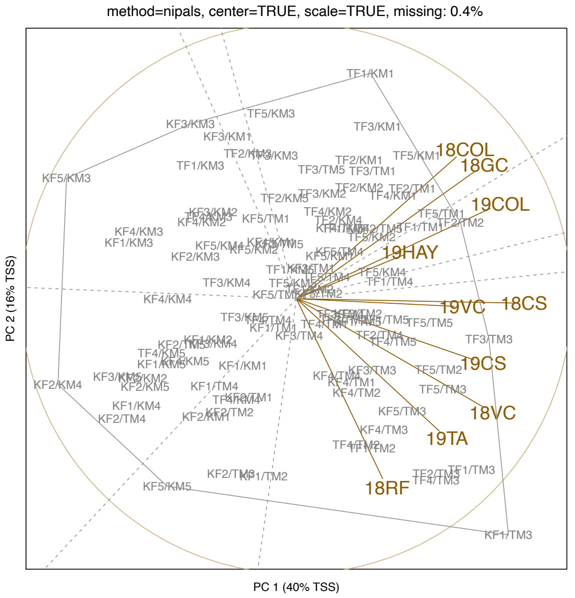

Figure 7. Genotype plus genotype × environment (GGE) biplot based on grain yield of sorghum hybrids evaluated in ten environments. GGE biplot was environment-centered and singular value scaling via nipals algorithm. Information includes the first and second principal components (PC) and the which-won-where polygon. Hybrids are identified in gray, while environments in dark orange.

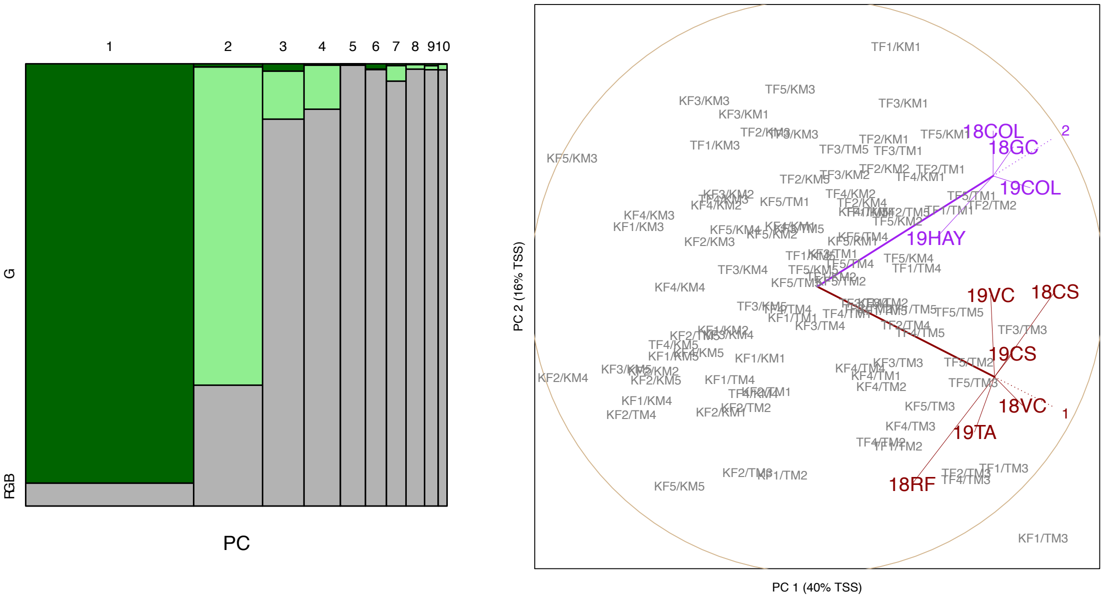

Figure 8. Mosaic plot and genotype plus genotype × block of environments (GGB) biplot of the multi-environment trial visualizing the two-way partitioning of the total sums of squares (TSS) into genotype (G), genotype × blocks of environments (GB), and residuals (R) along each GGB biplot axis for yield. The two mega-environments are denoted 1 (ME1) and 2 (ME2). ME1 included Texas environments labeled 18RF (Monte Alto 2018), 18VC (Victoria 2018), 18CS (College Station 2018), 19TA (Taft 2019), 19VC (Victoria 2019), 19CS (College Station 2019), while ME2 accounted for Kansas environments labeled 18GC (Garden City 2018), 18COL (Colby 2018), 19HAY (Hays 2019), and 19COL (Colby 2019). Hybrids are labeled according to each female x male combination. TF and TM codes for female and males derived from Texas A&M sorghum breeding programs, respectively, whereas females and males derived from Kansa breeding programs are identified as KF and KM respectively. Number following the inbred line origin identifies different parents.

4.4.3. Adaptability and stability of grain sorghum hybrids derived from public breeding programs

Grain sorghum hybrids derived from both the Texas AgriLife and Kansas programs not only developed some of the best hybrids but these combinations also resulted in highly stable genotypes (Figures 9, 10). For instance, within the subtropical environment (i.e., ME1), the highest yielding hybrid was a Kansas/Texas AgriLife combination (KF1/TM3) (Figure 9). More important, this hybrid had good stability (Figure 9). According to Finlay-Wilkinson (1963), linear regressions of genotype yield parallel to the population mean yield regression defines dynamic stability. Also, the authors described the ideal genotype as the one that outperforms others in both low- and high-yielding environments. The BFW regression method applied herein shows that KF1/TM3 has dynamic stability and represents a potential ideal genotype (Figure 9). For instance, KF1/TM3 presented one of the highest yields for 18RF and 19CS environments (Figure 9). Therefore, KF1/TM3 exemplifies a crucial characteristic that determines high-performing and stable grain sorghum hybrids: adapted to marginal areas and responsive to more favorable environments (Monk et al., 2014).

Within temperate ME2, the top two performers resulted from crosses involving Texas inbred lines (TF2/TM2 and TF2/TM1) (Figure 11). That implies that elite lines developed outside a particular target environment can produce good grain sorghum hybrids in another ME. It is common knowledge that it is easier to move germplasm from lower to higher latitudes (i.e., from South to North in the U.S.) than in the opposite direction (Rooney personal communication). Results presented herein support such a statement.

It is interesting to note that the top performers within ME1 expressed above-average PHT and below-average DA across environments (Figure 10). This highlights the positive correlation between PHT and GY and also indicates that earlier hybrids are less likely to be affected by biotic and abiotic stresses (i.e., midge damage and drought). However, for ME2, the top performers presented below-average PHT (Figure 12). Results suggest that taller plants might be undesirable where abiotic stress, such as wind or drought can lead to lodging. For those environments, shorter plants are likely to suffer less from lodging and stalk breakage problems, which justify these outcomes.

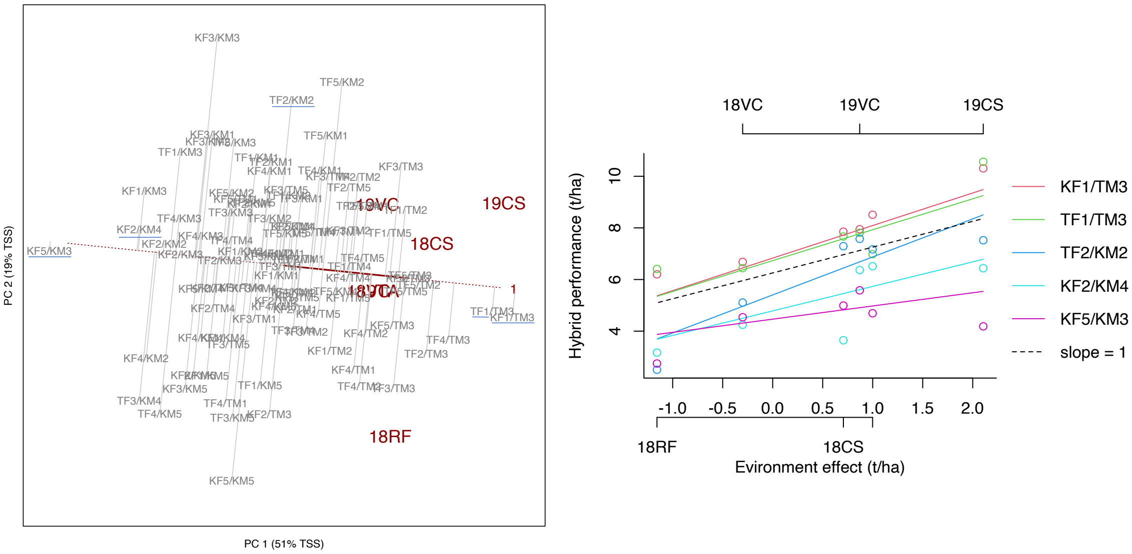

Figure 9. The "mean vs stability" representation of the genotype plus genotype × environment (GGE) biplot based on mega-environment 1 (left), and Bayesian Finlay-Wilkinson (BFW) regression (right) for grain yield. For GGE, data were environment-centered and not scaled; the biplot was focused on singular value partitioning using nipals algorithm and grain yield scores projections on the average environment coordinate displayed the stability of each hybrid. Underlined hybrids in the GGE biplot indicated the top and bottom two and an intermediate less stable hybrid. The underlined hybrids were assessed using BFW regression to further investigate static and dynamic stability.

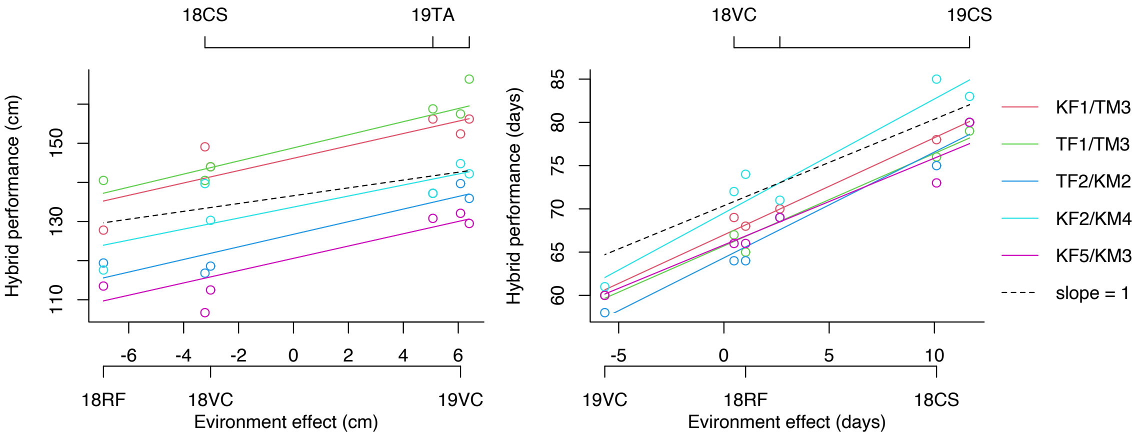

Figure 10. Bayesian Finlay-Wilkinson (BFW) regression of hybrids based on the top and bottom two hybrids and an intermediate less stable hybrid to assess plant height (left) and days to anthesis (right) stability of grain sorghums within mega-environment 1.

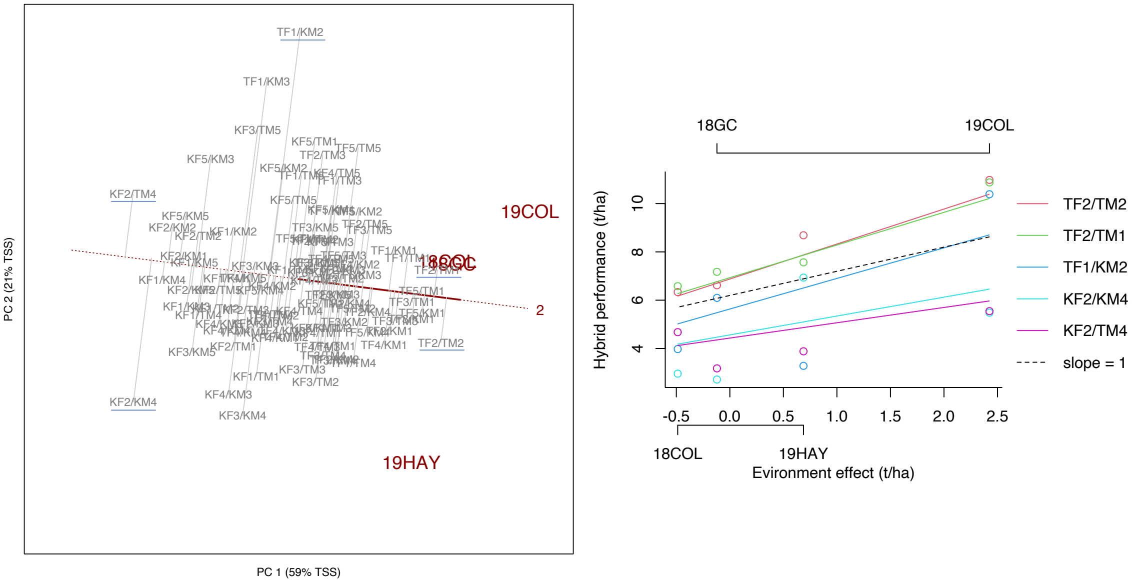

Figure 11. The “mean vs stability” representation of the genotype plus genotype × environment (GGE) biplot based on mega-environment 2 (left), and Bayesian Finlay-Wilkinson (BFW) regression (right) for grain yield. For GGE, data were environment-centered and not scaled. The biplot was genotype-focused singular value partitioning using nipals algorithm. Grain yield score projections on the average environment coordinate displayed the stability of each hybrid. Underlined hybrids in the GGE biplot indicated the top and bottom two hybrids and an intermediate less stable hybrid. The underlined hybrids were assessed using BFW regression to further investigate static and dynamic stability.

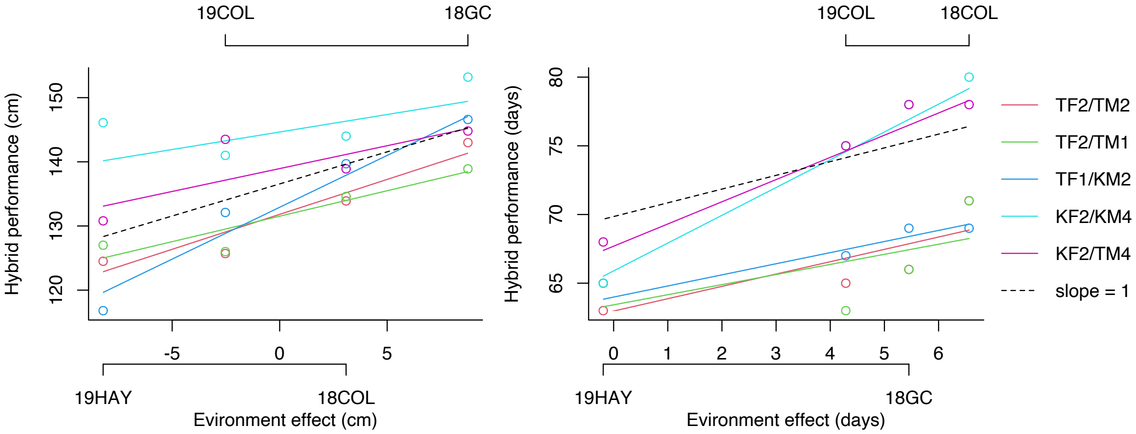

Figure 12. Bayesian Finlay-Wilkinson (BFW) regression of hybrids based on the top and bottom two hybrids and an intermediate less stable hybrid to assess plant height (left) and days to anthesis (right) stability of grain sorghums within mega-environment 2.

The Bayesian Finlay-Wilkinson applied herein presents an opportunity for sorghum breeders to assess the stability of genotypes appropriately. Among several advantages, such a method considers more flexible models when compared to fixed-effects models for allowing the inclusion of pedigree or genomic information, ultimately generating more reliable estimates and providing alternatives to manage the often MET challenge of missing values (Lian and de los Campos, 2016).

4.4.4. Genomic predictions modeling GxE within ME

Genomic predictions of grain sorghum hybrid performance within ME were inconsistent (Table 14). When compared to the prediction without ME, the leave-one-out female CV scheme presented higher prediction accuracies within ME1, while the leave-one-out male CV scheme presented better predictions in ME2 for GY and PHT. Overall, there was a slight increase in predicting GY and DA within ME1. These results indicate that closely related environments can increase prediction accuracies (Cuevas et al., 2016;

Crossa et al., 2019). However, other effects can negatively impact the potential of genomic predictions within ME.

The accuracy of genomic prediction models is influenced by many effects, including heritability of the trait, the size and structure of the training population, and the quality of the phenotypic data. Although ME can increase the correlations between environments, there are limited advantages for predicting untested genotypes because it reduces the number of hybrids present in the training set. While predictions without ME consider the entire data set, predictions within a ME only include information from that ME. As such, developing prediction models for specific ME is recommended when the target environment is well defined. For other situations, where predictions of untested hybrids involve new environments, reaction norm principles applied to genomic prediction models indicate promising results as both genomic and environmental information are included in the model (Jarquín et al., 2014; Li et al., 2018; Costa-Neto et al., 2020; Buntaran et al., 2021).

Table 14. Prediction accuracies of genomic GxE effects model for grain yield (GY), plant height (PHT), and days to anthesis (DA). Predictions are presented for genomic GxE effects model with and without partitioning data into mega-environments. Mega-environment 1 included environments from Texas year-location combination, while mega-environment 2 accounted for Kansas year-location combination as determined by Genotype plus genotype × environment (GGE) biplot. Numbers in bold show the overall prediction accuracy across females and males.

	Parents	Prediction accuracy without ME[‡]			Prediction accuracy in ME1			Prediction accuracy in ME2		
		GY	PHT	DA	GY	PHT	DA	GY	PHT	DA
Female	A05071	0.65	0.85	0.76	0.86	0.89	0.84	0.35	0.78	0.65
	A08140	0.59	0.85	0.69	0.75	0.89	0.81	0.36	0.80	0.56
	AARCH11051	0.67	0.91	0.76	0.79	0.92	0.79	0.51	0.90	0.72
	AARCH11129	0.37	0.87	0.69	0.51	0.91	0.77	0.17	0.79	0.56
	AARCH11136	0.69	0.82	0.86	0.71	0.80	0.90	0.64	0.83	0.79
	AARCH11146	0.63	0.79	0.81	0.75	0.81	0.79	0.45	0.77	0.83
	AHF14	0.60	0.80	0.83	0.66	0.83	0.88	0.46	0.75	0.80
	AKS118	0.71	0.82	0.84	0.78	0.83	0.85	0.54	0.78	0.83
	ATx3408	0.68	0.90	0.85	0.73	0.89	0.84	0.59	0.88	0.90
	ATx645	0.67	0.83	0.62	0.74	0.81	0.81	0.58	0.87	0.35
	Mean	0.63	0.84	0.77	0.73	0.86	0.83	0.47	0.81	0.70
Male	EON361	0.55	0.68	0.74	0.40	0.65	0.73	0.76	0.72	0.78
	R07178	0.64	0.84	0.83	0.68	0.83	0.81	0.59	0.87	0.85
	R08304	0.51	0.83	0.80	0.51	0.81	0.77	0.45	0.86	0.83
	RARCH11001	0.62	0.83	0.75	0.52	0.76	0.81	0.82	0.89	0.68
	RARCH11002	0.71	0.86	0.75	0.74	0.82	0.76	0.63	0.90	0.71
	RARCH11028	0.47	0.69	0.74	0.44	0.68	0.76	0.56	0.72	0.70
	RARCH11055	0.58	0.81	0.63	0.44	0.80	0.65	0.78	0.82	0.57
	RARCH11056	0.56	0.77	0.83	0.48	0.76	0.86	0.71	0.79	0.80
	RTx436	0.49	0.74	0.67	0.52	0.76	0.77	0.45	0.69	0.54
	RTx437	0.58	0.81	0.73	0.61	0.83	0.84	0.53	0.78	0.59
	Mean	0.57	0.79	0.75	0.53	0.77	0.77	0.63	0.80	0.70
Overall mean		**0.60**	**0.82**	**0.76**	**0.63**	**0.82**	**0.80**	**0.55**	**0.81**	**0.70**

[‡] Prediction accuracies presented in Fonseca et al. (2021a)

4.5. Conclusion

The identification of ME reduces GxE effects, provides better information about MET, and continues to be a common practice among plant breeders interested in both genotype and environment assessment. A MET for grain sorghum hybrids grown across Texas and Kansas confirmed the existence of two distinct ME that are consistent to historically different regions of adaptation generally designated as subtropical and temperate environments. Hybrid combinations between elite inbred lines adapted to each of those regions indicate a potential benefit of using such germplasm to generate high-performing and stable grain sorghum hybrids across locations. Plant breeders, growers, and seed companies are encouraged to explore such crosses to produce promising new products. Genomic enabled prediction models were inconsistent in their ability to predict hybrid performance within ME. Further research is needed to understand additional factors affecting the prediction of grain sorghum hybrids.

4.6. References

Acosta-Pech, R., J. Crossa, G. de los Campos, S. Teyssèdre, B. Claustres, et al. 2017. Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. Theor. Appl. Genet. 130(7): 1431–1440. doi: 10.1007/s00122-017-2898-0.

Alves, F.C., G. Galli, F.I. Matias, M.S. Vidotti, J.S. Morosini, et al. 2021. Impact of the complexity of genotype by environment and dominance modeling on the predictive accuracy of maize hybrids in multi-environment prediction models. Euphytica 217(3). doi: 10.1007/s10681-021-02779-y.

Alves, F.C., Í.S.C. Granato, G. Galli, D.H. Lyra, R. Fritsche-Neto, et al. 2019. Bayesian analysis and prediction of hybrid performance. Plant Methods 15(1): 14. doi: 10.1186/s13007-019-0388-x.

Ansarifard, I., K. Mostafavi, M. Khosroshahli, M. Reza Bihamta, and H. Ramshini. 2020. A study on genotype–environment interaction based on GGE biplot graphical method in sunflower genotypes (*Helianthus annuus* L.). Food Sci. Nutr. 8(7): 3327–3334. doi: 10.1002/fsn3.1610.

Balestre, M., V.B. dos Santos, A.A. Soares, and M.S. Reis. 2010. Stability and adaptability of upland rice genotypes. Crop Breed. Appl. Biotechnol. 10(4): 357–363. doi: 10.1590/S1984-70332010000400011.

Basnet, B.R., J. Crossa, S. Dreisigacker, P. Pérez-Rodríguez, Y. Manes, et al. 2019. Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. Plant Genome 12(1): 1–13. doi: 10.3835/plantgenome2018.07.0051.

Bates, D., M. Mächler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1). doi: 10.18637/jss.v067.i01.

Becker, H.C. 1981. Correlations among some statistical measures of phenotypic stability. Euphytica 30(3): 835–840. doi: 10.1007/BF00038812.

Becker, H.C., and J. Leon. 1988. Stability Analysis in Plant Breeding. Plant Breed. 101(1): 1–23. doi: 10.1111/j.1439-0523.1988.tb00261.x.

Buntaran, H., J. Forkman, and H.-P. Piepho. 2021. Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. Theor. Appl. Genet. 1: 3. doi: 10.1007/s00122-021-03786-2.

Burgueño, J., J. Crossa, P.L. Cornelius, and R.C. Yang. 2008. Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. Crop Sci. 48(4): 1291–1305. doi:

10.2135/cropsci2007.11.0632.

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52(2): 707–719. doi: 10.2135/cropsci2011.06.0299.

Coe, M.T., K.M. Evans, K. Gasic, and D. Main. 2020. Plant breeding capacity in U.S. public institutions. Crop Sci. 60(5): 2373–2385. doi: 10.1002/csc2.20227.

Comstock, R.E., and H.F. Robinson. 1952. Estimation of average dominance of genes. Estim. Aver. Domin. genes.

Cooper, M., C.D. Messina, D. Podlich, L.R. Totir, A. Baumgarten, et al. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65(4): 311–336. doi: 10.1071/CP14007.

Costa-Neto, G., R. Fritsche-Neto, and J. Crossa. 2020. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity (Edinb). doi: 10.1038/s41437-020-00353-1.

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2): 713–724. doi: 10.1534/genetics.110.118521.

Crossa, J., R. Fritsche-Neto, O.A. Montesinos-Lopez, G. Costa-Neto, S. Dreisigacker, et al. 2021. The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. Front. Plant Sci. 12(April): 1–6. doi: 10.3389/fpls.2021.651480.

Crossa, J., J.W.R. Martini, D. Gianola, P. Pérez-Rodríguez, D. Jarquin, et al. 2019. Deep Kernel and Deep Learning for Genome-Based Prediction of Single Traits in Multienvironment Breeding Trials. Front. Genet. 10(December): 1–13. doi: 10.3389/fgene.2019.01168.

Crossa, J., R.C. Yang, and P.L. Cornelius. 2004. Studying crossover genotype x environment interaction using linear-bilinear models and mixed models. J. Agric. Biol. Environ. Stat. 9(3): 362–380. doi: 10.1198/108571104X4423.

Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez, et al. 2016. Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. Plant Genome 9(3): 1–20. doi: 10.3835/plantgenome2016.03.0024.

Dalló, S.C., A.D. Zdziarski, L.G. Woyann, A.S. Milioli, R. Zanella, et al. 2019. Across year and year-by-year GGE biplot analysis to evaluate soybean performance and stability in multi-environment trials. Euphytica 215(6): 113. doi: 10.1007/s10681-019-2438-x.

FAOSTAT, 2021. Food and Agriculture Organization of the United Nations. FAOSTAT statistical database. Available online at: http://www.fao.org/faostat/en/#data/QC/visualize

Finlay, K., and G. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Aust. J. Agric. Res. 14(6): 742. doi: 10.1071/AR9630742.

Fonseca, J.M.O., Klein, P.E., Crossa, J., Pacheco, A., Pérez-Rodríguez, P., Perumal, R., Klein, R.R., and Rooney, W.L. 2021a. Assessing combining abilities, genomic data, and GxE interactions to predict hybrid grain sorghum performance. The Plant

Genome (Submitted).

Fonseca, J.M.O., Perumal, R., Klein, Klein, P.E., R.R., and Rooney, W.L. 2021b. Combining abilities and elite germplasm enhancement across US public sorghum breeding programs. Crop Science. (Submitted).

Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58(3): 453–467. doi: 10.1093/biomet/58.3.453.

Gauch, H.G. 1988. Model Selection and Validation for Yield Trials with Interaction. Biometrics 44(3): 705. doi: 10.2307/2531585.

Gauch Jr, H.G. 2013. A simple protocol for AMMI analysis of yield trials. Crop Sci. 53(5): 1860–1869. doi: doi.org/10.2135/cropsci2013.04.0241.

Gauch, H.G., and R.W. Zobel. 1997. Identifying Mega-Environments and Targeting Genotypes. Crop Sci. 37(2): 311–326. doi: 10.2135/cropsci1997.0011183X003700020002x.

González-Barrios, P., L. Díaz-García, and L. Gutiérrez. 2019. Mega-environmental design: Using genotype × environment interaction to optimize resources for cultivar testing. Crop Sci. 59(5): 1899–1915. doi: 10.2135/cropsci2018.11.0692.

Henderson, C.R. 1984. Applications of linear models in animal breeding.

Hu, X. 2015. A comprehensive comparison between ANOVA and BLUP to valuate location-specific genotype effects for rape cultivar trials with random locations. F. Crop. Res. 179: 144–149. doi: 10.1016/j.fcr.2015.04.023.

Huehn, M. 1990. Nonparametric measures of phenotypic stability. Part 1: Theory. Euphytica 47(3): 189–194. doi: 10.1007/BF00024241.

Jarquín, D., J. Crossa, X. Lacaze, · Philippe, D. Cheyron, et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127(3): 595–607. doi: 10.1007/s00122-013-2243-1.

Jarquin, D., R. Howard, J. Crossa, Y. Beyene, M. Gowda, et al. 2020. Genomic prediction enhanced sparse testing for multi-environment trials. G3 Genes, Genomes, Genet. 10(8): 2725–2739. doi: 10.1534/g3.120.401349.

Kofoid, K.D., and T.L. Harvey. 2005. Registration of Greenbug Resistant Sorghum Germplasm Lines KS 116 A/B through KS 120 A/B. Crop Sci. 45(2): 802. doi: 10.2135/cropsci2005.0802.

Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82(13). doi: 10.18637/jss.v082.i13.

Lado, B., P.G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2016. Modeling genotype × Environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci. 56(5): 2165–2179. doi: 10.2135/cropsci2015.04.0207.

Laffont, J.L., K. Wright, and M. Hanafi. 2013. Genotype plus genotype × block of environments biplots. Crop Sci. 53(6): 2332–2341. doi: 10.2135/cropsci2013.03.0178.

Li, X., T. Guo, Q. Mu, X. Li, and J. Yu. 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. 115(26): 6679–6684. doi: 10.1073/pnas.1718326115.

Lian, L., and G. de los Campos. 2016. FW: An R Package for Finlay–Wilkinson

Regression that Incorporates Genomic/Pedigree Information and Covariance Structures Between Environments. G3 Genes|Genomes|Genetics 6(3): 589–597. doi: 10.1534/g3.115.026328.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, et al. 2015. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 Genes, Genomes, Genet. 5(4): 569–582. doi: 10.1534/g3.114.016097.

Mbulwe, L., G.C. Peterson, J. Scott-Armstrong, and W.L. Rooney. 2015. Registration of sorghum germplasm Tx3408 and Tx3409 with tolerance to sugarcane aphid [*Melanaphis sacchari* (Zehntner)]. J. Plant Regist. 10(1): 51–56. doi: 10.3198/jpr2015.04.0025crg.

Miller, F.R., T.F. Dusek, K.L. Prihoda, and L.W. Rooney. 1992. Registration of RTx436 sorghum parental line. Crop Sci. 32(6): 1518–1518. doi: 10.2135/cropsci1992.0011183x003200060059x.

Monk, R., C. Franks, and J. Dahlberg. 2014. Sorghum. Yield gains in major US field crops. Wiley Online Library. p. 293–310

Nassar, R., and M. Huhn. 1987. Studies on Estimation of Phenotypic Stability: Tests of Significance for Nonparametric Measures of Phenotypic Stability. Biometrics 43(1): 45. doi: 10.2307/2531947.

NASS (U. S. Department of Agriculture National Agricultural Statistics Service). 2021. *2020* State agriculture overview Texas and Kansas. Available online at: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=T

EXAS and https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=KANSAS (accessed Mar 25, 2021).

NASS (U. S. Department of Agriculture National Agricultural Statistics Service) (2020). *Crop production*. Available online at: https://www.nass.usda.gov/Publications/Todays_Reports/reports/crop1020.pdf (accessed feb 5 2021).

Nielsen, D.C., and M.F. Vigil. 2018. Wheat yield and yield stability of eight dryland crop rotations. Agron. J. 110(2): 594–601. doi: 10.2134/agronj2017.07.0407.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58(3): 545–554. doi: 10.1093/biomet/58.3.545.

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198(2): 483–495. doi: 10.1534/genetics.114.164442.

Perumal, R., T. Tesso, K.D. Kofoid, R.M. Aiken, P.V. V. Prasad, et al. 2019. Registration of six grain sorghum pollinator (R) lines. J. Plant Regist. 13(1): 113–117. doi: 10.3198/jpr2017.12.0087crp.

Piepho, H.P. 1994. Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. 89(5): 647–654. doi: 10.1007/BF00222462.

Piepho, H.P. 1997. Mixed models with multiplicative terms for cultivar trial data. Adv. Biom. Genet: 215–220.

Piepho, H.-P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Springer-Verlag.

Piepho, H.P., and J. Möhring. 2005. Best Linear Unbiased Prediction of Cultivar Effects for Subdivided Target Regions. Crop Sci. 45(3): 1151–1159. doi: 10.2135/cropsci2004.0398.

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1–2): 209–228. doi: 10.1007/s10681-007-9449-8.

Podlich, D.W., and M. Cooper. 1998. Modelling plant breeding programs as search strategies on a complex response surface. Asia-Pacific Conference on Simulated Evolution and Learning. Springer. p. 171–178

R Core Team. 2020. R: A language and environment for statistical computing. https://www.r-project.org/.

Rakshit, S., K.N. Ganapathy, S.S. Gomashe, A. Rathore, R.B. Ghorade, et al. 2012. GGE biplot analysis to evaluate genotype, environment and their interactions in sorghum multi-location data. Euphytica 185(3): 465–479. doi: 10.1007/s10681-012-0648-6.

Resende, M.D.V., and R. Thompson. 2004. Factor analytic multiplicative mixed models in the analysis of multiple experiments. Rev. Matemática e Estatística, São Paulo, 22(2): 31–52. http://jaguar.fcav.unesp.br/RME/fasciculos/v22/v22_n2/A3_MDeon.pdf (accessed 23 October 2020).

Rooney, W.L. 2016. Sorghum: Production and Improvement Practices. 2nd ed. Elsevier

Ltd.

Rooney, W.L., F.R. Miller, and L.W. Rooney. 2003. Registration of RTx437 sorghum parental line. Crop Sci. 43(1): 445. doi: 10.2135/cropsci2003.4450.

Rosenow, D.T., L.E. Clark, G.C. Peterson, G.N. Odvody, and W.L. Rooney. 2021. Registration of Tx3440 through Tx3482 sorghum germplasm. J. Plant Regist. (1–9): plr2.20082. doi: 10.1002/plr2.20082.

Sharma, S.P., D.I. Leskovar, K.M. Crosby, and A.M.H. Ibrahim. 2020. GGE biplot analysis of genotype-by-environment interactions for melon fruit yield and quality traits. HortScience 55(4): 533–542.

Shelton, A.C., and W.F. Tracy. 2017. Cultivar development in the U.S. public sector. Crop Sci. 57(4): 1823–1835. doi: 10.2135/cropsci2016.11.0961.

Shukla, G.K. 1972. Some statistical aspects of partitioning genotype environmental components of variability. Heredity (Edinb). 29(2): 237–245.

Smith, A.B., L.M. Borg, B.J. Gogel, and B.R. Cullis. 2019. Estimation of Factor Analytic Mixed Models for the Analysis of Multi-treatment Multi-environment Trial Data. J. Agric. Biol. Environ. Stat. 24(4): 573–588. doi: 10.1007/s13253-019-00362-6.

Smith, A., B. Cullis, and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57(4): 1138–1147. doi: 10.1111/j.0006-341X.2001.01138.x.

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric.

Sci. 143(6): 449–462. doi: 10.1017/S0021859605005587.

Sorensen, D., and D. Gianola. 2007. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer Science & Business Media.

Technow, F., D. Podlich, and M. Cooper. 2020. Back to the future: Implications of genetic complexity for hybrid breeding strategies. bioRxiv: 2020.10.21.349332. doi: 10.1101/2020.10.21.349332.

Technow, F., T.A. Schrag, W. Schipprack, E. Bauer, H. Simianer, et al. 2014. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197(4): 1343–1355. doi: 10.1534/genetics.114.165860.

Wricke, G. 1962. Uber eine Methode zur Erfassung der okologischen Streubreite in Feldverzuchen. Z. pflanzenzuchtg 47: 92–96.

Wright, K., and J.-L. Laffont. 2020. gge: Genotype Plus Genotype-by-Environment Biplots. R Packag. version 1.6. https://cran.r-project.org/web/packages/gge/gge.pdf (accessed 12 May 2021).

Yan, W. 2015. Mega-environment Analysis and Test Location Evaluation Based on Unbalanced Multiyear Data. Crop Sci. 55(1): 113–122. doi: 10.2135/cropsci2014.03.0203.

Yan, W. 2016. Analysis and handling of G× E in a practical breeding program. Crop Sci. 56(5): 2106–2118.

Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Sci. 40(3): 597–605. doi: 10.2135/cropsci2000.403597x.

Yan, W., and M.S. Kang. 2002. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

5. CONCLUSIONS

This dissertation reported several significant events responsible for the successful improvement of grain sorghum hybrids and emphasized the major contributions made by the Texas A&M and Kansas State sorghum breeding programs throughout the process. The second chapter demonstrated that maintaining selectable genetic variation remains one of the most challenging factors in mature plant breeding programs. Nonetheless, the exchange of elite germplasm could become a modern practice for sorghum breeding programs to maximize variation and improve the agronomic performance of grain sorghum hybrids. Moreover, the study suggested that such a practice could foster collaborative efforts between public institutions to address their local challenges.

The third chapter explored classical and genomic models for predicting grain sorghum hybrid performance under a multi-environment trial. The inclusion of genomic information in kernel-based GBLUP models incorporated the natural population structure existent in a hybrid crop breeding scheme, allocating the genetic similarity to each specific genetic main effect. Further, the incorporation of the GxE interaction effect in kernel-based GBLUP models allowed borrowing information existing among the females and males involved in hybrid combinations and exploiting the genetic correlations among environments. Hence, the suitability of the Kernel-based GxE model improves the prediction capacity of grain sorghum hybrids, permits increasing selection intensity, and, ultimately, increasing the rates of genetic gain.

The fourth chapter expanded the benefits presented in chapter two. Hybrid combinations between elite inbred lines adapted to distinct mega-environments generate high-performing and stable grain sorghum hybrids across locations. Thus, plant breeders, growers, and seed companies should explore crossing elite germplasm adapted to diverse environments to produce promising new products.

APPENDIX A

Names, codes, and origin of grain sorghum elite lines used to generate hybrid combinations following a factorial II mating scheme. Information includes environment names and codes involved in the multi environment trial to assess grain yield, plant height, and days to anthesis of grain sorghum hybrids derived from U.S. public breeding programs.

Origin	Female	Codes	Male	Codes	Environment	Codes
Texas	A05071	TF1	EON361	TM1	Monte Alto 2018	18RF
	A08140	TF2	R07178	TM2	Victoria 2018	18VC
	ATx3447	TF3	R08304	TM3	College Station 2018	18CS
	ATx3408	TF4	RTx436	TM4	Garden City 2018	18GC
	ATx645	TF5	RTx437	TM5	Colby 2018	18COL
Kansas	AARCH11051	KF1	RARCH11001	KM1	Taft 2019	19TA
	AARCH11129	KF2	RARCH11002	KM2	Victoria 2019	19VC
	AARCH11136	KF3	RARCH11028	KM3	College Station 2019	19CS
	AARCH11146	KF4	RARCH11055	KM4	Hays 2019	19HAY
	AKS118	KF5	RARCH11056	KM5	Colby 2019	19COL