GENOMIC PREDICTION OF SWEET SORGHUM AGRONOMIC PERFORMANCE UNDER DROUGHT AND IRRIGATED ENVIRONMENTS IN HAITI

By

MARIE DORVAL

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2019

To my lovely family, friends and mentors

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to USAID through the AREA project for funding this research and giving me the opportunity to broaden my knowledge so I can contribute to the expansion of research in my country. My sincere thanks also go to my advisor: Dr. Geoffrey Meru for his motivation, patience and immense knowledge. His guidance helped me through my research and writing my thesis. I could not have imagined having a more supportive advisor for my studies. Besides, I am also grateful for every member of my thesis committee: Dr. Gael Pressoir, Dr. John Erickson and Dr. Wilfred Vermeris for their guidance and valuable comments throughout my study at the University of Florida. Their encouragement and thoughtful questions motivated me to widen my knowledge and research perspectives. I would also like to thank the CHIBAS staff (agronomists, field crew, administrators) for their participation in setting up the experiments and collecting the phenotypic data. Without their precious support, it would not be possible to complete the field trials on time and generate phenotypic data of good quality. I am also immensely grateful for Kansas State University for generating the genotypic data. I would like to acknowledge the support and help provided by Dr. Ferrao Luis Felipe and Rampazo Amadeu Rodrigo in the statistical analysis. In addition, I would like to thank the staff of the Tropical Research Center (TREC) for their support and to have welcomed me in this big family. Many thanks to my fellow labmates (Riphine, Vincent, Yuquin) for their stimulating discussions and accepting nothing less than excellence from me. Last but not the least, I would like to express my sincere gratitude to my family: my parents, my brothers and sisters, and my fiancé for supporting me spiritually throughout this study and in my life in general.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aux_traits	Auxiliary traits
CNSA	National Center for Food Security
FAO	Food and Agriculture Organization of the United Nations
GEBVs	Genomic Estimated Breeding Values
G x E	Genotype x Environment
GS	Genomic Selection
GY	Grain Yield
H	Days to Heading
JW	Juice Weight
LW	Leaf Weight
M	Days to Maturity
PAM	Word Food Program
PH	Plant Height
PWS	Pre-flowering Water Stress
SD	Stem Diameter
SW	Stem Weight
TGL	Total Green Leaf
TLN	Total Leaf Number
TSN	Total Stem Number
TSS	Concentration of soluble solids
USAID	United States Agency for International Development
VWS	Vegetative Water Stress

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

GENOMIC PREDICTION OF SWEET SORGHUM AGRONOMIC PERFORMANCE UNDER DROUGHT AND IRRIGATED ENVIRONMENTS IN HAITI

By

Marie Dorval

August 2019

Chair: Geoffrey Meru
Major: Horticultural Sciences

High-throughput phenotyping remains costly and inaccessible to most plant breeding programs. Over the last decade, genomic selection (GS) has gained momentum as a tool for predicting genetic gain in plant breeding populations, while lowering costs associated with phenotyping. Different statistical models and approaches have been developed to implement GS in plant breeding, and strategies that promote accurate and resource-efficient prediction are of increasing interest. Since its establishment in 2010, the sweet sorghum breeding program at CHIBAS, Haiti, has led efforts to develop and release cultivars resilient to abiotic and biotic stress. Among abiotic constraints, drought stress is the most limiting since growers depend on erratic rainfall for sorghum production in Haiti. The goal of the present study was to predict the genomic estimated breeding values of a sweet sorghum breeding population (n=250) under contrasting environments in Haiti using four statistical models (Bayes A, B, C and Bayesian Ridge Regression (BRR)). We evaluated twelve sorghum traits and performed within and across irrigated, pre-flowering and vegetative water stress prediction scenarios. Overall, the four methods showed similar results, however Bayes B and BRR

were superior in prediction accuracy and computation time, respectively. Generally, prediction accuracy was higher for within-environment (0.31 to 0.7) than across-environment (0.06 to 0.7) involving vegetative water stress scenarios. Prediction accuracy varied substantially for all traits, with total green leaf showing the highest mean value (0.70), and grain yield showing the least (0.49). Overall, there was no improvement in the prediction accuracy of grain yield with multi-traits genomic selection.

CHAPTER 1
INTRODUCTION

The predicted explosion in human population over the next few decades will place immense pressure on world food supply, as well as deplete already dwindling petroleum reserves. This challenge, coupled with global warming and food-fuel competition, requires innovation and development of high-yielding well adapted crops, particularly dual-purpose crops that can be used for both food and fuel. One such crop is sweet sorghum (*Sorghum bicolor* (L.) Moench). Sweet sorghum is a source of grain and stem for sugar, alcohol, syrup, jaggery, fodder, fuel, bedding, roofing, fencing, paper and chewing (Ratnavathi et al. 2011). In Haiti, sweet sorghum is mainly cultivated for its grains. However, there is a growing interest among growers for multipurpose varieties that can be used to diversify household income, particularly in the rural areas (Leclerc et al. 2014). The sorghum grains are used for human alimentation and beverage production such as Maltha-H and Kinanm (a Haitian beer). Sorghum is the third most important cereal (100,000 tons) in Haiti, after maize (205,000 tons) and rice (145,000 tons) (FAO and PAM 2017). The Centre department (31,340 ha) leads in production of sorghum in Haiti, followed by Artibonite (30,596 ha) and South department (19,729 ha) (William 2016).

Due to its environmental resilience, sorghum is an important crop for ensuring world food security, especially in developing nations such as Haiti. Unlike other bioenergy/ food crops such as sugarcane (*Saccharum* spp.), corn (*Zea mays*), sugar beet (*Beta vulgaris*), wheat (*Triticum aestivum*), cassava (*Manihot esculenta*) and potato (*Solanum tuberosum*), sorghum is well adapted to marginalized arid and semiarid regions due to its high water, radiation and nutrient use efficiency (Mathur et al. 2017).

However, tolerance to different environmental conditions varies by genotype, thus specific cultivars do well in one environment and not the other. In a country like Haiti with variable agro-ecological zones (low land to high lands), it is paramount to develop varieties with wide adaptability. However, conventional breeding for new superior varieties is time-consuming and expensive, often requiring decades of phenotyping and selection. Nevertheless, with contemporary breeding tools such as genomic selection, crop-breeding cycles can be significantly reduced, and genetic gains improved each cycle. The goal of the current project was to use genomic selection as a tool for predicting agronomic performance of a breeding population under drought and irrigated conditions in Haiti.

Statement of the Problem

Drought is a major limiting factor in agriculture and is considered the most important cause of yield reduction in crop plants (Boyer 1982; Tuberosa et al. 2003). Agriculture in Haiti is rainfall dependent, which makes water availability (quantity and distribution) unpredictable across seasons. The result is sometimes unusually long periods of drought, frequent floods and pest infestation, leading to catastrophic crop losses (CNSA 2012). Climate change has not helped, with erratic, intense, early rains, and prolonged drought periods. For example, the 2017 spring cropping season experienced insufficient and poorly distributed rainfall, which led to 18% reduction in sorghum production, 15% for maize and 6% for legumes (FAO and PAM 2017), thus highlighting the need to develop more drought-tolerant sorghum varieties. Although sorghum is both a drought and heat resistant plant (Nasidi et al. 2010), the level of tolerance is genotype dependent (Assefa et al. 2010). Given the economic importance of sorghum in Haiti, there is a clear and urgent need to develop high-yielding and

drought-tolerant sorghum cultivars widely adapted to various agro-ecological zones in Haiti.

Significance of the Study

Sweet sorghum is the third most important cereal in Haiti (Leclerc et al. 2014). It is mainly cultivated for its grain, which is used as food and for production of alcoholic beverages. Due to its agronomic and economic importance in the country, breeding efforts to alleviate production challenges (biotic and abiotic) were initiated in 2010 at CHIBAS, Haiti. Through this initiative, several superior sweet sorghum varieties have been released to growers. However, given the variation in agro-ecological zones and erratic rainfall patterns in the country, there is still need for new varieties adapted to harsh environments, particularly marginal semi-arid regions.

Traditional breeding for drought tolerance at CHIBAS is resource intensive and expensive due to long selection process and heavy labor requirement for phenotyping. Moreover, breeding for drought tolerance is known to be challenging for breeders regardless of many decades of research, as most of drought tolerance-related traits are polygenic (Kidanemaryam 2019; Zhang et al. 2015).The overarching goal of the breeding program is to integrate modern plant breeding tools to; 1) expedite varietal development and release through improved selection efficiency, and 2) reduce phenotyping costs. Due to its relatively small genome size (700 Mb) and short life cycle, sorghum is a model for other complex grass cereals and a target for genomics-assisted breeding (Bekele et al. 2013). Genomic selection is a process where large numbers of molecular markers (mostly SNPs) across the genome are used to estimate genetic variance for a given trait in a population, as opposed to focusing on a single Quantitative trait locus (Brown et al. 2014). Genomic selection integrates all genetic

effects (both large and small-effect QTL) into genomic estimated breeding values (GEBV), which is a representation of the genetic value of individuals (Stich and Inghelandt 2018; Windhausen et al. 2012; Desta and Ortiz 2014; Heffner et al. 2009). Using genomic selection in breeding can have many advantages including reduction in phenotyping cost and early generation plant selection. As genotyping costs continue to decrease, genomic selection will allow increased selection intensity, and therefore efficient utilization of available genetic resources (Biochard et al. 2016). However, the selection of stable lines can be challenging due to the presence of genotype × environment (G × E) interactions that can negatively affect the heritability of the traits and response to selection (Roorkiwal et al. 2018). Heritability of a trait, which is the proportion of total phenotypic variation that is due to genetic variation, is influenced by environmental factors, particularly for quantitative traits (Visscher et al. 2008). Thus, the performance of a group of lines can change from one environment to another. Hence accounting and modeling for G × E interaction in genomic prediction models could help breeders select multiple traits in the same or different environment at the same time.

Using a breeding population with genotypic and phenotypic data, the goal of the current project was to evaluate four statistical models for the estimation of GEBVs of 250 genotypes under drought and irrigated conditions in Haiti. Through this project, improved multipurpose sweet sorghum varieties (grain, feed, alcoholic beverage) will be released for the sub-humid and drought prone environments. Moreover, the current research project will allow the identification of the best scenario for implementation of genomic selection in sweet sorghum breeding in Haiti.

Objectives of the Study

The current project aimed to estimate the prediction accuracy of twelve sorghum traits within and across different environments (irrigation and water stress) using four different prediction models.

The following are the specific objectives:

1. Phenotypically characterize a sweet sorghum breeding population (n=250) for agronomic performance across drought and irrigated environments

2. Determine the best prediction model for an accurate prediction within environment

3. Estimate the prediction accuracy of the twelve sorghum traits within and across environments

Hypotheses

Hypothesis 1: Predictive accuracy within environments will be higher than across environments

Hypothesis 2: Variation in prediction accuracy among the models will be marginal but substantial among the traits

Hypothesis 3: The prediction accuracy of grain yield can be improved when a genetically correlated indicator trait with higher heritability is considered within the genomic prediction model.

CHAPTER 2
LITERATURE REVIEW

Environmental Adaptation

Sweet sorghum (*Sorghum bicolor* (L.) Moench) is a C4 plant that belongs to the Poaceae family, which, in addition to grains for human and animal consumption, also accumulate high level of sugars in its stems, usable for many products including syrup, rum (in Haiti) and bioethanol. It is a low water-demanding plant (about 400 to 700 mm of rain) (House 1987). Sorghum converts CO_2 into sugars efficiently, has high nitrogen and solar radiation use efficiency, short growth cycle (3-4 months), high drought tolerance (will survive with less than 300 mm of water), salinity and alkalinity tolerance (Tari et al. 2012; House 1987; Assefa et al. 2010).

Nutrition and Use

Sweet sorghum is a multipurpose crop. The juice extracted from its stems serves as raw material for syrup, jaggery and bio-ethanol production (Figure 2-1). Moreover, the bagasse, which is a byproduct of stem juice extraction, is a source of energy and organic fertilizer (Mathur et al. 2017). Sweet sorghum, which has significant sugar content in the stems, is closely related to sugarcane and is widely cultivated in some countries as a complementary crop to sugarcane for bioethanol production, as it has a shorter life cycle compared to sugarcane (Reddy et al. 2005; Burks et al. 2013). Although it is generally considered as an annual grass, sorghum is grown as a perennial crop and harvested several times a year in the tropics, to prevent soil degradation (Cox et al. 2010). The grain of sweet sorghum is also an important source of nutrients whose value relative to that of other major cereals is summarized in Table 2-1. It is naturally gluten free (safe for people with gluten intolerance), and contains significant level of

antioxidants, which help fight against cancer, diabetes, heart and neurological diseases (De Morais Cardoso et al. 2015; Shen et al. 2018). However, it is not recommended to consume freshly germinated sorghum because of the presence of significant amounts of dhurrin, a cyanogenic glycoside. The hydrolysis of this compound produces a powerful toxin: prussic acid or cyanogenic acid (HCN), which is known as a poison for human and animals (Panasiuk and Bills 1984). Sorghum grain can be consumed in a variety of forms that vary from region to region. In general, it is consumed as whole grain or processed into flour, from which traditional meals are prepared (Shen et al. 2018). These meals include flat bread, mostly unleavened and prepared from fermented or unfermented dough in Asia and parts of Africa, thin or thick fermented or unfermented porridge, mainly consumed in Africa, boiled products similar to those prepared from maize grits or rice and preparations deep-fried in oil (Léder 2004). Another important utilization of sorghum, especially in Africa and now in Haiti, is for nutritional and alcoholic beverage preparation. For example, in Haiti the sorghum grain is used in the production of Malta H, a nutritional beverage, and beer.

Sorghum Production in Haiti

Many sorghum varieties are grown in Haiti. Sorghum production in Haiti used to be dominated by Guinea grain varieties in marginal areas; but these have since disappeared because of sugarcane aphid (*Melanaphis sacchari*) proliferation. Currently, most varieties grown in Haiti were released by the breeding program at CHIBAS and have moderate to high levels of total soluble solids (14 to 17 °Brix). Currently, the most commonly cultivated variety is Papèpichon, a sugarcane aphid resistant variety released by CHIBAS. Before the proliferation of sugarcane aphid in 2015, Papèsèk used to be the most common varieties adopted by the Haitian farmers but has

disappeared. Papèsèk is a dual-use variety (forage and grain) from Central American breeding programs. It is a stay-green variety, very tolerant to drought and close to a yield of 4.4 tons ha^{-1}. Papèsèk has soluble solids concentration close to 16 °Brix (Leclerc et la. 2014). Another well-known variety is Dekabès. While released by CHIBAS as a sweet sorghum variety that gives excellent rum, Dekabès has never been adopted because its grain is too floury (floury endosperm while farmers/consumers require a vitreous/corneous endosperm). Although it has comparable yield (3 tons ha^{-1}) to Papèsèk, it is more sensitive to water stress. Dekabès stems contain more sugar than those of Papèsèk and can be used to produce syrup and rum (Leclerc et al. 2014). Dekabès is also a stay-green and sugarcane aphid resistant line. These two varieties have been used by the plant breeding program from CHIBAS to create new varieties combining high yield, high stem sugar levels, drought tolerance and insect resistance. Although Centre department (31,340 ha), Artibonite (30,596 ha) and South (19,729 ha) are the biggest producers, other regions such as West (19,560 ha), Nippes (12,049 ha), North West (6,400 ha), South East (6 ,135 ha), North (1,571 ha) and North East (431 ha) have modest sorghum production (William 2016). About 300,000 farmers rely on this crop as a source of food and income (Flecher 2016; Gabriel 2016). Two different types of sorghum are grown in Haiti: photoperiodic (the local variety) and non-photoperiodic (improved variety). The photoperiodic and non-photoperiodic sorghum differ in cycle duration, which depends on day length and planting date. The non-photoperiodic varieties flower rapidly, regardless of day length, while the cycle duration of photoperiod sensitive varieties change depending on day length and the sowing date (Wolabu and Tadege 2016). The local variety is widely cultivated in areas receiving

about 400 and 800 mm of rainfall per year (William 2016) (Figure 2-2). The photoperiodic sorghum is planted during June or July and can have a grain yield of about 0.8 tons ha^{-1} without any inputs. In contrast, the non-photoperiodic variety is planted all year long and can have a combined yield of 10 tons ha^{-1} over three harvests (William 2016).

Breeding Goals for Sorghum

Generally, sweet sorghum breeding targets improvement of grain and/or fodder yield, reduced maturity, enhanced nutrition (micronutrients and low HCN level) (Pineli et al. 2015; Dahir et al. 2015), and tolerance to abiotic factors (e.g. drought and salinity), diseases, e.g. blight, downy mildew, rust, smut, anthracnose and charcoal rot (Cassady et al. 1962; Erpelding 2010; Mofokeng et al. 2017) and insect-pests (e.g. sugarcane aphid (*Melanaphis sacchari*), sorghum midge (*Contarinia sorghicola*), sugarcane (*Diatraea saccharalis*) and corn borers (*Ostrinia nubilalis*), stem borer (*Chilo partellus*) and shoot fly (*Atherigona soccata*) (Sharma et al. 1993). These goals can be achieved using conventional breeding methods (e.g. pedigree method, pure line selection, back cross breeding, and mass selection) (Chaurasla 2015; Vaksmann et al. 2008, Piper and Kulakow 1994) or modern breeding methods such as genomic selection (Fernandes et al. 2018; Oliveira et al. 2018), marker-assisted selection (MAS) (Murray et al. 2009; Olweny et al. 2012), or transgenic approaches (Che et al. 2018, Jiang et al. 2013).

Biotic Factors Affecting Worldwide Sorghum Production

Biotic factors cause significant annual crop losses in the world. Economically important foliar diseases affecting sweet sorghum production include anthracnose (caused by *Colletotrichum sublineola*), leaf blight (caused by *Exserohilum turcicu*), sooty stripe (caused by *Ramulispora sorgi*), bacterial leaf strip (caused by *Burkholderia*

andropogonis), head smut (caused by *Sporisorium reilianum*), rust (caused by *Puccinia purpurea*), sorghum coal (caused by *Sphacellotheca sorghi*) and downy mildew (caused by *Peronosclerospora sorghi*) (Bonnet et al. 2012; White et al. 2011). Fungal rots caused by *Fusarium sp.* (Fusarium root rot) and *Macrophomina phaseolina* (charcoal rot) are the most common. Economically significant viral diseases include maize dwarf mosaic and sugarcane mosaic.

Insect pests affecting sweet sorghum production include sugarcane aphid (*Melanaphis sacchari*), chewing caterpillars (*Spodoptera frugiperda*), and sorghum midge (*Stenodiplosis sorghicola* or *Contarinia sorghicola*). Bird pests also cause devastating damage to crop throughout the growing season. However, most significant damage is typically observed during early flowering and late flowering. The most prevalent bird pests in sorghum include weavers (*Ploceus cucullatus*), parrots (*Aratinga* sp.), sparrows (*Passer* sp.), quelea (*Quelea quelea*), crows (*Corvus* sp.), blue-black grassquit (*Volatina jacarina*), pigeons (*Patagioena spicazuro*), and doves (*Columbina talpacoti*) (Priyavratha and Nakasimha, 1953; Melo and Cheschini, 2012).

Abiotic Factors Affecting Worldwide Sorghum Production

Nutrient deficiency, high salinity, aluminum toxicity, waterlogging, temperature stress and drought are the most important abiotic stresses limiting sorghum production around the world (Tari et al. 2012). These factors generally result in combined yield losses of up to 1 ton ha^{-1} (Rao 2004).

The key elements for sorghum nutrition are: nitrogen, phosphorous, potassium, and zinc (Erickson et al. 2012; Singh et al. 2012; Wortmann et al. 2013; Adams et al. 2015). Nitrogen is a major nutrient for all crops and its deficiency is a major constraint to sorghum production globally. Economic losses due to nitrogen deficiency depend on the

growth stage of the plant. In sorghum, losses are most significant during anthesis where between 16 to 30% of florets can abort (Rao 2004). Phosphorous is most likely the second nutrient essential for sorghum production optimization (Wortmann et al. 2013). Few symptoms are clearly recognizable with mild phosphorous deficiency, except for lack of vigor, thin plants with late flowering. When severe phosphorous deficiency occurs, it slows the plant growth and dark red to purple overtones develop on the surface of older leaves (Grundon et al. 1987). Similar to phosphorous deficiency, plants with potassium deficiency show a lack of vigor and late flowering. When this deficiency is severe, it induces dwarfism in plants with the older leaves showing marginal necrosis. However, the necrosis development pattern differs greatly among cultivars (Grundon et al. 1987). Zinc is a vital element for the optimization of sorghum production. Zinc deficiency inhibits nitrogen uptake and utilization, which reduces sorghum yields and lowers plant biomass (Roberson 2013).

Sorghum is considered to have moderate tolerance to salt stress (Azhar and McNeilly 1987). Salt stress can result from high concentrations of minerals and toxic ions, which can decrease the percentage of germination or increase the duration of germination in sweet sorghum (Tari et al. 2012).

Aluminum toxicity is major challenge for sorghum production, especially in soils with low pH (Tari et al. 2012; Magalhaes et al. 2007). To cope with aluminum toxicity, sorghum plants use an exclusion mechanism, chelation of the metal ion by organic acids in the rhizosphere. In 2007, a gene that plays a key role in aluminum toxicity tolerance in sorghum was identified and belongs to the multidrug and toxic compound

extrusion (MATE) family, an aluminum- activated citrate transporter (Magalhaes et al. 2007).

Waterlogging conditions occur as a result of flooding, especially in soils with poor drainage. The frequency of storms and heavy rain in the tropical and sub-tropical regions create a waterlogging environment that can impact plant metabolism and negatively affect the soil texture. The effects of waterlogging depend on the age of the plant; however, after 30 days of plant growth, shoot growth is not significantly affected. To cope with waterlogging, sorghum plants develop new nodal roots and form aerenchyma from the roots to the stalks (Tari et al. 2012).

Extreme temperatures (high and low) are limiting factors to sorghum production and are related to planting dates (Tari et al. 2012). A late planting can negatively affect the stem sugar content and biomass yield (Erickson et al. 2011), specifically in arid environments (Almodares and Mostafi Darany 2006). Sweet sorghum is highly sensitive to low temperature, which can inhibit seed germination (soil temperature below 10 ˚C), seedling emergence and overall plant growth (Tari et al. 2012). On the other hand, high temperature can repress biomass production and sugar yield (Tari et al. 2012). Although sorghum is well adapted to heat stress, this adaptation is genotype dependent (Tack et al. 2017). The temperature threshold beyond which sorghum yields starts to decline is about 33 ^{o}C (Tack et al. 2017).

Water stress is a major limiting factor in agriculture and is considered the most important cause of yield reduction in crop plants especially cereals (Boyer 1982; Tuberosa et al. 2003). Water is an essential element for perpetuation of both animal and plant life. Water represents almost 90% of the plant body on a fresh-weight basis and

performs several vital functions during plant growth, nutrition, seed germination, photosynthesis, multiplication of soil organisms and acts as a solvent for nutrient absorption from the soil. Therefore, the ability of plants to resist water stress is economically important. Depending on their response to water availability, plants can be classified as mesophytes (adapted to moderate supply of water), hydrophytes (require water-logged habitat) and xerophytes (tolerate dry conditions). Even though sorghum is considered as a drought tolerant plant compared to other major crops (Table 2-2), it is still susceptible to water stress in a genotype dependent manner. According to a study conducted by Miller and Ottman (2010), water stress can negatively affect radiation interception, leaf-number, plant height and grain yield. Studies conducted by the sweet sorghum breeding team at CHIBAS show that all sweet sorghum varieties currently under development are sensitive to drought, each at a different level. Some varieties have significantly reduced yield in contrasting environments (irrigated vs water stress), while others show stable yield across the environments. However, across varietal types, water stress does not negatively affect the concentration of total soluble solids. There are three adaptive mechanisms that a sorghum plant uses to cope with drought stress: drought escape (early maturity, leaf rolling, and remobilization of stem reserves), drought avoidance (adjustment of leaf area, increasing root systems) and drought tolerance where adapted cultivar usually exhibits light green and erect leaves (Rao 2004). Drought tolerance indices such as mean productivity, stability tolerance index, harmonic mean, yield stability index, tolerance index, and stress susceptibility index can be estimated to select superior genotypes under drought and irrigated conditions (Menezes et al. 2014).

Stay-green Trait in Sorghum

An important drought adaptation mechanism in sorghum is the stay-green trait. Stay-green trait is characterized by impaired or delayed chlorophyll degradation pathways. Quantitative traits loci (QTLs) associated with stay-green traits are routinely targeted in sorghum breeding programs to improve crop tolerance to stress and eventually increase plant biomass and grain yield (Borrell et al. 2014; Meru 2010; Thomas and Ougham 2014). In sorghum, functional stay-green is physiologically and genetically complex, resulting in a multitude of expression patterns and environmental effects that depends on genotype origin (Thomas and Howarth 2000; Thomas and Ougham 2014). Many sorghum varieties have leaves that senesce after grain maturity under normal field conditions. However, some genotypes remain green while containing greater stem carbohydrates and higher grain weight (Thomas and Howarth 2000; Duncan et al. 1981). Two well-known sources of stay-green for sorghum are B35 and E36-1 (Thomas and Howarth 2000; Thomas and Ougham 2014). B35 is derived from Ethiopian Durra and Nigerian Landraces and is widely used in genetic studies and breeding programs (Meru 2010; Thomas and Ougham 2014). Genetic mapping studies in populations derived from crosses with B35 have allowed the identification of four major stay-green QTLs (Stg2, Stg1, Stg3, and Stg4). Those QTLs are responsible for up to 54% of phenotypic variance. Stg2 and Stg1 are located on chromosome 3, Stg3 on chromosome 2 and Stg4 on chromosome 5 (Kim et al. 2005; Xu et al. 2000). E36-1 is derived from Ethiopian zera-zera germplasm which is also a source of sorghum resistance to sugarcane aphid (Thomas and Ougham 2014). Stay-green QTLs have been found to play an important role in reducing tillering and the size of upper leaves, which eventually reduces leaf transpiration and enhance grain yield under drought

(Borrell et al. 2014). The positive correlation between grain yield and green leaf area at maturity (GLAM) indicates the important contribution of stay-green to grain yield under post-anthesis drought (Borrell et al. 2000). A study conducted by Borrell and Hammer (2000) demonstrated that sorghum hybrids having the stay-green trait are photosynthetically superior compared to hybrids without the trait. Consequently, tandem selection for enhanced grain yield and stay-green is important for environments with post-anthesis drought (Borrell et al. 2014). Furthermore, stay-green appears to increase lodging resistance and disease resistance in sorghum hybrids (Borrell et al. 2000).

Relationship between Grain Yield and Stem Sugar in Sorghum

Little is known about the molecular mechanisms underlying the relationship between grain yield and stem sugar in sweet sorghum (Murray et al. 2009). Observations in some sweet sorghum germplasm having high biomass/ grain yield but low stem sugar indicate possible trade-off/ negative correlation between the two traits (Makanda et al. 2011; Mathur et al. 2017; Murray et al. 2008; Erickson et al. 2012; Nebie et al. 2013). However, this relationship may be genotype dependent (Gutjahr et al. 2013a; Gutjahr et al. 2013b). Murray et al. (2008) showed colocalization of plant height, grain yield and stem sugar in sweet sorghum, and were able to tandemly improve grain yield and stem sugar in the population. In addition, the sweet sorghum breeding program at CHIBAS has succeeded in creating new high yielding (> 4Tons ha^{-1}) with high stem sugar content (>16 °Brix). Of importance is that the negative correlation between gain yield was more pronounced among family selections (only 5 F_3 families selected out of 125) than for recurrent selection (18 S_2 families selected out of 160), leading to the conclusion that the trade-off between grain yield and soluble solids concentration may be genotype dependent. Bernal et al. (2014) demonstrated that

sugar accumulation can be the result of genotype X environment interaction, while Miller and Ottoman (2010) showed that water stress reduced plant height and grain yield (due to decreased interception of solar radiation) in sweet sorghum but had no effect on stem sugar content (Gutjahr et al. 2013a).

Conventional and Molecular Breeding Approaches in Crops

Plant breeding is a process that aims to improve desirable traits in a crop species by manipulating the genetic factors governing that trait (Brown et al. 2014). This practice started about 10,000 years ago with the domestication of plants by humans, which led to the birth of agriculture. During this period, farmers selected plants with more grain thus accumulating yield alleles each successive generation. However, this process was limited by G x E interaction and the resulting selections were not uniform/ consistent. However, in the 19th century, following the theory of evolution by Charles Darwin and the laws of inheritance and chromosome segregation by Gregor Mendel, scientists used their understanding of genes to generate hybrids with complementary characteristics. These two big discoveries are the foundation for today's modern genetics. Genetic markers are DNA sequences with specific location on a chromosome associated with variations in observed traits (Raza et al. 2016). Over the last two decades, advancements in DNA sequencing technology has accelerated marker discovery and facilitated their application in plant breeding for DNA finger printing, genetic diversity, QTL/ fine mapping, marker-assisted selection (MAS), genome-wide association studies (GWAS) and genomic selection. MAS has been widely applied in plant breeding for improvement of simple monogenic traits which are controlled by few major genes. However, MAS is inefficient for improvement of quantitative traits which are controlled by many small-effect genes. In addition, some QTLs are population-specific, thus

limiting the application of MAS across breeding populations of different genetic backgrounds. Unlike MAS, GS simultaneously estimates marker effects across the entire genome and calculates the GEBVs for individual plants (Meuwissen et al. 2001). Genomic selection uses dense molecular markers across the genome, combined with phenotypic data, to predict breeding values of individuals with genotypic data, but not phenotypic data. This method allows efficient selection for quantitative traits, while shortening the breeding cycle and reducing phenotyping costs (Picard 2015). Genomic selection has been successfully applied to improve quantitative traits such as yield and disease resistance in major crops such as wheat, maize, rice and barley (Bassi et al. 2016; Grenier et al. 2016; Guzman et al. 2016; Heslot et al. 2012; Ornella et al. 2017; Spindel et al. 2015; Song et al. 2017; Spindel and Iwata 2018). Genomic selection in sweet sorghum can be beneficial for improvement of quantitative traits such as yield and drought tolerance. However, application of GS for sweet sorghum improvement is nascent, compared to other cereal crops such as rice, maize and wheat (Kulwal 2016; Hunt et al. 2018; Oliveira et al. 2018).

Factors to Consider for Genomic Selection

Genomic selection is a method of marker-assisted selection in which genetic markers covering the whole genome of an organism are used so that all quantitative trait loci (QTLs) are in linkage disequilibrium with at least one marker (Meuwissen et al. 2001; Goddard and Hayes 2007). This technology has become an important tool in animal and plant breeding mainly due to the large number of single of nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) discovered by genome sequencing (Burgueno et al. 2012; Goddard and Hayes 2007). Many models can be used to test the prediction accuracy for genomic selection of a population, including

GBLUP, RRBLUP, Bayesian and machine learning. The accuracy of prediction models depends on several factors such as heritability, the size of the training set, statistical models, linkage disequilibrium and the marker density (Costa 2015; Picard 2015). These five factors have a positive correlation with prediction accuracy. Heritability expresses the proportion of the phenotypic variance that is explained by the genetic variance (Visscher et al. 2008) and can be considered as the most important factor when estimating prediction accuracy (Zhang et al. 2017). The prediction accuracy usually increases with increased trait heritability (Lian et al. 2014). This is corroborated by a study conducted by Zhang et al. (2017) where populations with higher trait heritability had higher genomic prediction accuracy values. This correlation is more pronounced for simple traits, but less for complex traits with low heritability values. In general, the prediction accuracy increases as the size of the training population increase, and it is related to the genetic relationship between the training and validation population (Liu et al. 2015). Similarly, prediction accuracy increases with an increase of the marker density, which means that high marker density is required to obtain accurate genomic prediction values, specifically for complex traits with low heritability (Zhang et al. 2017). Therefore, the distance between the markers should be less than 10-20 cM (Zhang et al. 2017). Another important consideration for genomic selection is G x E interaction, which is the physiological/ behavioral responses of individuals to different environmental conditions (Baye et al. 2011). Since prediction parameters vary depending on the testing environment, it is imperative to phenotype the training population across all target environments. Ultimately, genomic selection assigns a breeding value (GEBV) to a plant, which is an estimate of the ability of that individual to

produce superior offspring. This estimate is based on phenotypic measurement taken on the individual plant or more commonly, on several of its relatives. Generally, genomic selection is performed on a single trait, however, correlated traits can be used to predict complex traits with low heritability in a process called indirect or multi-trait genomic selection and trait-assisted selection (Fernandes et al. 2017; Velazco et al. 2019; Schulthess et al. 2015). Several studies have used simulated data to demonstrate how multi-trait models can increase the prediction ability of low-heritable traits or traits that are expensive or difficult to measure (Jia and Jannink 2012; Guo et al. 2014). More recently, other studies have assessed multi-trait GS in several crops using empirical phenotypic data (Schulthess et al. 2016; Rutkoski et al. 2016; Wang et al. 2016).

Sorghum Breeding Program in Haiti

The sorghum breeding program at Chibas, Haiti, was borne out a need for locally adapted, high-yielding and biotic tolerant varieties. Initial funding for the program was provided by the French government (ANR 2010-2014), and most recently by the Canadian (Global Affairs; 2014-2017) and U.S. governments (SMIL-USAID; 2017-2018). The long-term goal for the sweet sorghum breeding program is to apply new technologies (e.g. marker-assisted selection, genome-wide association studies and genomic selection) to accelerate the cultivar development process. From this breeding program, dual purpose lines (high grain yield (up to 4Tons ha^{-1}), high stem sugar concentration, (> 17°Brix)), non-photoperiodic, resistant to sugarcane aphid were released. These varieties are currently being tested for drought resistance (selecting for stay-green trait) and will be released for cultivation in drought-prone regions. These dual-purpose lines will serve as raw material to produce alcoholic beverage (rum), sorghum malt, pop sorghum and new products such as cereal bars. In addition, the

program seeks to evaluate the ratooning ability (up to three ratoons with no yield loss) of stay-green varieties (no tillering and high ratooning). Ratooning would allow farmers to do two cropping out of one planting (savings on land prep and weeding). A study conducted by Harlan and de Wet (1972) classified the primary gene pool, *Sorghum bicolor* (L.) Moench into five major races (bicolor, guinea, caudatum, kafir, and durra) and ten intermediate races, based on morphology. The varieties developed by CHIBAS are closer to the Caudatum group than any other group (Durra, Bicolor Asie, Bicolor AE, Guinea AAA, Guinea AO and Kafir). A study using 1,660 SNP markers revealed that there was less genetic diversity among CHIBAS varieties than that observed for Caudatum varieties in East Africa (Charles 2017). Traces of selection were identified between the CHIBAS lines and the Caudatum group of East Africa and between Chibas lines and a sorghum diversity panel that included 967 genotypes representing worldwide sorghum diversity. This panel also included the five main sorghum races, the ten intermediate races and the four wildtype races. Of these loci, two were found on chromosome 6, the chromosome on which the RMES1 gene conferring resistance to *Melanaphis sacchari* is located (Charles 2017).

Table 2-1. Sorghum nutrients per 100 g edible portions at 12% moisture

	Sorghum	Brown rice	Maize	Wheat	Pearl millet	Finger millet
Protein (g)	10.9	7.9	9.2	11.6	11	6
Fat (g)	3.2	2.7	4.6	2	5	1.5
CHO (g)	73	76	73	71	69	75
Crude Fiber (g)	2.3	1	2.8	2	2.2	3.6
Ash (g)	1.6	1.3	1.2	1.6	1.9	2.6
Energy (kcal)	329	362	358	348	363	336
Calcium (mg)	27	33	26	30	25	35
Iron (mg)	4.3	1.8	2.7	3.5	3	5
Thiamin (mg)	0.3	0.41	0.38	0.41	0.3	0.3
Niacin (mg)	2.83	4.31	3.57	5.05	2	1.4
Riboflavin (mg)	0.14	0.04	0.19	0.1	0.15	0.1

Adapted from Encyclopedia of Life Support Systems (EOLSS), 2004

Figure 2-1. The various utilities of sweet sorghum crop

Table 2-2. Water requirement of several crops

Plant	Water requirement (kg of water/kg of dry mater
Sorghum	332
Maize	358
Barley	434
Wheat	514

Adapted from Chaurasla 2015

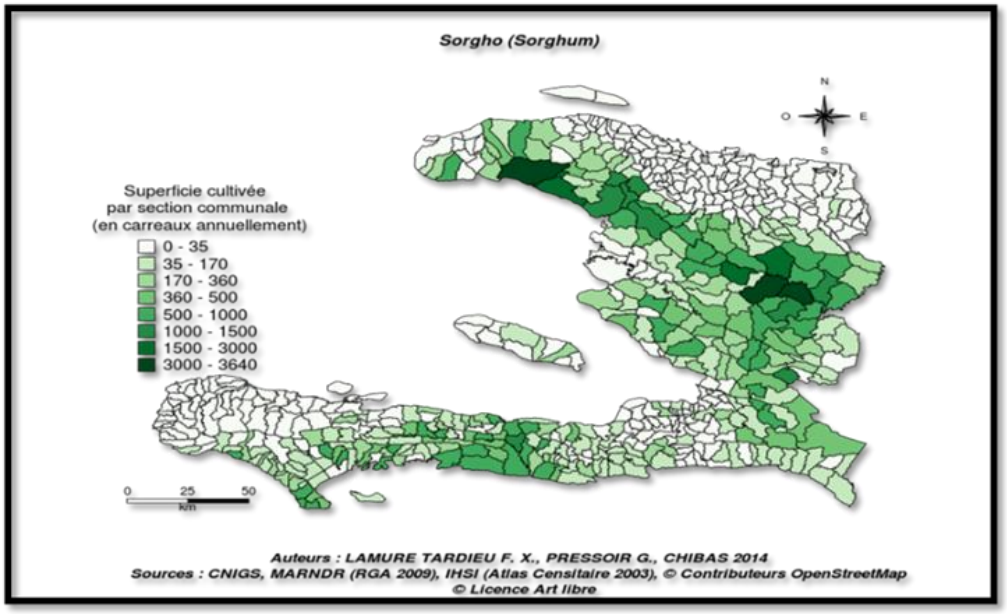

Figure 2-2. Sorghum cultivation area in Haiti (Pressoir and Lamure, personal communication)

CHAPTER 3
MATERIALS AND METHODS

Plant Material

A population of 250 lines developed at CHIBAS, which are offspring resulting from crossing between parents derived from CIRAD (French Agricultural Research Centre for International Development) and local varieties (Charles 2017) were evaluated under irrigated and water-stressed environments. The C1 variety was used as a check because of its high yield and resistance to sugarcane aphid. C1 is called Papèpichon, a variety developed and released by CHIBAS. It is currently the most widely cultivated sorghum variety in Haiti (over 90% of acreage).

Population Structure

The genetic structure of a population, which is defined as the total genetic diversity and its distribution within a population, determines a population's capacity to be improved or otherwise changed by selection (Gilleard and Redman 2016). The 250 sorghum lines could be divided into three populations. The lines in the first population were mainly derived from a phenotypic recurrent selection (Figure 3-1). The initial population was developed by crossing F_2 (BC_1 BF 95 11/110) (carrying ms3) x ICSV 25280. This population had a broader diversity, however many crosses that did not carry resistance to sugarcane aphid died. The second population included lines originating from a backcross selection where the initial population (S0) from the recurrent selection was subsequently crossed to: Coludo Nevado, 00-SB-FSDT-427, IS23563, WILEY, CIR-1/OG2-4G-1G-M-M, PCR-2>723C-1-M-1, Papèsèk, Dekabès (IS 23572) to introgress the resistance to sugarcane aphid (*Melanaphis sacchari*). The few

lines forming the third population were from a hybridization between Papèsèk (Centa S3) and Dekabès (IS 23572).

Experimental Design

The 250 genotypes were planted in a randomized complete block design (RCBD) into three fields. The study was conducted at one of the experimental stations of CHIBAS (Center for Bioenergy and Sustainable Agriculture) located in the Plaine du Cul de Sac (Croix des Bouquets-Haiti). Plaine du Cul de Sac is an area of about 360 km^2, with a length of 32 km and 25 km wide, it is bound on the North and South by high mountains, on the West by the Gulf of la Gonâve on the banks of which lies the Haitian capital Port-au-Prince and the plain of Arcahaie which extends to the West. This region is located at 26 meters from sea level, latitude 18°37'46.7"N, longitude 72°13' 33.9"W. The climate data of Croix des Bouquets during the period of experiment are presented in Table 3-1.

Application of Treatments

The three water regimes were assigned as irrigation, vegetative water stress and pre-flowering water stress. Each water regime was replicated four times. Each genotype was planted on 4 rows with a space of 70 cm between rows and 25 cm within rows. Each plot had a length of 2.8 m and a width of 3.5 m. There were 304 plots per replicate and 60 plants per plot as some parent lines were also planted during the experiment (Figure 3-2). The plant population density was 57,142 plants ha^{-1}, estimated using the formula fitting in the Equation 3-1 from Adebooye et al. (2006):

$$P_p = \frac{10000\ m^2\ X\ number\ of\ seeds\ per\ stand}{Product\ of\ spacing\ (m^2)} \quad (3\text{-}1)$$

The blocks under irrigation treatment were separated from those under drought stress by a 5-meter buffer, which was also a border (Figure 3-3). The experiment was planted from November 2017 to July 2018 and no fertilization was applied to the plants. Sowing was carried out in November 2017 and February 2018.

The fully irrigated treatment experiment was watered every 8 days throughout the growing season. For the pre-flowering water stress, the plots were watered every height days and irrigation was stopped just before flowering initiation (stopped at the vegetative stage). For the vegetative water stress, plots were grown on residual soil moisture (no irrigation and no rain). The latter is considered as a harsher stress as the soil has a slight salinity in addition to the water stress treatment. Irrigation was applied to the field by a system of manual irrigation where the water was pumped using an electric pump and brought to the field from canals prepared by farm workers. The soil moisture content and water deficit of sorghum were estimated. Plants were kept free from weeds by regular manual weeding. The soil moisture content was estimated during the period of the experiment using two different methods. The first method was a conventional approach based on the use of a hydro-sensor moisture probe, where the bars of the sensor are inserted into the ground and after a few seconds, the reading was recorded. The other method was based on a gravimetric approach where the soil moisture content is estimated by measuring the difference in weight of a soil sample before and after drying (Petropoulos et al. 2013).

Phenotypic Data Collection

Data were collected on phenological parameters (i.e. germination rate, plant vigor, homogeneity of plots, anthesis (days), flowering (days), and maturity (days)), growth parameters [plant height, stem diameter, total leaf number, stay-green (total

green leaf, leaf weight), and total stem number], soluble solids concentration and yield (stem weight, juice weight, and grain yield). Gemination rate was estimated one week after the sowing date as the number of holes having a seedling over the total number of holes. At the same time the plant vigor was estimated using a scale of 1 to 9, with 1 denoted highly vigorous and 9 represented poor vigor. Plot homogeneity was also estimated looking at the physical appearance of plants in the same plots using a similar scale as for plant vigor, with 1 representing highly homogeneous and 9 denoted highly heterogenous. Heading time and flowering time were observed and defined as the number of days from sowing until anthesis and 50% of the plants flowering in the plot, respectively. After flowering, plant height was measured from the ground to flag leaf sheath on 10 plants selected randomly in the two central rows using a graduated polyvinyl chloride pipe. Stem diameter was measured on ten plants randomly selected in the two central rows using a digital caliper. Two different measures of diameter, taken at the node, were recorded for each plant, one at the base of the panicle and the other at the base of the stem. Number of stems was counted on all plants in the two central rows, then extrapolated to obtain the total number of stems per plot. The number of leaves and green leaves were counted respectively on five plants randomly selected in the central rows during harvesting. Stem weight (leaf blade plus sheath) was measured by weighing the stems with and without leaves using a digital scale. Leaf weight was obtained as the difference between stem weight with leaf and without leaf. Generally, sorghum plots are harvested when 80% of the plants are at physiological maturity. For this study, the harvesting operation was performed 52 days after anthesis. In sorghum, the physiological maturity is confirmed by the presence of the black scar found in the

hilum or by crushing the grain of sorghum in the mouth to make sure that the grain is dry. Once mature, the panicles were harvested from all four rows of each plot. Then we checked the number of total panicles for each plot. The number of filled and unfilled panicles harvested for each plot was counted and weighted separately. Panicle weight was recorded as the sum of the weight of filled and unfilled panicles. Grain yield given in kg $plot^{-1}$ was estimated as 80% of the panicles weight harvested at grain physiological maturity, and then converted to tons ha^{-1}. After weighing the stems without leaves, the stems were milled, using a motor mill handled by two people placed at opposite sides. The weight and the volume of the extracted juice were measured using a scale and a graduated cylinder, respectively. Finally, the concentration of soluble solids (^{o}Brix) was measured with an Atago brand refractometer graduated from 0 to 33 %. This measure was determined by adding a few drops of subsample juice from each plot on the slide of the refractometer.

Generation of Genotypic Data

Young leaf tissues were collected from seedlings of each of the 250 progeny lines and their parents in a field at one of the experimental stations of CHIBAS. The leaf samples were harvested from the first five seedlings within each varietal plot using 96-well plates. The DNA was extracted using the Qiagen Plant DNAeasy DNA extraction kit. The DNA samples were shipped to Kansas State University for sequencing analysis using genotyping by sequencing technology (GBS). Genotyping by sequencing is a set of genomic analyses that allow the discovery of SNP markers through next generation sequencing (NGS) (Chung et al. 2017; He et al. 2014; Elshire et al. 2011). The genotyping by sequencing procedure includes several steps. First, the genomic DNA is digested with restriction enzymes, then ligated with barcode adapters, followed by PCR

amplification and sequencing analysis of the amplified DNA pool. Read alignment and SNP calling of genotyping by sequencing datasets can be performed using various bioinformatics pipelines (He et al. 2014). For this study, the Illumina Nextseq500 with 384X Sample Multiplexing and *Ape*KI Restriction Enzyme was used to generate the SNPs (Elshire et al. 2011). About 160,066 SNPs were identified from 1283 individuals based on the depth of the sequencing (number of reads >2), missing data and the frequency of the alleles.

Phenotypic Data Analysis

Twelve traits of the 250 sorghum genotypes were evaluated in fully irrigated, vegetative water stress, and pre-flowering water stress environments. The observations that fell outside of the interval [Q_1-1.5 X (Q_3-Q_1); Q_3 + 1.5 x (Q_3-Q_1)] were considered as outliers (about 40 genotypes) and treated as missing data with Q1, Q2 and Q3 representing the first quartile, second quartile and third quartile, respectively in the boxplot distribution for each trait. To meet the assumptions of normality, heading time, maturity and juice weight data were log transformed. An appropriate univariate linear mixed model was fitted using restricted maximum likelihood estimation with the Asreml-r 4.1 package in R software (Butler, 2018) to extract the best linear unbiased estimates (BLUEs) using the Equation 3-2:

$$\mathrm{y} = \mu + X_I b_I + X_G b_G + X_{(I\times G)} b_{(I\times G)} + Z_B u_B + Z_P u_P + \varepsilon \quad (3\text{-}2)$$

where y is the response variable, μ is a common intercept, X_I, X_G, $X_{(I\times G)}$, Z_B, and Z_P are the incidence matrices for the respective effects of irrigation, genotypic, interaction irrigation × genotypic, block, and main plot. b_I, b_G, and $b_{(I\times G)}$ are the respectively fixed effects for irrigation, genotype, and irrigation × genotype interaction. u_B and u_P are the

respectively random effects for block and main plot being $u_B \sim MVN(0, I_4\sigma_B^2)$ where I_4 an identity matrix of order four and σ_B^2 the variance component for the block effect, and being $u_P \sim MVN(0, I_{12}\sigma_P^2)$ where I_{12} an identity matrix of order twelve and σ_P^2 the variance component for the main plot effect, ε is the residual effect being $\varepsilon \sim MVN(0, I_N\sigma_e^2)$ where N is the total number of observations and *MVN* (multivariate normality test). After fitting the model, the eBLUEs (estimated best linear unbiased estimates) were obtained for the genotypic effects (b_G) which were later used in the genomic prediction study. About 42 genotypes were excluded for this analysis, two of them lacked genotypic information (C35_192_1 and D14_311_2), the other forty genotypes have missing information for at least one trait for all replications in either one of the treatments. Therefore, there was no way to model the effect of the trait for those genotypes in a given treatment using only the phenotypic information.

Computation of Heritability

The genomic heritability of the twelve traits under the different treatments was estimated using the univariate linear mixed model as in Equation 3-3:

$$y = \mu + X_I b_I + Z_{(I\times G)}u_{(I\times G)} + Z_B u_B + Z_P u_P + \varepsilon \quad (3\text{-}3)$$

where y is the response variable, μ is a common intercept, X_I is the incidence matrix for the irrigation effect. b_I is the fixed effect for the irrigation effect. $Z_{(I\times G)}$, Z_B, are Z_P are the respectively random effects for the irrigation × genotype interaction, block, and main plot. $u_{I\times G}$ is the random effect for the interaction irrigation × genotype where $u_{I\times G} \sim MVN(0, D_I \otimes K)$ being $\D_{I}$ a diagonal variance-covariance structure (heterogenous diagonal and off-diagonal equal zero) of dimension three, K a genomic relationship matrix built as in Van Raden (2008), and $\sigma^2_{(I\times G)}$ is the variance component for the

interaction term confounding with the genotypic variance. u_B and u_P are the respective random effects for block and main plot being $u_B \sim MVN(0, I_4\sigma_B^2)$ where I_4 an identity matrix of order four and σ_B^2 the variance component for the block effect, and being $u_P \sim MVN(0, I_{12}\sigma_P^2)$ where I_{12} an identity matrix of order twelve and σ_P^2 the variance component for the main plot effect, ε is the residual effect being $\varepsilon \sim MVN(0, D_E \otimes I_{Ng})$ being $\D_{I}$ a diagonal variance-covariance structure (heterogenous diagonal and off-diagonal equal zero) with of dimension three and N_g is the number of genotypes times the number of blocks. The heritability was computed for each irrigation level as in Equation 3-4:

$$h^2 = \sigma_G^2/(\sigma_G^2 + \sigma_E^2) \quad (3\text{-}4)$$

where σ_G^2 is the variance component for the given irrigation level in D_I matrix, and σ_E^2 is the variance component for the given irrigation level in D_E matrix using the Asreml-r 4.1 package in R software (Butler, 2018). Using the phenotypic data, the broad-sense heritability was also computed for each irrigation treatment using the same formula as in Equation 3-4, whereσ_G^2 is the variance component of the genotypes for the given irrigation level and σ_E^2 is the variance component of the replicates for the given irrigation treatment using the 'lmer' function in the R package lme4, v1.1-7 (Bates et al. 2016).

Genomic Data Analysis

The Single nucleotide polymorphisms (SNPs) were filtered using VCFtools package in R software based on minimum allele frequency of 1%, biallelic markers, maximum population missing data of 30%. After filtering, a total of 29,072 markers and 208 individuals were used in the genomic selection analysis.

Population Structure Analysis

The principal component analysis (PCA) and k-means clustering analysis were combined to evaluate the population structuring using the R software. The visualization of genetic differentiation among genotypes was performed by drawing a PC\ score plot in which for each genotype the value of PC1 and PC2 are plotted against each other. The population structure was also inferred using a heatmap based on genomic pedigree information.

Fitting of Genomic Selection Models

Prediction ability of genomic selection was evaluated using four models: BRR, Bayes A, Bayes B, Bayes C. These models differ with respect to assumptions about the marker effects. Bayesian methods depend on the inference of marker effects where the sum of estimated effects identified the genetic value of an individual (Ferrão et al. 2017). These models differ in the priors used for the regression while having a Gaussian distribution with a mean vector represented by a regression on the markers and a common residual variance (Ferrão et al. 2017).

- BRR model: The Bayesian ridge regression is the Bayesian version of RRBLUP. This model assumes that all marker effects are normally distributed and have identical genetic variance as in Equation 3-5 (Meuwissen et al. 2001):

$$g_i \sim N(0, \sigma_g^2),\ \sigma_g^2 = \sigma_{g1}^2 = \sigma_{g2}^2 = \cdots = \sigma_{gm}^2 \quad (3\text{-}5)$$

The prior distribution of the genetic variance of the markers follows a scaled inverted chi-squared distribution variance. The SNP effects are random, normally distributed, have identical variance and is given by the Equation 3-6:

$$\sigma_{gi}^2 | v_g,\ S_g \sim v_g S_g \sigma X_{vg}^{-2} \quad (3\text{-}6)$$

Where v_g is degree of freedom and S_g a scale (Meuwissen et al. 2001; Mota et al. 2018).

- Bayes A model: Unlike BRR, Bayes A assumes that the markers have different variances, are normally distributed and follow a scaled inversed χ2 distribution with degrees of freedom v_a and scale parameter S_a^2 as in the Equation 3-7:

$$S_a^2 = \frac{\tilde{\sigma}_a^2(v_a - 2)}{v_a}, \qquad \tilde{\sigma}_a^2 = \frac{\tilde{\sigma}_s^2}{(1-\pi)\sum_{k=1}^{K} 2p_k\,(1-p_k)} \tag{3-7}$$

Where pk corresponds to the allele frequency of the kth SNP, $\tilde{\sigma}_a^2$ is the variance of a given marker and $\tilde{\sigma}_s^2$ is the additive genetic variance that is explained by the SNPs (Resende et al. 2012; Habier et al. 2011; Meuwissen et al. 2001).

- Bayes B model: This method uses a prior having a high density (π) at $\sigma_{gi}^2 = 0$ and an inverted chi-square distribution for $\sigma_{gi}^2 > 0$ and assumes that marker effects have identical and independent mixture distributions with the equation 3-8:

$$P(\sigma_{gi}^2, gi|y^*) = p(\sigma_{gi}^2|y^*) \text{ X } p\ (gi|\sigma_{gi}^2, y^*) \tag{3-8}$$

where y* denotes the data y corresponding to the mean and all other genetic effects except gi, $\sigma_{gi}^2 = 0$ with probability π and gi represents the genetic effects of the SNPs at the i^{th} 1-cM segment (Meuwissen et al. 2001).

- Bayes C model: This model assumes the normalization of the distribution effects of one fraction of the SNPs (π) and that the other fraction of SNPs (1- π) has zero effects and is expressed by this equation:

$$y = 1\mu + ZMQq + e \tag{3-9}$$

where M is the design matrix of scaled SNP genotypes; Q is a diagonal matrix with indicators on the diagonal that are 1 if the SNP has an effect (with prior probability π) and 0 if it has no such effect (with prior probability $(1 - \pi)$); q is a vector of SNP effects

(q_j) assumed to be normally distributed, i.e. qj~N(0,σ2q)qj~N(0,σq2) with probability π and 0 otherwise (Habier et al. 2011; Iheshiulor et al. 2017).

All the models were fitted using the Bayesian Generalized Linear Regression (BGLR) package (Pérez and de Los Campos, 2014; Oliveira et al. 2018). For each model, the sampler algorithm was running for a total of 30,000 iterations, with 5,000 discarded as burn-in. We used a cross-validation procedure to compare the models, where the whole population was randomly partitioned into two subsets (75% training set and 25% testing set). The cross-validation (2-fold) was repeated 100 times to obtain accurate estimates of the average prediction correlation. For each cross-validation set, we began by fitting the genomic selection models on 156 genotypes (training set), to estimate marker effects based on genotypic and phenotypic information. These marker effects then provided genomic estimates of the breeding values of the remaining 52 individuals (testing set), based only on genotypic information. The principal component analysis was included into the genomic selection models to correct for a population structure. Finally, the Pearson's correlation between the genomic estimated breeding values and the breeding values estimates provided estimates of the prediction accuracy of the models. Thus, two-step approaches were used; the phenotypic model was first fitted, then the best linear unbiased estimate values (BLUEs) were used in the genomic prediction studies. The genomic prediction referred to traditional scenarios when a common treatment effect is extracted across the three conditions. This treatment effect is interpreted as the estimated breeding value across environments, where the higher values ultimately mean more 'stable' genotypes. We then performed a stratified analysis which involved predictions for each treatment. Here, we present a scenario where the

breeder would like to train a model in one condition to predict genotype performance in a new one. In addition, different multi-trait prediction models were computed for grain yield in order to test the improvement of the prediction accuracy across environments. For this analysis, the models were re-trained in the irrigated scenario and predicted for the water stress environments. The procedure is similar to across-environment scenarios except that correlated traits such as leaf weight, plant height, stem weight, total green leaf were included as auxiliary traits in the model. These traits were chosen based on their positive, significant, consistent correlation and genetic covariance with grain yield across the three environments. The computation of the genomic prediction analysis was performed using the HiPerGator cluster at University of Florida.

Table 3-1. Temperature and rainfall data of Croix des Bouquets during the time of the experiment

	Temperature (oC)			Rainfall	
Months	Min	Max	Average	Rain (mm)	Days
Nov-17	22	31	26	41.38	18
Dec-17	22	28	24	1.71	2
Jan-18	20	29	23	10.18	12
Feb-18	19	29	23	0.55	1
Mar-18	20	29	24	13.12	12
Apr-18	21	29	25	29.06	23
May-18	22	30	25	9.47	6

Adapted from WorldWeatherOnline.com

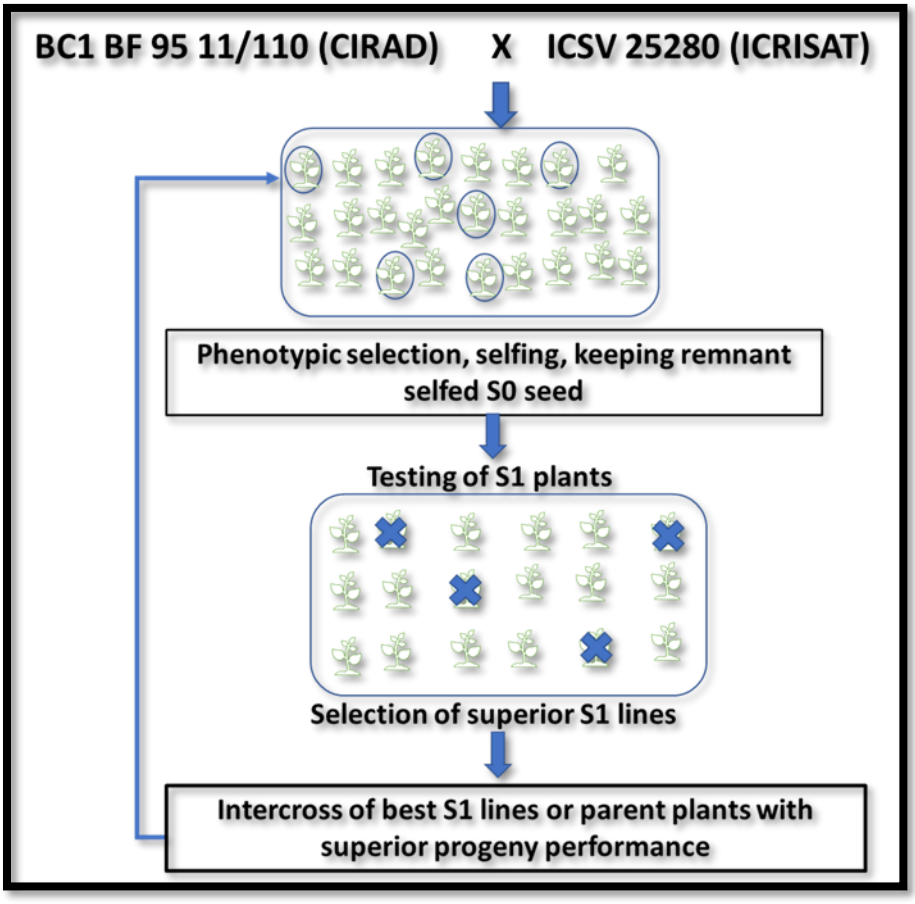

Figure 3-1. Phenotypic recurrent selection diagram of the sorghum breeding program from CHIBAS

3.5	1	3.5	1	3.5	1	3.5	1	3.5	1	3.5	1	3.5	1	3.5
1		2		3		4		5		6		7		8
Bordure		Bordure		Bordure		Bordure		Bordure		Bordure		Bordure		Bordure
T		T		T		T		T		T		T		T
E001		E035		E069		E103		E137		E171		E205		E239
E002		E036		E070		E104		E138		E172		E206		E240
E003		E037		E071		E105		E139		E173		E207		E241
E004		E038		E072		E106		E140		E174		E208		E242
E005		E039		E073		E107		E141		E175		E209		E243
E006		E040		E074		E108		E142		E176		E210		E244
E007		E041		E075		E109		E143		E177		E211		E245
E008		E042		E076		E110		E144		E178		E212		E246
E009		E043		E077		E111		E145		E179		E213		E247
E010		E044		E078		E112		E146		E180		E214		E248
E011		E045		E079		E113		E147		E181		E215		E249
E012		E046		E080		E114		E148		E182		E216		E250
T		T		T		T		T		T		T		T
E013		E047		E081		E115		E149		E183		E217		E251
E014		E048		E082		E116		E150		E184		E218		E252
E015		E049		E083		E117		E151		E185		E219		E253
E016		E050		E084		E118		E152		E186		E220		E254
E017		E051		E085		E119		E153		E187		E221		E255
E018		E052		E086		E120		E154		E188		E222		E256
E019		E053		E087		E121		E155		E189		E223		E257
E020		E054		E088		E122		E156		E190		E224		E258
E021		E055		E089		E123		E157		E191		E225		E259
E022		E056		E090		E124		E158		E192		E226		E260
E023		E057		E091		E125		E159		E193		E227		E261
T		T		T		T		T		T		T		T
E024		E058		E092		E126		E160		E194		E228		E262
E025		E059		E093		E127		E161		E195		E229		E263
E026		E060		E094		E128		E162		E196		E230		E264
E027		E061		E095		E129		E163		E197		E231		E265
E028		E062		E096		E130		E164		E198		E232		E266
E029		E063		E097		E131		E165		E199		E233		E267
E030		E064		E098		E132		E166		E200		E234		E268
E031		E065		E099		E133		E167		E201		E235		E269
E032		E066		E100		E134		E168		E202		E236		E270
E033		E067		E101		E135		E169		E203		E237		E271
E034		E068		E102		E136		E170		E204		E238		E272
T		T		T		T		T		T		T		T
tampon		tampon		tampon		tampon		tampon		tampon		tampon		tampon

Figure 3-2. View of one block

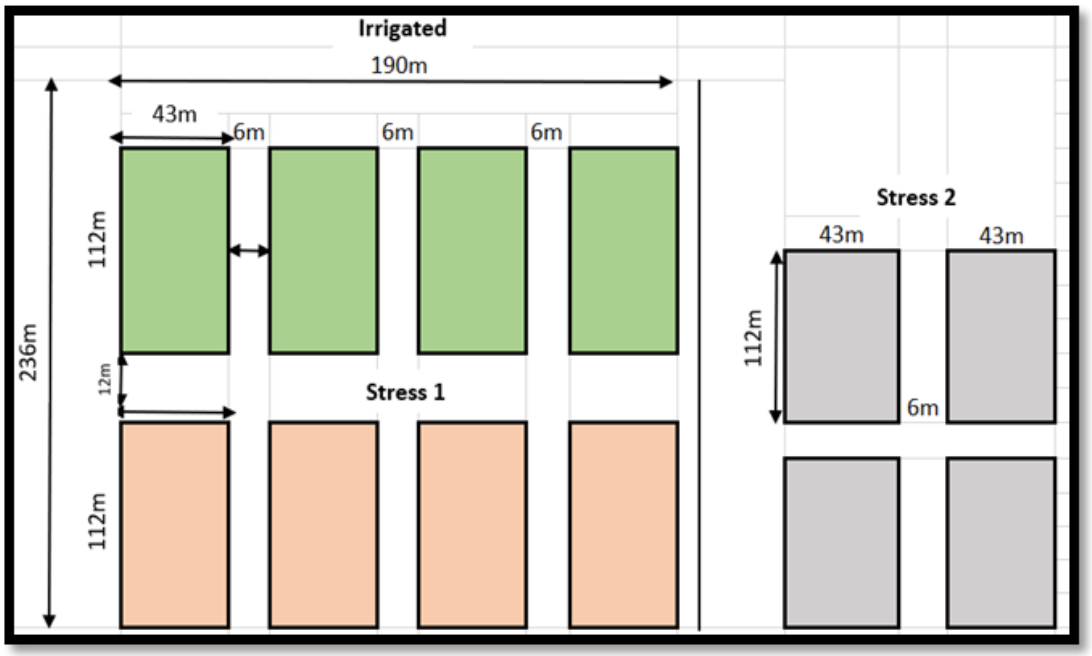

Figure 3-3. View of the experimental design

CHAPTER 4
RESULTS

Phenotypic Characteristics of the Population

The distribution of the phenotypic values varied substantially across the three environments, which showed that the impact of water stress was quite different among the genotypes. As might be expected, the performance of the lines was substantially higher in the irrigated environment than in the two water stress conditions. Generally, the genotypes were more affected by the vegetative water stress environment than by the pre-flowering water stress (Table 4-1, Figures 4-1, 4-2). The vegetative water stress was severe and led to substantial reductions in almost all the traits as compared to the values observed under fully irrigated and pre-flowering water stress conditions. The pre-flowering water stress environment was moderate, with a modest reduction in the performance of the genotypes as compared to the values obtained in the fully irrigated environment. The difference in the severity of the stress may be attributed to the exposure time and the application time of the water stress. The combined analyses of variance showed significant genotype x treatment interaction for all the traits (Table 4-2). The environment (treatment x replicates) mean squares were highly significant for all traits, meaning that the field was probably heterogeneous. Genotypic differences among the replicates were non-significant for the traits except for total leaf number and soluble solids concentration. The interaction between genotypes x treatment was highly significant for all traits (Table 4-2).

Heritability

Genetic characterization and selection of specific traits mainly depends on the magnitude of the heritability of the trait. The heritability can change over time due to changes in variance of genetic values. Thus, we calculated the broad-sense and genomic heritabilities for all twelve traits in the three environments. The estimated broad-sense heritability ranged from 0.06 to 0.43 for vegetative water stress, from 0.31 to 0.51 for the pre-flowering water stress, and from 0.19 to 0.71 for the irrigated environment (Table 4-3). Overall, the highest broad-sense heritability estimates were observed for the irrigated condition, where the highest values were obtained for days to heading (0.71), maturity (0.71), plant height (0.43) and soluble solids concentration (0.43). The genomic heritability estimates were higher than the broad-sense heritability, with the highest values obtained for heading (0.82), maturity (0.82), total soluble sugar (0.61), grain yield (0.61) and juice weight (0.60) in the pre-flowering water stress condition. Overall, the values varied from 0.16 to 0.65 for vegetative water stress, 0.29 to 0.82 for the irrigated condition and 0.35 to 0.82 for pre-flowering water stress (Table 4-3).

Phenotypic Correlation

Tandem selection of multiple traits in superior genotypes requires positive and consistent correlations among those traits. Significant phenotypic correlations were found among almost all pairs of traits across the environments (Figures 4-3 to 4-5). Phenotypic correlations varied in terms of magnitude across the environments. Grain yield was positively and significantly correlated with all eleven traits in the irrigated environment. For the water stress environments

(vegetative and pre-flowering water stress), no significant correlations were observed between grain yield, heading and maturity. The strongest positive correlation between grain yield and leaf weight was observed for the pre-flowering water stress environment, followed by the vegetative water stress environment. Significant positive correlations between grain yield and soluble solids concentration were observed for all three environments. The linear relationship between maturity and heading in the irrigated and pre-flowering water stress environments can be explained by the fact that in these environments, the plots were homogeneous, and the genotypes were harvested about 45 days after heading (Figures 4-3 to 4-5).

Genotypic Results

Population Structure

A hypothetical number of subpopulations between three and four were tested using a combination of PCA and k-means clustering analyses. Based on pedigree information of the sorghum breeding lines from CHIBAS, the number of subpopulations is likely three (Charles 2017). The lines coded with the letter C are from the first generation of phenotypic recurrent selection and those coded with letter D are from the second generation (Figure 4-6). From the results of the two-dimensional plot obtained in the PCA analysis, the first principal component (Dim1) explained 52.2% of the variance and the second component (Dim2) accounted for 14.5 % of the variance (Figure 4-6).The heatmap also revealed that the strength of genomic relationships among the different breeding lines is relatively high (Figure 4-7).

Chromosomal Distribution of SNPs

The distribution of 29,072 SNPs across the ten sorghum chromosomes was checked after filtering for missing values (30%) and minor allele frequency (1%). The average number of SNPs per chromosome was about 2,907, with the highest number of SNPs found in chromosome 3, and the lowest in chromosome 9 (Figure 4-8). This distribution showed that the SNPs covered the entire sorghum genome and was sufficient for the genomic prediction analysis.

Prediction Accuracy of the Models within Environment

Prediction accuracy for the models was estimated by performing correlation analysis between the breeding values predicted by genomic selection and the estimated breeding values. Although the models tested differ mainly in their assumptions about the variances of markers effects, they yielded similar predictive abilities for the twelve traits. Even though differences between the models were small, Bayes B showed a modest improvement over the three other models for all the traits except for soluble solids concentration (Figure 4-9). The predictive values obtained were medium (0.49) to high (0.70) and, in general high prediction accuracy values are practical for genomic selection. On average, the highest predictive abilities were obtained for total green leaf (0.70), total soluble solids concentration (0.68), leaf weight (0.68), total leaf number (0.68), stem weight (0.65), stem diameter (0.64), juice weight (0.63) and heading (0.60) while grain yield had the lowest value (0.49) (Table 4-4).

Computational Feasibility

Considering the relatively small differences found between the models, differences in computation time may be an important factor that determines the

model of choice in practical applications, especially for programs with low computational resources. On average, BRR (14 minutes) was the fastest while, Bayes B, that is based on variable selection required more time (21 minutes) to run each replicate during cross-validation using the HiPerGator cluster (Figure 4-10). Using a single computer, these models can take between 13 to 16 hours to run (Mota et al. 2018).

Prediction Accuracy of the Models across-Environments

As the BRR had the least computation time, it was used to test the predictive performance across different cross validation scenarios (Figure 4-11). Overall, cross-validation using within-environment scenarios yielded higher predictive performances than across-environments scenarios involving vegetative water stress environment (Figure 4-11, scenarios 1-3 and 6-9). Most scenarios that involved vegetative water stress (water 2) had lower predictive accuracies (Figure 4-11, scenarios 6,7,8,9). However, soluble solids concentration, leaf weight, total green leaf and total leaf number showed high predictive capacity across all scenarios (Figure 4-11). When the model was trained in the fully irrigated treatment to predict the pre-flowering water stress, high predictive values were observed for all the traits except for grain yield and total stem number (Figure 4-11, scenario 5). Similarly, high predictive values were also observed when the model was trained in the pre-flowering water stress environment to predict the irrigated condition (Figure 4-11, scenario 4).

Multi-trait Prediction Models

Predictive accuracy for grain yield was low for almost all the scenarios (Figure 4-11). Improvement of prediction accuracy was attempted by

including correlated traits, e.g. leaf weight, plant height, total green leaf, and stem weight as auxiliary traits in the models (Tables 4-4; Figures 4-3 to 4-5). Those traits had highly significant and consistent correlation with grain yield across the three environments. High genetic covariances were also observed between grain yield and the auxiliary traits in the drought environments (Table 4-7). The prediction accuracy varied from 0.36 to 0.37 for the pre-flowering water stress and from 0.13 to 0.16 for the vegetative water stress environment. The results suggested no improvement of prediction accuracy of grain yield across-environment based on multi-trait models.

Table 4-1. Summary statistics of the 12 phenotypic traits presented by treatment

Traits	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Pre-flowering water stress						
Plant height	114	140	153	152.5	164.7	192.5
Stem diameter	12	14.18	15.47	15.44	16.15	18.24
Heading	65	78	84	84.88	90	123
Maturity	110	123	129	129.9	135	168
Total leaf number	7	22	27	26.68	31	47
Total green leaf	4	16	19	19.31	31	47
Total stem number	3	19	21	20.74	24	31
Soluble solids con.	4.67	9.75	11.75	11.74	13.5	17.38
Juice weight	0.18	1.17	1.73	1.75	2.88	4.89
Stem weight	0.99	8.65	10.52	10.49	12.94	21.05
Leaf weight	0.2	4.5	6.39	6.29	7.84	12.04
Grain yield	1.67	3.32	3.71	3.74	4.15	5.22
Vegetative water stress						
Plant height	105.1	12.2	129.6	130.5	137.6	171
Stem diameter	14.09	16.97	17.69	17.72	18.43	20.6
Heading	74	91	96	97.71	102	143
Maturity	108	140	146	145.7	152	185
Total leaf number	15	25	29	29.34	33	46
Total green leaf	12	19	22	22.75	26	38
Total stem number	8	13	16	15.97	18	29
Soluble solid con.	5.85	10.89	12.02	12.22	13.14	18.2
Juice weight	0.29	1.46	2.22	2.28	2.94	5.9
Stem weight	2.48	4.92	6.28	6.57	8.16	14.8
Leaf weight	1.12	2.72	4.01	4.11	4.98	12.7
Grain yield	0.37	2.03	2.27	2.33	2.61	3.92
Irrigated						
Plant height	122.5	151.1	162.2	162	172.7	209.2
Stem diameter	12.75	14.88	15.54	15.5	16.22	18.73
Heading	62	75.5	80	80.1	84	101
Maturity	107	120.5	125	125.1	129	146
Total leaf number	13	28	32	31.8	36	50
Total green leaf	9	20	24	23.99	28	45
Total stem number	13	23	25	24.36	26	32
Soluble solid con.	9	12.5	14.5	14.13	15.5	18
Juice weight	0.42	1.65	2.07	2.11	2.57	3.96
Stem weight	4.19	11.12	13.11	12.95	15.05	20.61
Leaf weight	3.88	6.98	8.62	8.51	9.82	13.86
Grain yield	3.25	4.12	4.42	4.44	4.73	5.76

Plant height (cm), stem diameter (mm), heading and maturity (days), total leaf number, total green and total stem number (number), soluble solids concentration (^{o}brix), Juice weight (kg/m^2), stem weight and leaf weight (kg/ha), grain yield (T/ha).

Table 4-2. Mean squares from the combined analyses of variance for evaluated traits across environments

	Source of variation						
	Genotypes (G)	**Treatment (T)**	**Replicates (R)**	**G x T**	**G x R**	**T x R**	**Residuals**
df	248	2	3	478	744^{ns}	6	1426
PH	1854***	236070***	12578***	404***	246^{ns}	3102***	370734
SD	8.40***	1550.42***	158.70***	2.44**	1.9^{ns}	76.34***	2
H	396***	76600***	1147***	73***	42^{ns}	438***	44
M	354***	95052***	161*	245***	50^{ns}	419***	55
TLN	257.6***	6017.3***	806***	69.1***	56.8***	603.2***	50.2
TGN	194.7***	5145***	543.5***	53.2***	45^{ns}	429***	40.7
TSN	74.3***	62202.9***	144.5***	34.6***	21.1^{ns}	276.4***	22.6
TSS	31.7***	3450.3***	180.5***	6.1***	4.1*	96.9***	3.6
JW	2.34***	287.88***	16.33***	0.86***	0.56^{ns}	4.21***	0.58
SW	57.9***	7684.6***	182.1***	15.8***	8.51^{ns}	157.5***	9
LW	25.19***	2564.89***	9.84^{ns}	7.78***	5.06^{ns}	63.60***	5.25
GY	2.15***	562.51***	9.31***	0.95***	0.58^{ns}	13.46	0.57

*, **, *** Significant at the 0.05, 0.01 and 0.001 alpha levels, respectively.

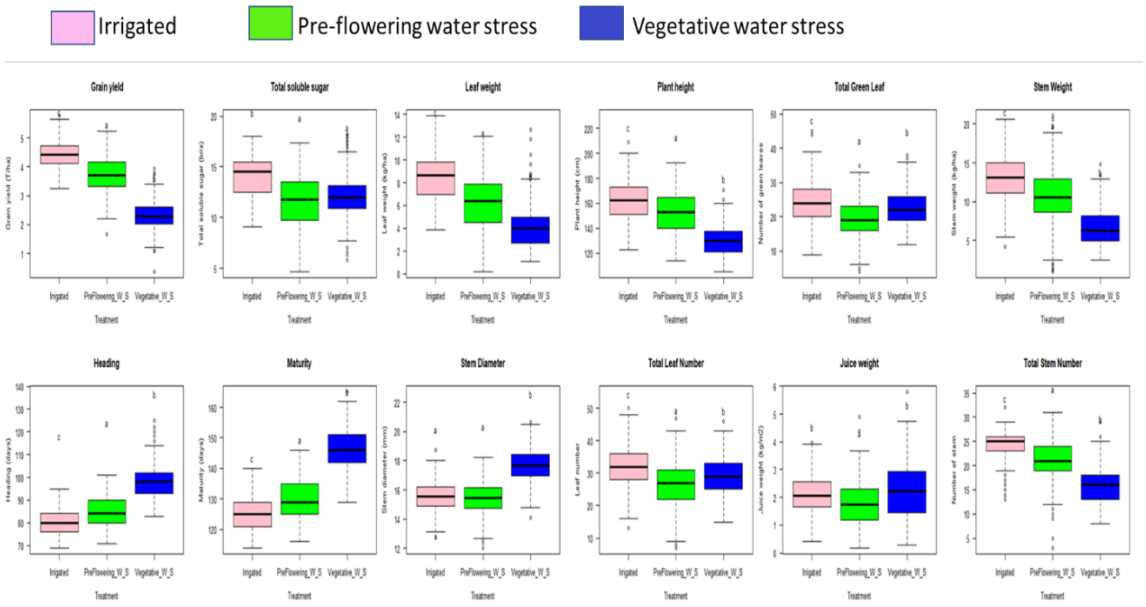

Figure 4-1. Graphical distribution of the traits by treatments

Table 4-3. Genetic and phenotypic variance, Broad-sense heritability, Standard error and genomic heritability by treatment

Traits	Genetic variance	Residual	Phenotypic variance	Broad-sense heritability	Standard error	Genomic heritability
			Pre-flowering water stress			
PH	159.27	355.7	514.97	**0.31**	3.97	**0.53**
SD	0.7	1.59	2.29	**0.31**	0.04	**0.35**
H	40.13	38.48	78.61	**0.51**	0.52	**0.82**
M	40.12	38.47	78.59	**0.51**	0.52	**0.82**
TLN	29.67	42.61	72.28	**0.41**	0.58	**0.48**
TGL	19.7	34.07	53.77	**0.37**	0.45	**0.4**
TSN	10.62	22.88	33.5	**0.32**	0.29	**0.43**
TSS	4.6	4.76	9.36	**0.49**	0.08	**0.61**
JW	0.32	0.65	0.97	**0.33**	0.05	**0.6**
SW	7.84	10.44	18.28	**0.43**	0.18	**0.54**
LW	3.22	5.14	8.36	**0.39**	0.1	**0.5**
GY	0.37	0.75	1.12	**0.33**	0.04	**0.61**
			Vegetative water stress			
PH	114.11	149.39	263.5	**0.43**	1.99	**0.43**
SD	0.51	2.97	3.48	**0.15**	0.05	**0.27**
H	33.95	84.55	118.5	**0.29**	0.96	**0.6**
M	8.12	119.72	127.84	**0.06**	0.42	**0.65**
TLN	14.23	50.39	64.62	**0.22**	0.52	**0.17**
TGL	9.86	42.35	52.21	**0.18**	0.39	**0.17**
TSN	5.76	28.82	34.58	**0.16**	0.25	**0.22**
TSS	1.6	2.88	4.48	**0.36**	0.06	**0.32**
JW	0.08	0.28	0.36	**0.23**	0.04	**0.19**
SW	2.78	7.09	9.87	**0.28**	0.11	**0.23**
LW	1.6	4.51	6.11	**0.26**	0.08	**0.22**
GY	0.06	0.39	0.45	**0.14**	0.04	**0.16**
			Irrigated			
PH	185.23	250.47	435.7	**0.43**	3.19	**0.51**
SD	0.63	1.41	2.04	**0.31**	0.04	**0.29**
H	30	12.21	42.21	**0.71**	0.14	**0.82**
M	30	12.21	42.21	**0.71**	0.14	**0.82**
TLN	21.76	64.13	85.89	**0.25**	0.65	**0.28**
TGL	18.15	50.32	68.47	**0.26**	0.53	**0.33**
TSN	4.43	15.73	20.16	**0.22**	0.16	**0.31**
TSS	2.71	3.62	6.33	**0.43**	0.07	**0.45**
JW	0.24	0.73	0.97	**0.24**	0.03	**0.37**
SW	6.24	9.09	15.33	**0.41**	0.13	**0.5**
LW	2.18	5.85	8.03	**0.27**	0.08	**0.37**
GY	0.13	0.55	0.68	**0.19**	0.03	**0.35**

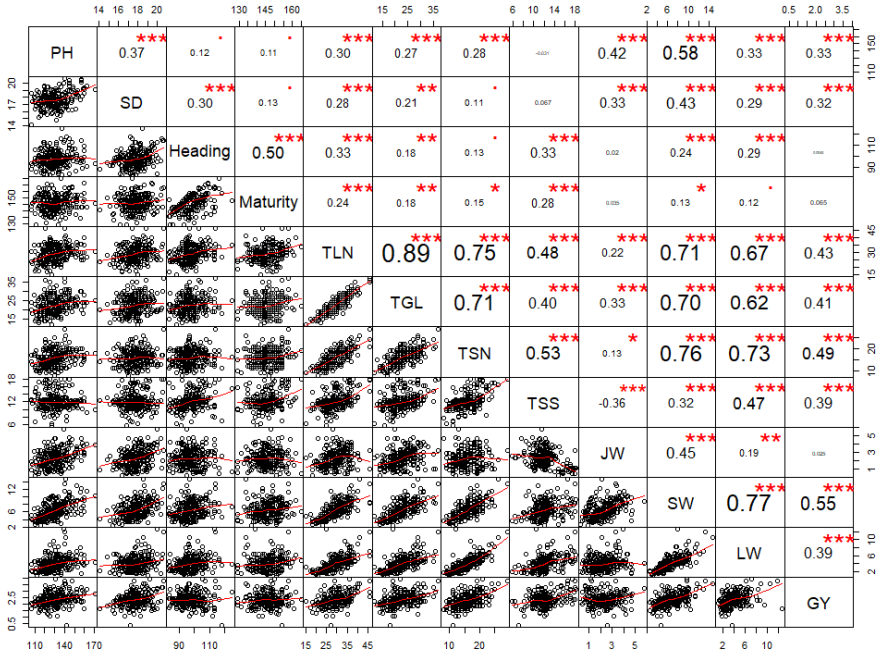

Figure 4-2. Phenotypic correlation between the twelve traits in vegetative water stress condition

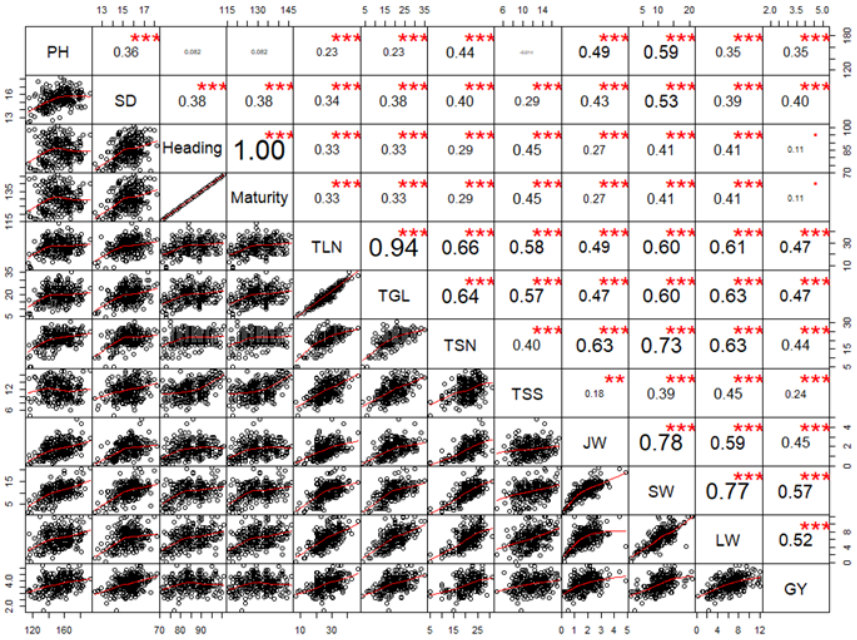

Figure 4-3. Phenotypic correlation between the twelve traits in pre-flowering water stress condition

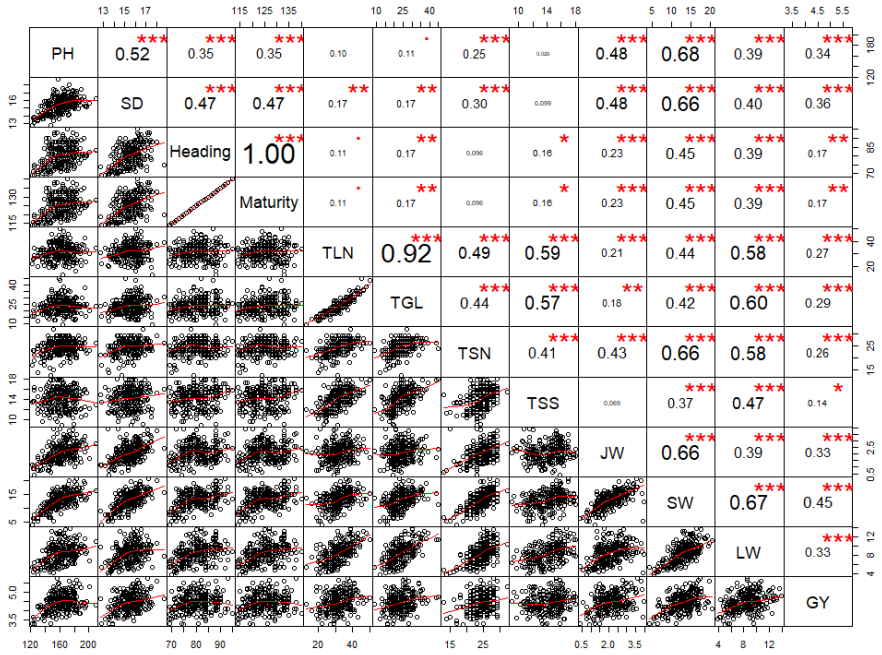

Figure 4-4. Phenotypic correlation between the twelve traits in irrigated condition

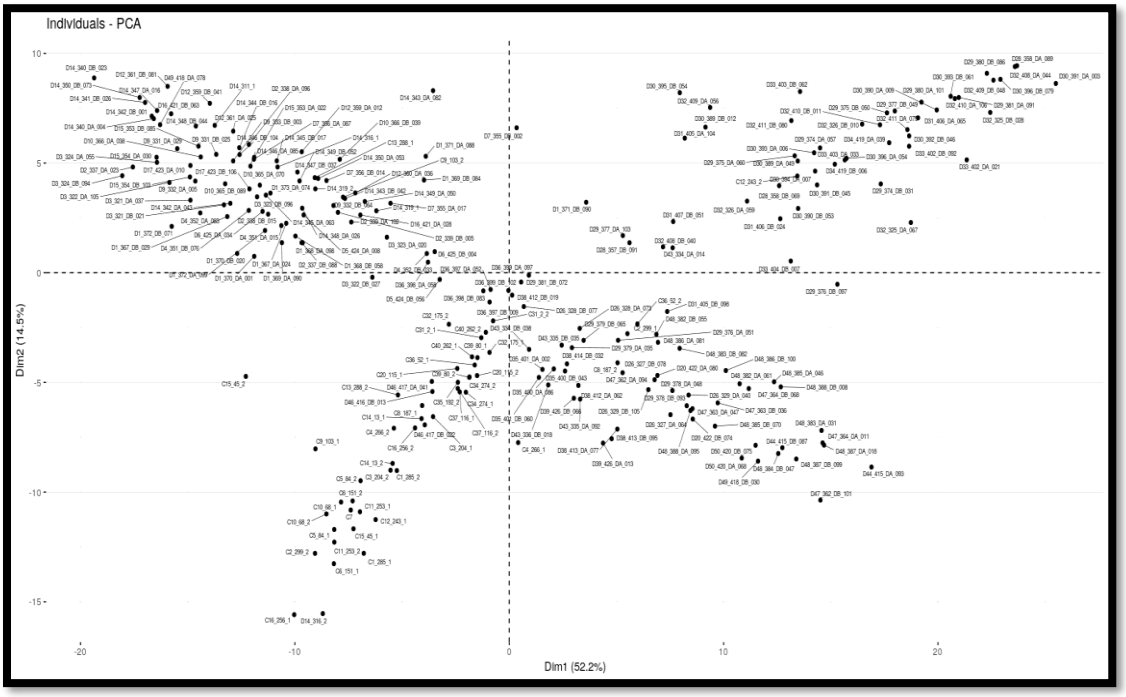

Figure 4-5. Principal component score plot obtained from Principal component analysis (PCA) for 208 lines and data of 29072 SNPs, PC1: 52.2 % and PC2: 14.5%

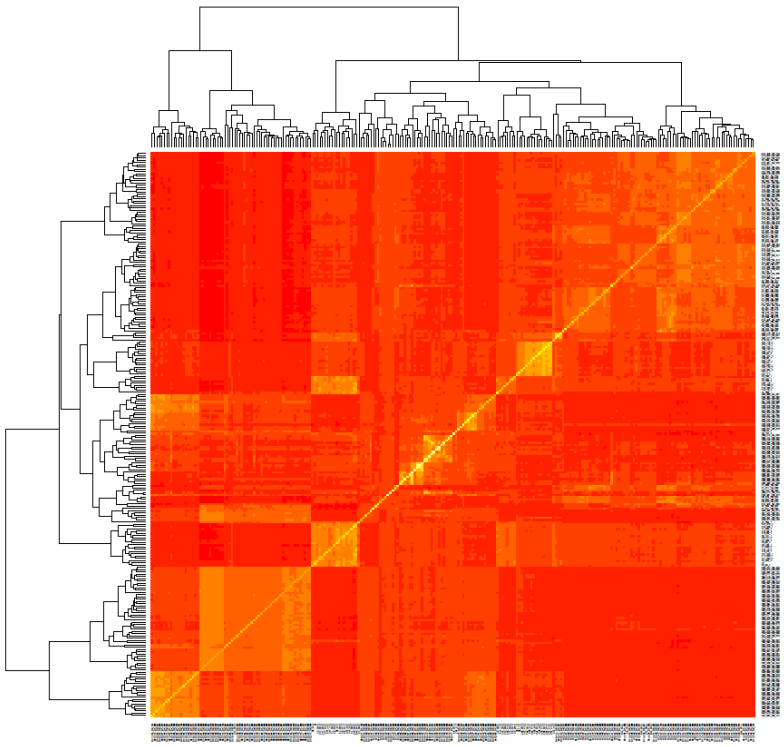

Figure 4-6. Heatmap of the genomic relationship matrix from CHIBA's sweet sorghum breeding lines showed the subdivision of the breeding population, at the bottom and right side are the name of the genotypes

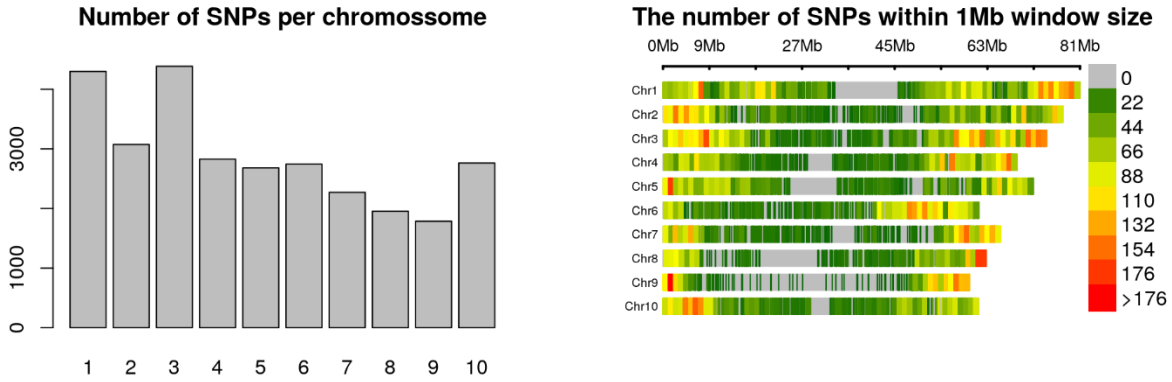

Figure 4-7. Distribution of the 29,072 along the ten sorghum chromosomes SNPs (left), the number of SNPs within 1 Mb window size (right)

Table 4-4. Average prediction accuracy by traits and models

Trait	Bayes A	Bayes B	Bayes C	BRR
Soluble solids concentration	0.739368	0.511396	0.736653	0.736499
Grain yield	0.490796	0.491373	0.482639	0.481198
Heading	0.597978	0.599583	0.595495	0.596233
Juice weight	0.632285	0.632314	0.629028	0.629876
Leaf weight	0.684383	0.684510	0.682436	0.683309
Maturity	0.541581	0.542610	0.541116	0.541886
Plant height	0.532908	0.533622	0.524404	0.523821
Stem diameter	0.639178	0.639478	0.638274	0.637813
Stem weight	0.654837	0.655519	0.651161	0.651939
Total leaf number	0.679599	0.680321	0.679481	0.679807
Total Stem number	0.594613	0.594769	0.593353	0.593671
Total green leaf		0.699868	0.698784	0.699030

Table 4-5. Average running time per trait and model in minutes for 100 replicates

Trait	Bayes A	Bayes B	Bayes C	BRR
Soluble solids concentration	18.00	15.89	19.61	11.33
Grain yield	17.63	22.62	18.34	13.23
Heading	17.89	23.44	18.76	15.15
Juice weight	20.35	17.59	19.53	11.28
Leaf weight	17.13	21.49	17.56	12.36
Maturity	17.47	23.38	18.70	16.34
Plant height	17.31	20.26	19.10	18.98
Stem diameter	16.56	21.64	19.08	18.96
Stem weight	20.36	17.44	17.42	11.04
Total leaf number	18.95	21.39	17.24	16.54
Total Stem number	20.47	23.85	16.46	11.39
Total green leaf		23.60	16.40	16.56
Average	**18.38**	**21.05**	**18.18**	**14.43**

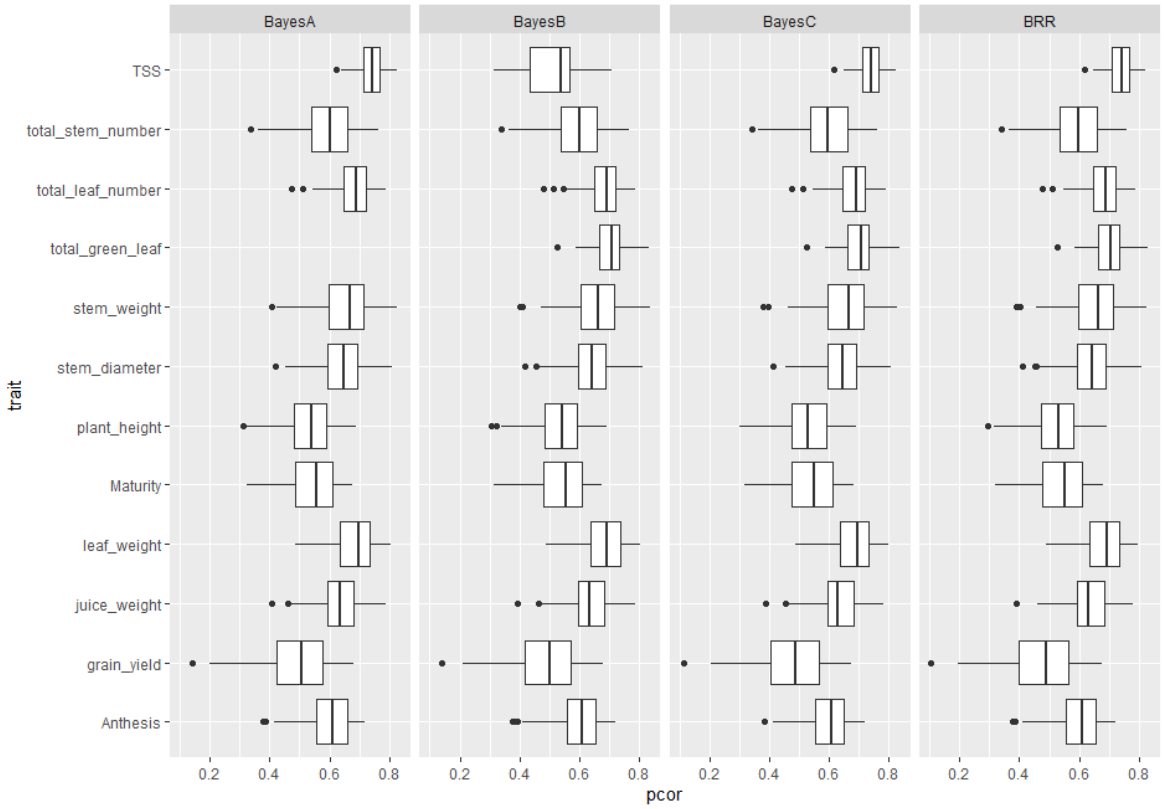

Figure 4-8. Predictive ability of the models

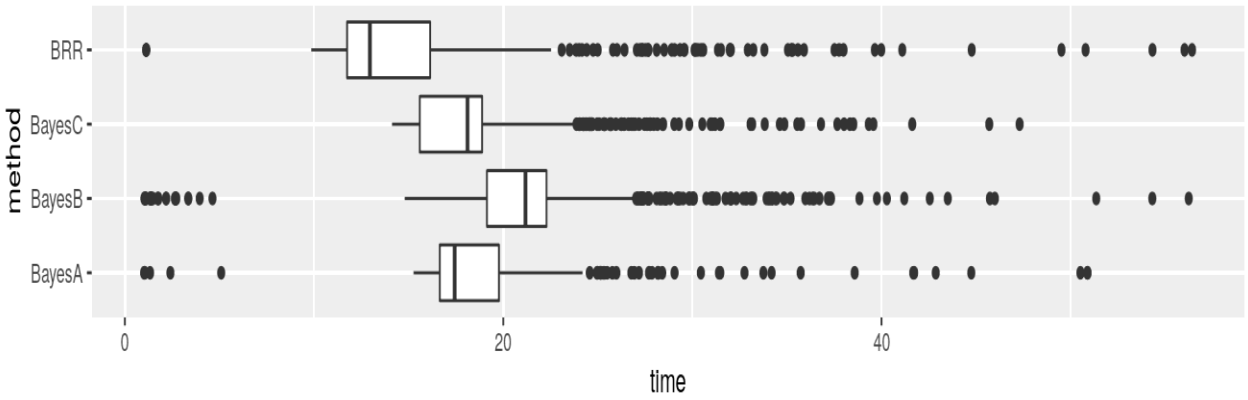

Figure 4-9. Total runtimes in minutes for fitting the four genomic prediction models in all cross-validation runs

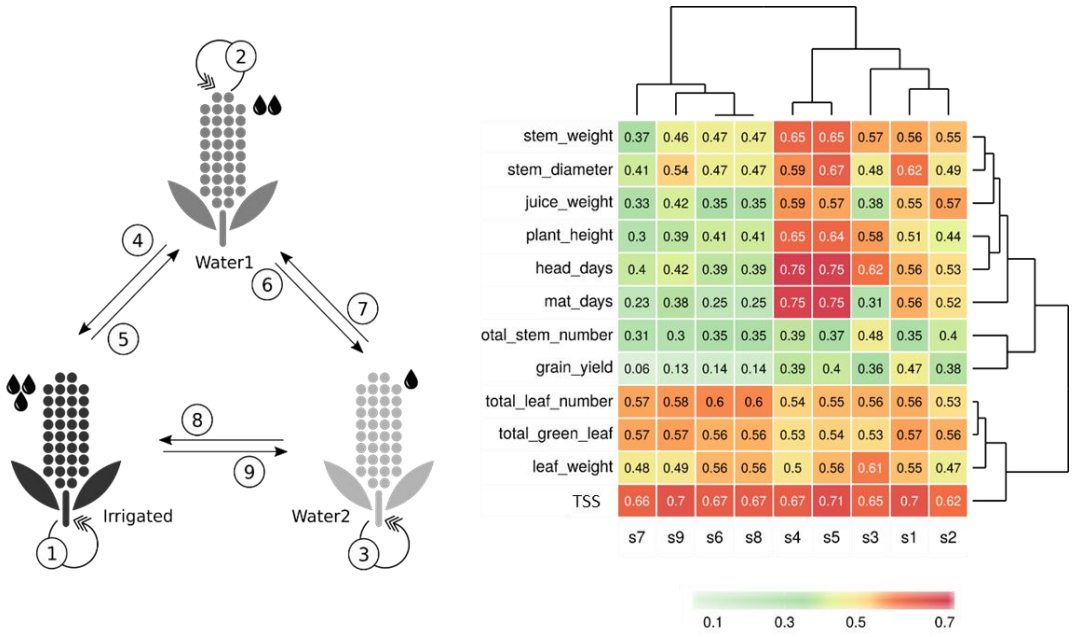

Figure 4-10. Genomic prediction of experimental scenarios

Scenarios 1 -3 involved GS performance within treatments, scenarios 8-9 compare GS performance across irrigated and water stress 2 conditions (vegetative water stress), scenarios 4-5 assess GS performance across irrigated and water stress1 (pre-flowering water stress), scenarios 6-7 compare GS performance across water stress1 and water stress2.

Table 4-6. Genetic covariance between grain yield and associated traits

	GY	LW	SW	PH	TGL	TLN
GY	1	0.61	0.70	0.41	0.43	0.51
LW	0.31	1	0.89	0.60	0.87	0.88
SW	0.44	0.74	1	0.81	0.66	0.71
PH	0.18	0.26	0.70	1	0.31	0.39
TGL	0.24	0.93	0.60	0.12	1	0.98
TLN	0.29	0.94	0.68	0.18	0.99	1

Values for vegetative water stress are below the diagonal and those above the diagonal are for pre-flowering water stress

Table 4-7. Prediction accuracy of multi-traits models

		Prediction accuracy	
Trait	**aux_traits**	**PWS**	**VWS**
Grain yield	LW	0.37	0.13
Grain yield	PH	0.36	0.13
Grain yield	TGL	0.37	0.13
Grain yield	SW	0.36	0.13
Grain yield	LW+ PH	0.36	0.13
Grain yield	LW+TGL	0.37	0.14
Grain yield	LW+ SW	0.36	0.13
Grain yield	PH+TGL	0.36	0.13
Grain yield	PH+SW	0.36	0.13
Grain yield	TGL+SW	0.36	0.13
Grain yield	LW+PH+TGL+SW	0.36	0.16

CHAPTER 5
DISCUSSION

Phenotypic analysis detected significant ($P<0.05$) differences among the sorghum genotypes across the three environments. G x E was highly significant, indicating that the impact of the water treatment varied among genotypes. According to Borrell et al (2006), the genetic variation between the genotypes can be due to morphological and physiological modifications. Low heritability values were observed for grain yield across the three environments. The low heritability estimates (broad-sense or genomic) for grain yield is due to the direct and indirect accumulative effects of yield components on grain yield (Bello et al. 2007). The low heritability of grain yield across environments was consistent with findings from previous studies conducted on sorghum biomass (Velazco et al. 2019; Kenga et al. 2006). However, for some other studies, high heritability estimates were observed for grain yield (Al-Naggar et al. 2018; Hamidou et al. 2018), giving the possibility to successfully perform direct selection for this trait. The difference between the two heritability estimated values may be explained by the factors that affect the phenotypic variances of the genotypes across the treatments. For example, the grain yield was slightly affected by midge and the plants were completely stunted in the vegetative water stress treatment. Leaf weight (stay-green trait) was positively correlated with grain yield under fully irrigated and water stress conditions, supporting its potential utility for indirect selection for grain yield improvement under drought stress. The results of the present study are in agreement with previous experiments in regard to positive correlation between grain yield and most of the traits (Kenga et al. 2006). The positive significant correlation observed between grain yield and soluble solid

concentration suggested that the trade-off between those traits may be genotype dependent (Gutjahr et al. 2013a) and that tandem improvement of grain yield and stem sugar content is possible in sorghum (Murray et al. (2008). These results also suggested that the sorghum breeding program at CHIBAS is on the right track to develop multi-purpose sorghum varieties (food, feed, and fuel).

Discovery and application of molecular markers such as SNPs and simple sequence repeats has revolutionized modern plant breeding for many species. Although, GBS can be disadvantageous due to the amount of missing data generated (Beissinger et al. 2013), it remains one of the quickest and most cost-effective genotyping tools (Poland and Rife, 2012). The number of SNPs discovered through GBS in the current study provided adequate genome coverage, allowing model training and testing for genomic selection. Similar studies have successfully employed GBS derived SNPs in sorghum for genomic selection studies (Habier et al. 2009; Mulder et al. 2012; Oliveira et al. 2018).

The accuracy of prediction models in genomic selection is important for the success of the breeding program (Gianola 2013; Crossa et al. 2017). In the current study, there were minimal differences in prediction accuracies across the models tested, perhaps due to similarities in biological and statistical assumptions across the models (Gianola et al. 2013; Hayes et al. 2009; De los Campos et al. 2013; Ferrão et al. 2018). Bayes B had the highest prediction accuracy compared to other models. According to Meuwissen et al. (2001), the modest improvement of Bayes B over the other methods is due to its great flexibility to work with contrasting genetic architectures. In addition, some authors have reported that methods where the effect of SNPs are assumed to have a

small variance with a normal distribution (e.g. BRR) are less efficient than methods that allow mutations with moderate to large effects (e.g. Bayes A and Bayes B) (Meuwissen et al. 2001; Mota et al. 2018; Gapare et al. 2018). Different studies on comparison and adjustment of statistical models in genomic selection for different species have reported similar results across the models (Oliveira et al. 2018; Haile et al. 2019; Ferrão et al. 2018; Wang et al. 2015). De Los Campos et al. (2013) reported that the large gap between the number of observations and parameters can constrain the statistical learning of the models, which may lead to similar predictive performances of the models.

Computation time across models is an important consideration in cases where resources are limited (Ferrão et al. 2018; Mota et al. 2018). In such cases, the least computationally intensive model should be considered. In the current study, the mean elapsed time required by each tested model substantially differed across the models. In agreement with previous studies, Bayes B was the most resource intensive model, while BRR required the least resources (Mota et al. 2018; Ferrão et al. 2018). Since BRR yielded similar prediction accuracies as the other models, yet with less computational resources, it may be an efficient model for implementation of genomic selection in small breeding programs, especially in developing countries with limited resources (Guo et al. 2012; Ferrão et al. 2018; Mota et al. 2018).

Genotype-by-environment interaction is an important factor affecting both phenotypic and genomic selection accuracy (Haile et al. 2019). Studies on the impact of G x E in genomic selection for sorghum is limited. As expected, within-environment prediction (scenarios 1-3) provided higher predictive accuracy

values than across-environment predictions when the vegetative water stress environment is involved (scenarios 6-9), confirming the fact that expression of genotypes is generally dependent on the environmental conditions. Similar results were observed in previous study in coffee and lentil (Ferrão et al. 2018; Haile et al. 2019). Theoretically, the interaction between genotype and the environment occurs, because the final state of a trait is determined by its genetic make-up, which is expressed according to the conditions in which the genotype is developing (Malosetti et al. 2013). The genotypic variation is captured by the estimation of marker effects, which eventually influence the predictions (Ferrão et al. 2018). Considering the complex mechanisms and processes associated with phenotypic response across diverse and contrasting environments, there is need to develop rapid and resource-efficient analytical tools to help breeders perform accurate prediction (Malosetti et al. 2013). In the current study, modelling of the interaction terms (G x E) helped improve the prediction accuracy for stem weight, stem diameter, plant height, heading, maturity, leaf weight and soluble solids concentration across fully irrigated and pre-flowering water stress environments (scenarios 5 vs 2). Previous studies have demonstrated that modeling G x E can improve the accuracy of genomic predictions, thus suggesting that G x E interactions are important in genomic selection (Crossa et al. 2015; Jarquín et al. 2014; Haile et al. 2019). Generally, all scenarios that involved prediction of performance in the vegetative water stress environment revealed low prediction accuracies (scenarios 6, 7, 8, 9). These results suggested that the performance of the sorghum genotypes was more impacted by vegetative water stress. Consequently, further re-training of the models may be necessary to increase the

prediction accuracy. Such efforts would potentially help identify a set of markers with stable effects across environments, thus reducing G × E interactions (López-Cruz et al. 2015).

Grain yield and concentration of soluble solids are key target traits for the sorghum breeding program at CHIBAS. While all the models tested showed high predictive ability for soluble solids within and across environments, low predictive accuracies were observed for grain yield, especially in the water stress environments. This was expected because grain yield is a complex quantitative trait controlled by many genes of small effects, and heavily influenced by G x E (Schulthess et al. 2015; Velazco et al. 2019; Fernandes et al. 2017). However, previous studies have shown that prediction accuracy for grain yield across environments can be improved by including correlated traits in the models in a process called multi-trait genomic selection or trait-assisted genomic selection (Fernandes et al. 2017; Haile et al. 2019; Velazco et al. 2019). A phenotypic correlation assumes that phenotypic values for two traits are associated due to genetic and non-genetic causes (Bernardo 2010). The ideal scenario to exploit multi-trait genomic selection would be when a relatively highly heritable indicator trait is associated with the prediction of a relatively low heritable trait with scarce phenotypic records (Schulthess et al. 2015). Given the low prediction accuracy for grain yield observed in the current study, multi-trait genomic selection was attempted. Our results suggested no improvement in the prediction accuracy of grain yield based on multi-trait methods, contrary to previous studies that have shown superiority of this strategy over single-trait genomic selection (Schulthess et al. 2015; Guo et al. 2014; Jia and Jannink 2012). This result was not expected

as grain yield has low heritability and is moderately correlated with other traits. However, our findings are similar to those reported in sorghum and maize (Fernandes et al. 2017; Dos Santos et al. 2016). The lack of improvement of the multivariate models might be due to low- to- moderate magnitude correlation and moderate heritabilities between the traits studied (Dos Santos et al. 2016). In addition, a study conducted by Guo et al. (2014) suggested that multivariate genomic models can be superior to univariate models when unbalanced data scenarios are considered, which was not the case for this study. These authors recommended the application of multivariate model only in specific cases (Guo et al. 2014).

CHAPTER 6
CONCLUSIONS

Wide phenotypic variation was observed for all traits across the three environments in response to water availability. Among all the traits, grain yield was the least stable across the environments. Prediction accuracy across the models was very similar for all traits. However, Bayes B performed slightly better than the other models. The BRR model required the least computation resources and may be adopted in small breeding programs with limited resources, especially in developing countries like Haiti. It could also be concluded that single-trait models yielded similar prediction accuracy as multi-trait models for grain yield. Prediction accuracy for grain yield under water stress conditions was lower than that under fully irrigated conditions. The prediction accuracy for the secondary traits (leaf weight, maturity, heading, and plant height) were generally higher than the prediction accuracy of grain yield under almost all the scenarios. In a modern breeding scheme, multi-environment field trials might be considered in advanced generations in order to re-estimate the markers effects. The challenge is that multi-location field trials are laborious and costly. Therefore, using data from one site to predict the performance in independent sites is necessary for plant breeding. Although, within-environment prediction accuracy was higher than that across irrigated and vegetative water stress environments, the values were moderate to high when irrigated environment was used to predict trait performance in pre-flowering water stress environment. Considering the changes in environment (new diseases, new pests, soil degradation, and changes in water availability), these results suggested that across environment GS could be a potential strategy for rapid, accurate and

resource efficient prediction. Overall, accuracies of genomic prediction obtained in this study are encouraging for implementation of GS in small breeding programs. Nevertheless, further studies involving multi-trait modelling are required to assess usefulness of highly correlated traits (highly heritable, cheap and easy to sample) in improving prediction accuracies for grain yield in sorghum.

LIST OF REFERENCES

Adams, C. B., Erickson, J. E., & Singh, M. P. (2015). Investigation and synthesis of sweet sorghum crop responses to nitrogen and potassium fertilization. Field Crops Research 178:1-7.

Adebooye, O. C., Ajadi, S. O., & Fagbohun, A. B. (2006). An Accurate Mathematical Formula for Estimating Plant Population in a Four-Dimensional Field of Sole Crop. Journal of Agronomy, volume 5, issue 2, page 289-292.

Almodares, A., & Mostafafi Darany, S. M. (2006). Effects of planting date and time of nitrogen application on yield and sugar content of sweet sorghum, Journal of Environmental Biology 27(3) 601-605.

Al-Naggar, A. M. M., El-Salam, R. M. A., Hovny, M. R. A., & Yaseen, W. Y. S. (2018). Variability, Heritability, Genetic Advance and Interrelationships for Agronomic and Yield Traits of Sorghum B-Lines under Different Environments. Asian Journal of Biochemistry, Genetics and Molecular Biology 1(1): 1-13.

Assefa, Y., Staggenborg, S. A., & Prasad, V. P. V. (2010). Grain sorghum water requirement and responses to drought stress: A review. Online. Crop Management.

Azhar, F. M., & McNeilly, T. (1987). Variability for Salt Tolerance in Sorghum bicolor (L.) Moench. under Hydroponic Conditions. Journal of agronomy and crop science Volume159, Issue4 Pages 269-277.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.) Plant Science volume 242, Pages 23-36.

Bates, D., Maecechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2016). lme4: Linear Mixed-Effects Models using Eigen and S4, 1.17.

Baye, T. M., Abebe, T., & Wilke, R. A. (2011). Genotype–environment interactions and their translational implications. Personalized Medicine, 8(1), 59–70.

Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., et al. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193:1073–1081.

Bekele, W. A., Wieckhorst, S., Friedt, W., & Snowdon, R. J. (2013). High-throughput genomics in sorghum: From whole-genome resequencing to a SNP screening array. Plant Biotechnology Journal, 11(9), 1112-1125.

Bello, D., Kadams, A. M., & Simon, S. Y. (2007). Correlation and path coefficient analysis of grain yield and its components in sorghum. Nig J.Trop. Agric., A Publication of SAAT, FUT, Yola, Nigeria 3: 4-9.

Bernal, A. H., Ligarreto, M, G. A., & Hernández, R. S. (2014). Effects of the genotype and environment interaction on sugar accumulation in sweet sorghum varieties (Sorghum bicolor [L.] Moench) grown in the lowland tropics of Colombia. Agron. colomb., Volumen 32, Número 3, p. 307-314.

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

Boichard, D., Ducrocq, V., Croiseau, P., & Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. Comptes Rendus Biologies Volume 339, Pages 274-277.

Bonnet, J. J., Azar, M. S., & Lamy, J. D. (2012). Plan de gestion des pestes et des pesticides (PGPP).

Borrell, A. K., Hammer, G. L., & Henzell, R. G. (2000). Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. Crop Sci. 40:1037–1048.

Borrell, A. K., & Hammer, G. L. (2000). Nitrogen dynamics and the physiological basis of stay-green in sorghum. Crop Sci. 40:1295–1307.

Borrell, A., Jordan, D., Mullet, J., Henzell, B., & Hammer, G. (2006). Drought adaptation in sorghum. In: Drought adaptation in cereals. The Haworth Press, Inc., Binghamton, NY, USA, pp: 335-400.

Borrell, A. K., Mullet, J. E., George-Jaeggli, B., Oosterom, E. J. V., Hammer, G. L., Klein, P. E., & Jordan, D. R. (2014). Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. Journal of Experimental Botany, Volume 65, Issue 21, 1 Pages 6251–6263.

Boyer, J. S. (1982). Plant productivity and environment. Science 218, 4571, 443–448.

Brown, J., Caligari, P., & Campos, H. (2014). Introduction to plant breeding-revised and updated, 2nd edition, John Wiley and Sons.

Burgueno, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. Crop Science, 52(2), 707.

Burks, P. S., Felderhoff, T.J., Viator, H. P., & Rooney, W. L. (2013). The influence of hybrid maturity and planting date on sweet sorghum productivity during a harvest season. Agron. J. 105(1):263.

Butler, D. (2018). Asreml: Fits the Linear Mixed Model. R package version 4.1.0.98.

Casady, A. J., Heyne, E. G., & Hansing, E. D. (1962). Breeding Sorghum for Smut Resistance.

Chaurasla, N. K. (2015). Breeding sorghum. Department of plant breeding and genetics Assam Agricultural University Jorhat.

Charles, J. R. (2017). Caractérisation et évaluation des prédictions génomiques des lignées de sorgho développées par le Chibas en Haïti. Mémoire de fin d'études.

Che, P., Anand, A., Wu, E., Sander, J. D., Simon, M. K., Zhu, W., et al. (2018). Developing a flexible, high-efficiency Agrobacterium-mediated sorghum transformation system with broad application. Plant Biotechnology Journal, pp. 1–8.

Chung, Y. S., Choi, S. C., Jun, T., & Kim, C. (2017). Genotyping-by-Sequencing: A Promising Tool for Plant Genetics Research and Breeding. Hortic. Environ. Biotechnol. 58(5) : 425-443.

CNSA. (2012). Rapport final campagne de printemps sept.

Costa, P. M. D. A. (2015). Prediction of breeding values in sugarcane using pedigree and genomic information. Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doctor Scientiae.

Cox, T. S., Van Tassel, D. L., Cox, C. M., & DeHaan, L. R. (2010). Progress in Breeding Perennial Grains. Crop & Pasture Science 61 :513-521.

Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., & Pérez-Rodriguez, P. (2015). Extending the marker x environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. Crop Science 56 :1 :17.

Crossa, J., Pérez-Rodriguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De los Campos, G. et al. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science, Vol. 22, No.11.

Dahir, M., Zhu, K., Guo, X., Aboshora, W., & Peng, W. (2015). Possibility to Utilize Sorghum Flour in a Modern Bread Making Industry. Journal of Academia and Industrial Research (JAIR) Volume 4, Issue 4, ISSN: 2278-5213.

de Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193(2), 327-345.

De Morais Cardoso, L., Pinheiro, S. S., Martino, H. S., & Pinheirosant'Ana, H. M. (2015). Sorghum (*Sorghum bicolor* L.): Nutrients, bioactive compounds, and potential impact on human health. Crit. Rev. Food Technol, 57:372–390.

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci.19(9) :592-601.

Dos Santos, J. P. R., Vasconcellos, R. C. C., Pires, L. P. M., Balestre, M., & Von Pinho, R. G. (2016). Inclusion of dominance effects in the multivariate gblup model. PLoS One 11(4):1–21.

Duncan, R. R., Bockholt, A. J., & Miller, F. R. (1981). Descriptive comparison of senescent and non-senescent sorghum genotypes. Agronomy Journal 73, 849-853.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:1 10.

Erickson, J. E., Helsel, Z. R., Woodard, K. R., Vendramini, J. M. B., Wang, Y., Sollenberger, L. E., & Gilbert, R. A. (2011). Planting date affects biomass and brix of sweet sorghum grown for biofuel across Florida. Agronomy Journal 103-1827-1833.

Erickson, J. E., Woodard, K. R., & Sollenberger, L. E. (2012). Optimizing sweet sorghum production for biofuel in the southeastern US through nitrogen fertilization and top removal. Bioenergy Research 5:86-94.

Erpelding, J. E. (2010). Anthracnose resistance in sorghum breeding lines developed from Ethiopian germplasm. Online. Plant Health Progress doi :10.1094/PHP-2010-1123-02-RS.

FAO/PAM. (2017). Haïti : Rapport de la mission FAO/PAM d'évaluation des récoltes et de la sécurité alimentaire.

Ferrão, L. F. V., Ortiz, R., & Garcia, A. A. F. (2017). Genomic Selection: State of the Art. Genetic Improvement of Tropical Crops pp 19-54.

Ferrão, L. F. V., Ferrão, R. G., Ferrão, M. A. G., Fonseca, A., Carbonetto, P., Stephens, M., & Garcia, A. A. F. (2018). Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. Heredity 122, 261–275.

Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., & Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131 : 747.

Flecher, J. (2016). En route vers la production de nouvelles variétés de sorgho en Haïti. https://lenouvelliste.com/lenouvelliste/article/162184/En-route-vers-la-production-de-nouvelles-varietes-de-sorgho-en-Haiti.

Gabriel, S. (2016). Perte de 60% de la production de petit mil : une catastrophe annoncée pour Haïti ! http://ayibopost.com/perte-de-60-de-la-production-de-petit-mil-une-catastrophe-annoncee-pour-haiti/.

Gapare, W., Liu, S., Conaty, W., Zhu, Q-H., Gillespie, V., Llewellyn, D., Stiller, W., & Wilson, I. (2018). Historical Datasets Support Genomic Selection Models for the Prediction of Cotton Fiber Quality Phenotypes Across Multiple Environments. G3 (Bethesda) 8(5): 1721–1732.

Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. Genetics 194, 573–596.

Gilleard, J. S., & Redman, E. (2016). Genetic Diversity and Population Structure of *Haemonchus contortus*. Advances in Parasitology 93.

Goddard, M. E., & Hayes B. J. (2007). Genomic selection. J Anim Breed Genet.124(6) :323-30.

Grenier, C., Cao, T. V., Ospina, Y., Quintero, C., Châtel, M. H., Tohme, J., et al. (2016). Correction: Accuracy of Genomic Selection in a Rice Synthetic Population Developed for Recurrent Selection Breeding. PLOS ONE 11(5): e0154976.

Grundon, N. J., Edwards, D. B., Takkar, P. N., Asher, C. J., & Clark, R. B. (1987). Nutritional Disorders of Grain Sorghum. Australian Centre for International Agricultural Research. 1987 G.P.O. Box 1571.

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., & Su, G. (2014). Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:30.

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

Gutjahr, S., Vaksmann, M., Dingkuhn, M., Thera, K., Trouche, G., Braconnier S., & Luquet, D. (2013a). Grain, sugar and biomass accumulation in tropical sorghums. I. Trade-offs and effects of phenological plasticity Functional Plant Biology, 40: 342-354.

Gutjahr, S., Clément-Vidal, A., Soutiras, A., Sonderegger, N., Braconnier, S., Dingkuhn, M., & Luquet, D. (2013b). Grain, sugar and biomass accumulation in photoperiod sensitive sorghums. II. Biochemical processes at internodes level and interaction with phenology. Functional Plant Biology, 40, 355–368.

Guzman, C., Peña, R. J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., Rutkoski, J., Poland, J., & Battenfield, S. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. Applied & Translational Genomics volume 11, December 2016, Pages 3-8.

Habier, D., Tetens, J., Seefried, F-R., Lichtner, P., & Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution 42:5.

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:186.

Haile, T. A., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., & Bett, K. E. (2019). Genomic selection for lentil breeding: empirical evidence. bioRxiv 608406.

Hamidou, M., Souley, A. M., Kapran, I., Souleymane, O., Danquah, E. Y., Ofori, K., Gracen, V., & Ba, M. N. (2018). Genetic Variability and Its Implications on Early Generation Sorghum Lines Selection for Yield, Yield Contributing Traits, and Resistance to Sorghum Midge. Hindawi International Journal of Agronomy Volume 2018, Article ID 1864797.

Harlan, J. R. & de Wet, J. M. J. (1972). A simplified classification of cultivated sorghum. Crop Science 12:172-176.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443.

He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plants Science| Volume5 | Article 484.

Heffner, E. L., Sorrels, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. Crop Science - dl.sciencesocieties.org.

Heslot, N., Yang, H., Sorrells, M. E., & Jannink, J. (2012). Genomic Selection in Plant Breeding: A Comparison of Models Crop Sci. 52 :146–160.

House L. R. (1987). Manuel pour la sélection du sorgho 2e éd. Patancheru, Inde, ICRISAT.

Hunt, C. H., Eeuwijk, F. A. V., Mace, E. S., Hayes, B. J., & Jordan, D. R. (2018). Development of Genomic Prediction in Sorghum Crop Sci. 58 :690–700.

Iheshiulor, O. O. M., Woolliams, J. A., Svendsen, M., Solberg, T., & Meuwissen, T. H. E. (2017). Simultaneous fitting of genomic-BLUP and Bayes-C components in a genomic prediction model. Genetics Selection Evolution 49:63.

Jia, Y., & Jannink, J. L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. & Weeks, D.P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res. 41, e188.

Kenga, R., Tenkouano, A., Gupta, S. C., & Alabi, S. O. (2006). Genetic and phenotypic association between yield components in hybrid sorghum (*Sorghum bicolor* (L.) Moench) populations. Euphytica (2006) 150: 319–326.

Kidanemaryam, W. (2019). Review on Mechanisms of Drought Tolerance in Sorghum (Sorghum bicolor (L.) Moench) Basis and Breeding Methods. Acad. Res. J. Agri. Sci. Res. 7(2): 87-99.

Kim, J. S., Klein, P. E., Klein, R. R., Price, H. J., Mullet, J. E., & Stelly, D. M. (2005). Chromosome identification and nomenclature of Sorghum bicolor. Genetics 169, 1169–1173.

Kulwal, P. L. (2016). Association Mapping and Genomic Selection—Where Does Sorghum Stand?. In: Rakshit S., Wang YH. (eds) The Sorghum Genome. Compendium of Plant Genomes. Springer.

Leclerc, E., Pressoir, G., & Braconnier, S. (2014). « L'avenir prometteur du sorgho sucré en Haïti ». Field Actions Science Reports, Special Issue 9.

Léder, I. (2004). Sorghum and Millets, in Cultivated Plants, Primarily as Food Sources in György, F. (Ed.) Encyclopedia of Life Support Systems (EOLSS) Oxford: Eolss Publishers.

Lian, L., Jacobson, A., Zhong, S., & Bernardo, R. (2014). Genome wide Prediction Accuracy within 969 Maize Biparental Populations. Crop Sci. 54 :1514–1522.

Liu, H., Zhou, H., Wu, Y., Li, X., Zhao, J., Zuo, T., et al. (2015). The Impact of Genetic Relationship and Linkage Disequilibrium on Genomic Selection.

López-Cruz, M. A., Crossa, J., Bonnet, D., Dreisigacker, S., Poland J., Jannink, J. L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a markers x environment interaction genomic selection model. G3 (Bethesda) 5(4): 569–582.

Magalhaes, J. V., Liu, J., Guimarães, C. T., Lana, U. G. P., Alves, V. M. C., Wang, Y., et al. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nature Genetics volume 39, pages 1156–1161.

Makanda, I., Derera, J., Tongoona, P., & Sibiya, J. (2011). Development of sorghum for bioenergy: A view from the stakeholders and priorities for breeding dual purpose varieties African Journal of Agricultural Research Vol. 6(19), pp. 4477-4486.

Malosetti, M., Ribaut, J-M., & Eeuwijk, F. A. V. (2013). The statistical analysis of multi-environment data: modeling genotype-by environment interaction and its genetic basis. Front Physiol 4:44.

Mathur, S., Umakanth, A. V., Tonapi, V. A., Sharma, R., & Sharma, M. K. (2017). Sweet sorghum as biofuel feedstock: Recent advances and available resources.

Melo, J., & Cheschini, J. (2012). Damage caused by birds in sorghum (*Sorghum bicolor*) crops in Central Brazil. Bioagro 24(1) :33-38.

Menezes, C. B., Ticona-Benavente, C. A., Tardin, F. D., Cardoso, M. J., Bastos, E. A., Nogueira, D. W., et al. (2014). Selection indices to identify drought-tolerant grain sorghum cultivars. Genet. Mol. Res. 13 (4): 9817-9827.

Meru, G. M. (2010). Genotyping BC3F2 populations of four Ethiopian sorghum varieties for Stay Green QTLs introgression through Marker assisted selection with SSRs. thesis submitted in partial fulfillment of the requirements for the award of the degree of Master of Science (Biotechnology) of Kenyatta University.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4), 1819-1829. Retrieved from http://www.genetics.org/cgi/content/abstract/157/4/1819.

Miller, A. N., & Ottman, M. J. (2010). Irrigation Frequency Effects on Growth and Ethanol Yield in Sweet Sorghum. Agron. J. 102 :60-70.

Mofokeng, M. A., Shimelis, H., Laing, M., & Shargie, N. (2017). Sorghum [*Sorghum bicolor* (L.) Moench] breeding for resistance to leaf and stalk anthracnose, Colletotrichum sublineolum, and improved yield: Progress and prospects. AJCS 11(09) :1078-1085, ISSN:1835-2707.

Mota, R. R., Silva, F. F. E., Guimarães, S. E. F. F., Hayes, B., Fortes, M. R. S., Kelly, M. J. et al. (2018). Benchmarking Bayesian genome enabled-prediction models for age at first calving in Nellore cows. Livestock Science volume 211, pages 75-79.

Mulder, H. A., Calus, M. P.L., Druet, T., & Schrooten, C. (2012). Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. J Dairy Sci 95:876–889.

Murray, S. C., Sharma, A., Rooney, W. L., Klein, P. E., Mullet, J. E., Mitchell, S. E., & Kresovich, S. (2008). Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Science, 48(6), 2165.

Murray, S. C. (2008). Genetic and phenotypic diversity in sorghum for improvement as a biofuel feedstock. https://ecommons.cornell.edu/handle/1813/11052.

Murray, S. C., Rooney, W. L., Hamblin, M. T., Mitchell, S. E., & Kresovich, S. (2009). Sweet Sorghum Genetic Diversity and Association Mapping for Brix and Height. The Plant Genome 2 :48–62.

Nasidi, M., Akunna, J., Deeni, Y., Blackwood, D., & Walker, G. (2010). Bioethanol in Nigeria: comparative analysis of sugarcane and sweet sorghum as feedstock sources. Energy & Environmental Science 3, 1447–1457.

Nebie, B., Nanema, R. K., Kando, P. B., Traore, E. R., Labeyrie, V., Sawadogo, N., et al. (2013). Variation de caractères agromorphologiques et du Brix d'une collection de sorghos à tige sucrée du Burkina Faso. International Journal of Biological and Chemical Science 7(5) : 1919-1928.

Oliveira, A. A., Pastina, M. M., Souza, V. F., Parrella, R. A. C., Noda, R. W., Simeone, M. L. F., et al. (2018). Genomic prediction applied to high-biomass sorghum for bioenergy production. Mol Breeding 38:49.

Olweny, C., Okori, P., Abayo, G., & Dida, M. (2012). Unravelling the potential of sweet sorghum for sugar production in Kenya. Third RUFORUM Biennial Meeting 24 - 28 September 2012, Entebbe, Uganda.

Ornella, L., González-Camacho, J. M., Dreisigacker, S., & Crossa, J. (2017) Applications of Genomic Selection in Breeding Wheat for Rust Resistance. In: Periyannan S. (eds) Wheat Rust Diseases. Methods in Molecular Biology, vol 1659.

Panasiuk, O., & Bills, D. D. (1984). Cyanide Content of Sorghum Sprouts. Journal of Food Science Volume 49, Issue 3, Pages 791-793.

Pérez, P., & de los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics, 198(2), 483–495.

Petropoulos, G. P., Griffiths, H. M., Dorigo, W., Xaver, A., & Gruber, A. (2013). Surface Soil Moisture Estimation: Significance, Controls, and Conventional Measurement Techniques.

Picard, C. (2015). Evaluation d'une stratégie de sélection génomique dans 3 dispositifs expérimentaux chez la tomate.

Pineli, L., Zandonadi, R. P., Bothelho, R. B. A., Oliveira, V. R., & Figueiredo, L. F. (2015). The use of sorghum to produce gluten-free breads: A systematic review. Journal of Advanced Nutrition and Human Metabolism; 2: e944.

Piper, J. K., & Kulakow P. A. (1994). Seed yield and biomass allocation in Sorghum and F1 and backcross generations of S. bicolor X S. halepense hybrids. Can. J. Bot. 72: 468-474.

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

Priyavratha, R. S. B., & Nakasimha, D. V. R. (1953). Bird damage in jowar [sorghum]. Madras Agri. J., 40(10):466-467.

Rao, S. S. (2004). Abiotic stress factors affecting sorghum production and their Management. Physiological basis of genetic gain in tropical dryland sorghum.

Ratnavathi, C. V., Suresh, K., Kumar, B. S. V., Pallavi, M., Komala, V. V., & Seetharama, N. (2011) Study on genotypic variation for ethanol production from sweet sorghum juice. Biomass Bioenergy 34 :947-52.

Raza, S., Shoaib, M. W., & Mubeen, H. (2016). Genetic Markers: Importance, uses and applications. International Journal of Scientific and Research Publications, Volume 6, Issue 3, pages 221-223.

Reddy, B. V., Ramesh, P., Reddy, S., Ramaiah, B., Salimath, M., & Kachapur, R. (2005). Sweet sorghum—A potential alternate raw material for bioethanol and bioenergy. Int. Sorghum Millets Newsl.46 :79–86.

Resende, Jr. M. F. R., Muñoz, P., Resende, M. D. V., Garrick, D. J., Fernando, R. L., Davis, J. M., & Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (*pinus taeda* L.). Genetics, 190(4), 1503-1510.

Roberson, R. (2013). Zinc deficiency can reduce sorghum yields, lower test weights, Southeast Farm Press Daily

Roorkiwal, M., Jarquín, D., Singh, M. K., Bharadwaj, C., Rathore, A., Howard, R., et al. (2018). Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. Scientific Reports volume 8, Article number : 11701.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., & Crossa, J. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3: Genes, Genomes, Genet. 6:2799–2808.

Schulthess, A. W., Wang, Y., Miedaner, T., Wilde, P., Reif, J. C., & Zhao, Y. (2016). Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet 129:273-287.

Sharma, H. C., Agrawal, B. L., Vidyasagar, P., Abraham C. V., & Nwanze, K. F. (1993). Identification and utilization of resistance to sorghum midge, Contarinia sorghicola (Coquillet), in India. Crop protection Volume 12, Issue 5, Pages 343-350.

Shen, S., Huang, R., Li, C., Wu, W., Chen, H., Shi, J., et al. (2018). Phenolic Compositions and Antioxidant Activities Differ Significantly among Sorghum Grains with Different Applications. Molecules 23(5): 1203.

Singh, M. P., Erickson, J. E., Sollenberger, L. E., Woodard, K. R., Vendramini, J. M. B., & Fedenko, J. R. (2012). Mineral composition and biomass partitioning of sweet sorghum grown for biofuel in the southeastern U.S.A. Biomass and Bioenergy 47:1-8.

Song, J., Carver, B. F., Powers, C., Yan, L., Klápště, J., El-Kassaby, Y. A., et al. (2017). Practical Application of Genomic Selection in a Doubled-Haploid Winter Wheat Breeding Program, Mol Breeding 37: 117.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. PLoS Genet 11(2): e1004982.

Spindel J., & Iwata H. (2018). Genomic Selection in Rice Breeding. Rice Genomics, Genetics and Breeding pp 473-496.

Stich, B., & Inghelandt, D. V. (2018). Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. Front Plant Sci. 2018; 9: 159.

Tari, I., Laskay, G., Takács, Z., & Poór, P. (2012). Response of Sorghum to Abiotic Stresses: A Review. Journal of Agronomy Crop Science ISSN 0931-2250, pages 264-274.

Thomas, H., & Howarth, C. J. (2000). Five ways to stay green. Journal of Experimental Botany 51, 329–337.

Thomas, H., & Ougham, H. (2014). The stay-green trait. Journal of Experimental Botany, Volume 65, Issue 14, Pages 3889–3900.

Tuberosa, R., Salvi, S., Sanguineti, M. C., Maccaferri, M., Giuliani, S., & Landi, P. (2003). Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. Plant Soil 255: 35-54.

Vaksmann, M., Kouressy, M., Chantereau, J., Bazile, D., Sagnard, F., Touré, A., et al. (2008). Utilisation de la diversité génétique des sorghos locaux du Mali, Cahiers Agricultures vol. 17, n° 2 : 140-146.

Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., & Eeuwijk, F. A. V. (2019). Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theoretical and Applied Genetics pages 1–13.

Visscher, P. M., Hill, W. G. & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. Nature Reviews Genetics 9, 255-266.

Wang, X., Yang, Z., & Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. Science Bulletin, Volume 60, Issue 10.

Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. (2016). Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. Heredity 118:302–310.

White, J. A., Ryley, M. J., George, D. L., Kong, G. A., & White, S. C. (2011). Yield losses in grain sorghum due to rust infection. Australasian Plant Pathology Society 41(1).

William, M. (2016). La culture du sorgho en Haïti.lenouvelliste.com/lenouvelliste/.../La-culture-du-sorgho-en-Haiti. https://lenouvelliste.com/lenouvelliste/article/154393/La-culture-du-sorgho-en-Haiti.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 2, 1427–1436.

Wolabu, T. W., & Tadege, M. (2016). Photoperiod response and floral transition in sorghum. Plant Signal Behavior 11(12): e1261232.

Wortmann, C. S., Ferguson, R. B., Hergert, G. W., Shapiro, C. A., & Shaver, T. M. (2013). Nutrient Management Suggestions for grain sorghum. University of Nebraska-Lincoln Extension, Institute of Agriculture and Natural Resources.

Xu, W., Subudhi, P.K., Crasta, O. R., Rosenow, D.T., Mullet, J.E., & Nguyen, H.T. (2000). Molecular mapping of QTLs conferring stay-green in grain sorghum (Sorghum bicolor L. Moench). Genome 43, 461–469.

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017). Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. Frontiers in Plant Science, 8, 1916.

BIOGRAPHICAL SKETCH

Marie Dorval was born in Port-au-Prince, Haiti. She received her bachelor's degree in Agronomy from University Quisqueya in January 2016, Port-au-Prince. After her graduation, she was hired as a research assistant at CHIBAS (a research center working bioenergy and sustainable agriculture). One year later, she received a scholarship from USAID-AREA project to pursue a master's degree in Horticulture at the University of Florida. Marie joined the lab of Dr. Meru Geoffrey at the Tropical Research and Education Center (TREC). Her main focus of research was "Genomic prediction of sweet sorghum agronomic performance under drought and irrigated environments in Haiti". Her dream is that one day, her country Haiti becomes self-sufficient on the agricultural level for major crops production.