Genomics-enabled breeding for sorghum in West Africa

by

Fanna Maina Assane Mamadou

B.S., University of Blida, Algeria, 2009
M.S., University of Maradi, Niger, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

In semi-arid regions, staple crop productivity is affected by multiple environmental, biological and socio-economic factors threatening food security. Sorghum (*Sorghum bicolor*) is adapted to semi-arid and sub-humid zones of West Africa (WA). This crop is cultivated over large areas, corresponding to variable and diversified local contexts. The genetic basis of local adaptation and farmer preferences and their applications in breeding need further studies. Recent genotyping methods have provided access to high-density markers and their applications in breeding. In this thesis, genomic resources of WA germplasm were developed using the genotyping-by-sequencing method (GBS) to understand the genetic diversity and to identify quantitative trait loci (QTLs) associated with yield components under pre-flowering water deficit. Evidence of local adaptation in genomic regions linked to flowering time in sub-humid zones and balancing selection grain pigmentation were found. Phenotyping of the WA sorghum association panel (WASAP) was conducted under experimental water-deficit treatments. Significant variations of yield components were observed suggesting local adaptation and drought tolerance in the WASAP. Genome-wide association studies identified novel QTLs controlling vegetative biomass and grain weight under water deficit treatment. QTLs colocalizing with known genes in various traits were also identified. Furthermore, these genomic resources were used to develop diagnostic markers for resistance to *Striga* hermonthica, a weed parasite of grass crops, in which resistance is known to be associated with a deletion of a few genes. Using GBS data, single nucleotide polymorphisms in linkage disequilibrium with the deletion to generate breeder-friendly markers were selected. Analyses identified eight SNPs, converted to breeder-friendly markers and tested in biparental populations and diverse

germplasm using outsourced genotyping. The findings provide genetic resources to the scientific community and could facilitate genomics-enabled breeding of sorghum in sub-Saharan Africa.

Genomics-enabled breeding for sorghum in West Africa

by

Fanna Maina Assane Mamadou

B.S., University of Blida, Algeria, 2009
M.S., University of Maradi, Niger, 2013

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:
Major Professor
Geoffrey P. Morris

Copyright

Abstract

In semi-arid regions, staple crop productivity is affected by multiple environmental, biological and socio-economic factors threatening food security. Sorghum (*Sorghum bicolor*) is adapted to semi-arid and sub-humid zones of West Africa (WA). This crop is cultivated over large areas, corresponding to variable and diversified local contexts. The genetic basis of local adaptation and farmer preferences and their applications in breeding need further studies. Recent genotyping methods have provided access to high-density markers and their applications in breeding. In this thesis, genomic resources of WA germplasm were developed using the genotyping-by-sequencing method (GBS) to understand the genetic diversity and to identify quantitative trait loci (QTLs) associated with yield components under pre-flowering water deficit. Evidence of local adaptation in genomic regions linked to flowering time in sub-humid zones and balancing selection grain pigmentation were found. Phenotyping of the WA sorghum association panel (WASAP) was conducted under experimental water-deficit treatments. Significant variations of yield components were observed suggesting local adaptation and drought tolerance in the WASAP. Genome-wide association studies identified novel QTLs controlling vegetative biomass and grain weight under water deficit treatment. QTLs colocalizing with known genes in various traits were also identified. Furthermore, these genomic resources were used to develop diagnostic markers for resistance to *Striga* hermonthica, a weed parasite of grass crops, in which resistance is known to be associated with a deletion of a few genes. Using GBS data, single nucleotide polymorphisms in linkage disequilibrium with the deletion to generate breeder-friendly markers were selected. Analyses identified eight SNPs, converted to breeder-friendly markers and tested in biparental populations and diverse

germplasm using outsourced genotyping. The findings provide genetic resources to the scientific community and could facilitate genomics-enabled breeding of sorghum in sub-Saharan Africa.

Table of Contents

List of Figures

List of Tables

Acknowledgements

"Alone we can do so little; together we can do so much." Helen Keller

This work would not be possible without the help of many people who have surrounded accompanied, supported me.

I am deeply thankful to the Sorghum and Millet Innovation Lab for their financial support without which this journey and scientific discoveries would not be possible. To the "Institut de la Recherche Agronomique du Niger" (INRAN) for allowing me to pursue this program.

I would like to sincerely thank my Ph.D. advisor, Dr. Geoffrey Morris, for his unfailing support throughout these years. His availability, kindness, knowledge, scientific rigor, and precious advice have never ceased to accompany and guide me. Thank you for this very beautiful and rewarding academic journey. I am incredibly thankful for the passion and knowledge that you transmitted to me. This adventure will not be enjoyable and memorable without your full support.

My sincere thanks to my committee members: Dr. Abdoul-Aziz Saïdou for his guidance and excellent advice, Dr. Mary Guttieri for her availability and insightful comments and help, Dr. Krishna Jagadish for sharing with me his tremendous experience, and Dr. Ramasamy Perumal for his crucial remarks and help. Thank you! My gratitude also goes to Prof. Carolyn Ferguson, my outside chair committee.

Thanks to my lab mates Dr. Sophie Bouchet, Dr. Zhenbin Hu, Dr. Sandeep Marla, Dr. Olatoye Olalere, Jacques Faye, Dr. Terry Felderhoff, Dr. Kebede Muleta, Dr. Rubi Raymundo, Matt Davis, Dr. Sridevi Nakka, and Tarik Eluri. Thank you for the very beautiful and unforgettable moments spent together as well as the brainstorming sessions that we often had

during my Ph.D. You all were able to share with me your scientific knowledge, inspire me daily and give me your very relevant remarks.

To my colleagues from SAWAGEN: Dr. Abdou Harou, Dr. Aissata Mamadou, Dr. Bassirou Sine, Dr. Cyril Diatta, Dr. Daniel Fonceka, Dr. Falalou Hamidou, Dr. Jack Akata, Dr. Ndiaga Cisse, and Ousseini Ardaly. You all kept inspiring me and helping me. I would like to thank you for all your support! I am very much thankful to Dr. Jesse Lasky who hosted me at Penn State University.

I express my deepest gratitude to Kira Everhart-Valentin and Nancy Williams for their tremendous help in organizing travel and bureaucracy. Glad to work with you.

Because it takes a village to do a Ph.D., I would like to thank my friends and family in Kansas who have shown me full support during my stay in the state. Thank you, Assia and the Maroua! Thanks to you, I have a family in the Midwest. Thank you very much for these unforgettable and fun family moments.

Many thanks my family and friends! Thank you for everything!

Dedication

To those who have always been there in my life, those who have always believed in me, those who have always supported me and pushed me to go forward, those who will always be there even after this journey, those who keep inspiring me daily.

General Introduction

Smallholder farming systems are affected by various constraints limiting agricultural production in the context of growing populations in developing countries, notably Sahelian countries that have been highly exposed to climate change and climate variability over past decades (Sultan et al., 2019). Sorghum, along with pearl millet, is among the main staple food crops of these countries. Sorghum landraces, maintained and selected by smallholder farmers based on local practices and cultural preferences, contribute to food security in these countries (National Research Council, 1996; Walker and Alwang, 2015). Despite the global threat of agrobiodiversity erosion worldwide, a high diversity of sorghum and pearl millet landraces is maintained by farmers in West Africa (Deu et al., 2008; Bezançon et al., 2009). This diversity is a factor of resilience in the context of smallholder farming. Based on the model of pearl millet, previous studies of population genetics combined with association mapping showed that landraces evolve to adapt to climate change in response to environmental and farmers' selection (Vigouroux et al., 2011). Furthermore, the rapid changes in environmental and social contexts make it necessary to sustain the processes of selections made by farmers with the help of modern approaches to breeding. This need is critical since the agricultural response to growth of the human population is necessary due to land pressure and higher climate risks and will be based on sustainable productivity improvement.

Despite the progress and investment in crop improvement, only a few modern varieties are adopted by farmers. The lack of understanding and inclusion of crucial traits preferred by farmers contribute to this shortcoming (Ndjeunga et al., 2015; Walker and Alwang, 2015). The existing crop genetic diversity at the West African regional level could provide better insight into breeding programs, but this diversity remains insufficiently utilized.

In recent years, several studies were undertaken to identify the genetic basis of complex traits in crops and their subsequent utilization in crop improvement (Bernardo, 2008; Huang et al., 2010). With sequencing technology and statistical methods evolving rapidly, applications of marker-assisted selection and genomic selection are now possible and can be used in many breeding programs (Collard and Mackill, 2008; Bhat et al., 2016).

High-density genome-wide SNP markers provide a high-resolution mapping that facilitates marker discovery and application to crop improvement (Visendi et al., 2013; Thomson, 2014). However, the application of these frameworks to dissect the genetic and genomic bases of crop adaptation in local environments and characterize the genomics of farmer preferences is limited, especially in Sahelian countries. Sorghum is among the most cultivated cereals in West Africa and is an ideal crop for marginal lands (Olembo et al., 2010). Previous studies of West African collections held in genebanks were characterized thus, providing genomic resources of sorghum diversity in the region (Deu et al., 2006; Billot et al., 2013). Further studies analyzed the genetic bases for agronomic traits in sorghum (Cuevas et al., 2018; Olatoye et al., 2018; Faye et al., 2019).

Breeding programs of West Africa would benefit from recent technologies of outsourcing genotyping to better understand their germplasm and implement marker-assisted selection (Atlin et al., 2017). Unlocking the genetic basis of West African sorghum adaptation and response to stress through the development of trait predictive and cost-effective markers is of significance for the implementation of genomics-enabled breeding.

This thesis is organized into five chapters. The first chapter highlights a review of the state of crop improvement in West Africa with an emphasis on sorghum. Significant gaps and limitations in current sorghum breeding programs were discussed. Approaches and

methodologies for sorghum breeding and genomics that could assist in crop improvement for developing countries were addressed. The following chapters cover approaches for a better understanding of regional sorghum germplasm, discovering genomic regions associated with drought tolerance, and leverage genomic resources to assist breeding for parasitic weed resistance (Figure 1.1). Thus, to develop genomic resources for West African sorghum, the second chapter focuses on population genomics of sorghum germplasm to characterize its diversity existing in the West African country of Niger and to dissect the genetic basis for local adaptation and farmer preferences. Based on new genomic data developed, the third chapter focused on genome-wide association studies for a West African Sorghum Association Panel (WASAP) with phenotypic data obtained under a managed environment for well-watered and water deficit experiments. The fourth chapter describes the implementation of genomics-enabled breeding for the development of sorghum varieties resistant to the parasitic weed *Striga* hermonthica via low germination stimulant through the development of cost-effective Kompetitive Allele-Specific PCR (KASP) markers. The fifth chapter reviews future hypothesis-driven approaches to strengthen the capacity of varietal development, and the potential use of this methodology to assist small breeding programs.

Chapter 1 - Literature Review

The need for crop improvement in sub-Saharan Africa to ensure food security

Climate change is increasing variation in temperatures and precipitation. Increased temperatures and erratic rainfall will be prevalent in semi-arid regions (IPCC, 2014). Staple crops, mostly cultivated under traditional cropping systems with low inputs, will be affected by climate variability (Sarr, 2012). Direct consequences are observed and predicted by reduced food production throughout those regions that undermine food security (Lobell et al., 2008; Biasutti, 2019). Moreover, predicted human population growth will intensify food demand. In the absence of changes in crop productivity, this could highly increase the vulnerability of those regions.

The Sahel is a vast zone in Africa spanning from the desert of the Sahara in the North to the Sudanian climate zone in the South and is characterized by variability of precipitation, often uneven and unpredictable. The Sahel experienced periods of drought characterized by reduction and shifts in precipitation patterns for successive years (Glantz, 1987; Biasutti, 2019). The consequences of climate change in sub-Saharan Africa imply an increase in temperature, erratic rainfall leading to negative impacts on subsistence agriculture (Lobell et al., 2008; Enenkel et al., 2015). Additionally, West African countries will experience rainfall variability, increasing the vulnerability of the populations, dependent on agricultural production each year, and reducing the capacity of staple crops to maintain their yield at the end of the season. As a result, cultivated species are less productive and, at the same time, subject to other factors limiting production, such as insect attacks or invasive weeds (Mundia et al., 2019; Sultan et al., 2019). In the long term, staple crops may experience significant yield reductions in the smallholder farmers' environments threatening food and nutritional security (Lobell et al., 2008; Enenkel et al., 2015).

Another important constraint is land degradation and soil fertility management, particularly in the case of the Sahelian cropping systems which are largely based on low-input options (Abdoulaye and Sanders, 2005; Leiser et al., 2014).

Sustainable and efficient solutions are needed to tackle these constraints and threats to Sahelian agriculture and food security. These include a range of measures that will together contribute to the improvement and the resilience of crop productivity, notably the diversification of crop management options related to soil fertility management, pest management, cropping systems improvement and varietal diversity management. Designing new plant breeding strategies to improve major cereal crops in these regions (particularly pearl millet and sorghum) is among the promising options. Future plant breeding programs, in the context of sub-Saharan Africa with smallholder farming systems, have to be built with an emphasis on traits associated with local adaptation, resistance to biotic and abiotic stress, and with the socio-economic context. New sorghum varieties to be developed must satisfy and optimize the diverse requirements and preferences of the chain of end-users (farmers, food processors, traders, consumers) while responding to constraints linked to climate variability, local adaptation, and local crop management options. New technologies including molecular ones could enhance our understanding of the basis of trait variation as well as our capacities for breeding.

Next-generation technologies accelerate germplasm characterization and facilitate variety development

Next-generation sequencing provides an efficient method in exploiting genomic markers for the characterization of crops and their utilization in breeding programs (Varshney et al., 2005; Bhat et al., 2016). Genomic approaches used to dissect the genetic basis of relevant agronomic traits include an understanding of genome-wide patterns of nucleotide variation as

well as selection signatures shaping local adaptation and population differentiation (Meyer et al., 2016; Bandillo et al., 2017). Recent genotyping methods provide sequencing of a large number of genotypes at a low cost per sample. Genotyping-by-sequencing (GBS) takes advantage of reduced representation libraries using restriction enzyme cut sites and multiplexed samples (Elshire et al., 2011). Studies on population and quantitative genomics of staple crops were conducted using GBS (Morris et al., 2013; Lasky et al., 2015; Romero Navarro et al., 2017). Most of the collections characterized group global germplasms of staple crops. These collections generally provide great coverage of global diversity but do not include, for lack of cost-effectiveness, a large sample size to underpin the studies at smaller or median scales (regional scales). A smaller number of studies described germplasm collections to understand signatures of selections, patterns of local adaptation and evolution of cultivated species at a regional scale. On the other hand, exploring the genetic diversity has also made possible the application of genomic data in marker-assisted breeding and genomic selection (Collard et al., 2005; Collard and Mackill, 2008). Such methodologies can be implemented in breeding programs of West Africa through outsourced genotyping such as Kompetitive Allele-Specific PCR (KASP) (Burow et al., 2019).

Regional crop germplasms that harbor relevant sources of variation are maintained in genebanks. Genebanks play an important role in the conservation and distribution of germplasm over decades providing ex-situ genetic characterization and sources of germplasm for crop improvement (Singh et al., 2012; Upadhyaya et al., 2017; Milner et al., 2018). Several accessions that are in multiple genebanks and are freely available, could be explored to characterize global and regional germplasm where the genetic diversity remains untapped. The majority of the accessions being held, mostly composed of landraces, are important sources of resistance or

tolerance to stressors and adapted for farmers' environments. The potential of conservation in a local environment (in-situ) is also critical in understanding the dynamics of crop adaptation and domestication over the years (Teshome et al., 1999; Deu et al., 2008; Zonneveld et al., 2014). Altogether, both ex and in situ collection capture a large amount of genetic diversity, crop evolution, and adaptation to changing climates.

Sorghum crop: botanical classification, agroclimatic distribution, diversity, and constraints

Sorghum [*Sorghum bicolor* (L) Moench, 2n=20] is a staple crop for millions of people worldwide. It is a versatile C4 cereal of the Poaceae family (grass), probably originated from East Africa (Harlan and de Wet, 1972; Doggett, 1988; Smith and Frederiksen, 2000; Billot et al., 2013). Its wide adaptation allows for growing on marginal lands and in a wide range of environments compared to other cereal crops from humid and semi-arid regions around the world spreading across Africa, the Middle East, and Asia (Smith and Frederiksen, 2000; Paterson et al., 2009; Kimber et al., 2013). Mainly used as food in many countries, sorghum production ranks fifth among major world cereal crops, with a total production of 57 million tons in 2017 (FAO, 2020), is also a source of bioenergy and forage worldwide (Smith and Frederiksen, 2000).

Sorghum has five botanical types (bicolor, caudatum, durra, guinea, and kafir) and ten intermediate types distributed across the world, classified based on the panicle shape, spikelet morphology, and grain shape (Doggett 1988; Smith and Frederiksen 2000; Kimber et al. 2013). The bicolor type is the primitive type distributed in Africa and Asia. Caudatum type is mostly found in the western and eastern Africa and harbors agronomic importance due to its high yield, medium height, and seed quality (Smith and Frederiksen, 2000). The durra type, having a dense panicle, originating from eastern Africa is adaptive to dry environments with low rainfall, where

areas of production expand to West Africa, the Middle East, and South Asia (Smith and Frederiksen, 2000; Billot et al., 2013; Soler et al., 2013). The guinea type, having long and loose panicles, is cultivated in humid regions of West Africa, South Africa, and East Asia, and is characterized by being low yielding. The kafir type is predominantly found in South Africa, with a medium plant height and it was part of the US sorghum conversion programs (Stephens et al., 1967; Smith and Frederiksen, 2000). Sorghum landraces harbor genetic footprints of domestication and local adaptation traits that were used in crop improvement (Kimber et al., 2013; Morris et al., 2013; Lasky et al., 2015). This diversity and adaptation of sorghum may improve its resilience under specific agroclimatic zones and cultural practices and promising food and bioenergy for people around the world.

Sorghum genome

Crop reference genomes advance understanding of the evolution and function of genes (Jackson et al., 2011). Sorghum harbors a large phenotypic variability with five botanical types and ten intermediates (Harlan and de Wet, 1972). The sorghum genome fully sequenced with the US breeding line BTx623 was used for the whole genome shotgun (Paterson et al., 2009). Sorghum is a diploid crop (2n=20) with a genome size of 730 Mb and more than 36,000 genes (Paterson et al., 2009; McCormick et al., 2018). Its moderate genome size with ten chromosomes and low gene duplication has enabled understating of genetic structure and function of genes and is a model crop for C4 cereals (Moore and Ming, 2008; Paterson et al., 2009; Kimber et al., 2013). Sorghum exhibits an extensive genetic diversity worldwide that is still untapped by using recent methodologies. Genes associated with flowering time, plant height, and grain pigmentation were cloned and provided a step forward in the understanding of the versatile crop.

Sorghum has genome-based data available that unlock the potential of genetic diversity in the QTL Atlas (Mace et al., 2018).

A key abiotic constraint in sorghum: drought

Drought is a major abiotic constraint in sub-Saharan African rainfed agriculture. Drought in these regions can differ in the timing and severity throughout the crop growing season (Glantz, 1987). Major responses to drought were reported in staple crops such as the avoidance, the escape and the tolerance mechanism (Tuinstra et al., 1997; Barnabás et al., 2008). Studies have shown that under post-flowering drought stress, delayed leaf senescence would be favorable for sorghum during grain filling (Assefa et al., 2010; Beyene et al., 2015; Badigannavar et al., 2018). The stay-green trait is a delayed mechanism of senescence for some genotypes during grain filling under drought conditions (Blum et al., 1989; Vadez et al., 2011). During the terminal drought, stay-green lines maintain their photosynthetic activity longer than non stay-green lines (Borrell et al., 2014b). In sorghum, four major stay-green QTLs were reported allowing for marker-assisted selection of drought tolerance with known stay-green lines (Harris et al., 2006; Borrell et al., 2014a). Efforts to develop drought-tolerant varieties were since possible with the known stay-green genotypes and were marker-associated with drought tolerance (Jordan et al., 2012; Ouedraogo et al., 2017). However, existing phenotyping and screening approaches in semi-arid regions are not sufficient to improve locally preferred varieties in addition to the shortcomings of replicating smallholder farmers' field conditions. Also, the use of drought tolerance sources previously developed as donor parents might have undesirable traits for specific breeding programs with limited resources. This is the case of BTx642, a post-flowering drought-tolerant line with dwarf height, which is not the ideotype for multipurpose sorghum production (Borrell et al., 2014a). BTx642 has four stay-green alleles conferring tolerance to

post-flowering drought, and little is known about the alleles present in the West African sorghum germplasm. Dissecting the genetic basis of drought stress in diverse germplasm would elucidate appropriate approaches and markers for accelerating sorghum improvement in sub-Saharan Africa.

A key biotic constraint in sorghum: *Striga hermonthica*

Among all the biotic limiting factors, *Striga hermonthica* (known as witchweed) damages crop production the most in sub-Saharan Africa and tolerates a wide range of climatic and soil conditions leaving smallholder farmers vulnerable (Ejeta, 2007; Ejeta and Gressel, 2007). *Striga hermonthica* (2n=38) is an obligate root parasite that infests sorghum, millet, and corn. *Striga*, as an annual parasitic plant, requires a plant host to complete its life cycle (Scholes and Press, 2008; Westwood, 2013). *Striga* likely originated from Sudan and Ethiopia, also a center of origin of the host plant *Sorghum bicolor* (Nations, 1989; Scholes and Press, 2008). The germination of *Striga* seeds requires root exudates, secondary metabolites from the host known as strigolactones (Scholes and Press, 2008; Yoder and Scholes, 2010). Strigolactones are endogenous hormones that play a role in mycorrhizal symbiotic association and control of shoot branching in plants (Scholes and Press, 2008). Upon germination, *Striga* seeds develop haustorial initiation to attach to the host plant through the root cortex for parasite survival. This attachment establishes the parasite-host connection resulting in water and nutrient uptake by the parasite. The host's growth and development are reduced, causing severe damage. The weed parasite is able to perform photosynthesis and complete its life cycle by producing ~50,000 seeds from a single plant, viable in the soil for more than ten years (Ejeta, 2007; Ejeta and Gressel, 2007). Known *Striga* resistant lines have different mechanisms preventing parasite development. For instance, Framida and SRN39 produce low germination stimulants thus the *Striga* germination does not happen while

N13 has a mechanical barrier resistance that prevents the haustorial attachment to the host plant (Haussmann et al., 2000a; Ejeta and Gressel, 2007).

Integrated *Striga* management and control include the use of *Striga* resistant lines besides the limitation of *Striga* development and seed production (Haussmann et al., 2000b; Ejeta, 2007). The control of the weed by farmers consists of using trap crops, hoeing or crop rotation to reduce yield loss, prevent seed dissemination and propagation, but often abandonment of the field may be necessary (Haussmann et al., 2000b; Kountche et al., 2017). However, those controls are not sufficient to fight against the devastating parasite in multiple *Striga* prone areas. Thus, investigating the genetic diversity of crops is of significance in developing *Striga* resistant varieties for smallholder farmers. This could provide a great contribution to integrated management strategies.

Earlier breeding efforts in West Africa

Breeding efforts to develop new high yielding varieties with preferred agronomic characteristics have been made over the last decades, mainly with traditional breeding approaches. International research organizations started sorghum improvement programs in the early 1950s (J.-C. MAUBOUSSIN et al., 1977). With the late establishment of research centers in West Africa, sorghum breeding program efforts led to the development of new, resistant, and high yielding varieties. Targeted traits include characteristics for local adaptation, such as the wide range of maturity groups to fit the local environment, plant height, panicle length, and grain weight (Walker and Alwang, 2015). Increasing yield in specific regions was possible with improved sorghum varieties in the Sahel. For instance, IRAT 204 and SRN39 were released for their performance during drought stress and *Striga* infestation respectively across West Africa (Lynam et al., 2011). Some new varieties show a lack of adoption in specific regions resulting in

a lower rate of adoption. Estimates of adoption were based on the percentage of cultivated areas of the improved varieties in comparison with the total cultivated area (Olembo et al., 2010; Walker and Alwang, 2015). In Niger, improved varieties cover 15% of the total cultivated area (Walker and Alwang, 2015).

Hybrid breeding was also part of the West African breeding program (Andrews, 1975; Yohe et al., 2005; Christiansen, 2009). In West Africa, hybrid breeding has been successful in improving local landraces for grain yield (Andrews, 1975; Rattunde et al., 2013). In Niger, NAD1 hybrid (Tx623A x MR732) was among the released varieties for Sahelian agroclimatic zones through the collaboration with INTSORMIL in the early 1980s (Ministère de l'agriculture du Niger, 2012). The NAD1 yields over two tons per hectare, higher than the local check Mota Maradi (Kapran et al., 2007). Malian breeding program developed photoperiod-sensitive hybrid varieties from guinea landraces adapted to climatic zones with acceptable grain quality (Kante et al., 2017).

West African agriculture, most of the time, relies on rainfed cropping systems and subsistence agriculture, especially in smallholder farming systems. Crop improvement in sub-Saharan Africa was weakened by the heterogeneity and lack of adaptation to the agroclimatic zones (Lynam *et al.* 2011; Walker and Alwang 2015). Thus, research toward the development of varieties that fit the needs of local smallholder farmers remains a major priority. One of the criteria for farmers is to have improved varieties be the best fit for their agroclimatic zones while still having their preferred traits. In this context, studies have been conducted in different locations to capture local adaptation. Research has been done in Africa on sorghum and millet through collaborations with international programs such as INTSORMIL and internationally

established institutions including ICRISAT, ORSTOM (France, current IRD) and IRAT (France, current CIRAD).

In Niger, sorghum is the second most cultivated cereal after pearl millet. Local varieties have low yield potential and are susceptible to major stressors. Farmers rely mostly on landraces that are low yielding even if they hold desirable traits of local adaptation and local preferences (Deu et al., 2008). Traditional breeding has significantly contributed to sorghum improvement in Niger (Olembo et al., 2010). Efforts were made by INRAN to deliver improved varieties, being produced by seed companies, with acceptable yields under specific stressors such as drought tolerance (e.g. NAD1) and midge resistance (e.g. SSD 35). While the multiplication of improved varieties in Niger has been mainly done by public institutes, some of these varieties are being produced by seed companies in recent years. Furthermore, the lack of varieties that are adapted to multiple climatic zones or a broader adaptation in the released germplasm makes sorghum production challenging with a lower rate of adoption compared to the total cultivated area (Purdue University, 1992; Olembo et al., 2010; Walker and Alwang, 2015). One particular resistance trait is not sufficient to drive adoption, and farmers were not always sufficiently involved during the breeding process (e.g. specific needs, traits preferences). Taking into account relevant agronomic traits for farmers' preferences, cultural practices and traits associated with adaptation to local environments might enhance new variety suitability and thus, farmers' acceptance.

Future crop improvement strategies: lessons from earlier breeding programs

The questions that remain to be answered revolve around how to make small breeding programs more efficient, speed up crop improvement efforts, deliver new varieties to end-users that meet their needs, and ensure that farmers have the best varieties for their current and future

environments. Poor crop production in West Africa is often associated with poor soil fertility, and limited use of fertilizers by farmers (Abdoulaye and Sanders, 2005; Leiser et al., 2014). The limited use of recent advancements is one of the major drawbacks in small breeding programs (Bhat et al., 2016). Furthermore, the need for agroecological transition worldwide makes it critical to consider varietal traits that are suited for sustainable intensification and adapted to cropping systems with a balanced use of inputs, limiting particularly chemical (external) inputs. Besides slow varietal development, the lack of information about new varieties and their lack of accessibility remain prevalent (Walker and Alwang, 2015). Integrating multidisciplinary approaches is significant for successful crop improvement in sub-Saharan Africa to enhance variety suitability and to promote adoption and appreciation by end-users.

For collaborative research, both basic and applied research are essential for crop improvement (Chambers, 1994; Atlin et al., 2017). Knowledge of end-user preferences for improved varieties is of great importance for breeding in sub-Saharan Africa. Several breeding programs are using participatory rural appraisal (PRA), an approach to facilitate knowledge sharing and emphasize interactions with farmers, to guide selection and targets for specific end-users and environments. PRA investigates farmers' expectations, needs, and perceptions of new varieties through focus groups and semi-structured interviews to emphasize their role in variety selection (Chambers, 1994; Cornwall and Pratt, 2011). Thus, the adoption will be driven by popular demand instead of delivering varieties without consideration of farmer and stakeholder needs.

Strong collaboration with other disciplines has enhanced the effectiveness of breeding programs in many countries (Crossa et al., 2014; Hammer et al., 2014). Taking advantage of cost-effective marker technology, West African breeding programs should effectively use new

technologies through collaborations and partnerships. Strategies of genomic-assisted breeding may incorporate diverse studies and multi-environment trials for a better knowledge of the genetic basis of crop adaptation and marker development (Tuberosa and Salvi, 2006). Marker-assisted selection would be valuable in breeding programs to accelerate varietal development in targeted environments by outsourcing genotyping services such as KASP genotyping.

Research and development towards genomics-enabled breeding in West Africa

Under Agricultural Research Systems (NARS), breeding programs sought to improve staple crops by increasing yields under harsh environments and ensuring food security in a changing environment for a growing population. Some of the breeding programs lack capacity for demand-driven crop improvement (Atlin et al., 2017). Future research should focus more on understanding the needs of end-users through PRA on the one hand and accelerate breeding cycles to deliver new varieties on the other hand. PRA also orients scientists to identify, prioritize major constraints, and develop efficient breeding product profiles that guide variety development and participatory selection (Atlin et al., 2001; Ceccarelli and Grando, 2020).

Integrative approaches to strengthen the capacities of small breeding programs to develop and deliver better varieties to smallholder farmers are needed. Recently, international connections are not a major issue for some programs since collaborations are established for many years. Those studies allowed us to identify genomic regions associated with local adaptation to diverse agroclimatic zones, farmer preferences, and footprints of selection in West African sorghum. In the breeding perspective, local adaptation plays an essential role, especially when multiple precipitation patterns and photoperiods are involved (Kouressy et al., 2008). The genetic architecture of sorghum drought tolerance in West Africa remains challenging while

climate change is shifting precipitation patterns affecting periods of drought throughout the growing season.

The development of new varieties should account for multiple factors in a wide range of environments most of the time under complex cultural management. For *Striga hermonthica*, the mechanism of adaptation and selection pressure of the parasite are increasing the rate of infestation in the spatially variable farmer's field. Thus, it is not sufficient to maintain one or two *Striga* resistant genotypes for diverse environments in sub-Saharan Africa, SRN39 and N13 commonly used with a low rate of adoption (Walker and Alwang, 2015). Accounting for adaptation of targeted environment and farmer preferences, breeders should effectively select for *Striga* resistance that would drive adoption. Those two varieties have different mechanisms of resistance, production of low germination stimulants and mechanical barriers, have not been combined yet to enhance resistance under highly variable environments and diverse management systems in West African crop production. The low germination stimulant does not prevent haustorial attachment with the host roots facilitated by the root cortex (Yoshida Satoko and Shirasu Ken, 2009). The genetic resistance (both *LGS1* and mechanical barrier) in the West African environment might be significant with integrated management of *Striga*, as technical and social managements, and environmental interaction are involved.

Small breeding programs have limited capacity for multi-location trials to account for genotype by environment interactions. There is a need to define targeted population environments on the local scale and take advantage of the broad adaptation of the variety through collaborations. Future crop improvements may incorporate interdisciplinary teams and multi-environmental trials (Atlin et al., 2017). Building a strong network with stakeholders, researchers, and regional and international partnerships will be required to accelerate the varietal

development of staple crops in West Africa. We proposed a conceptual roadmap for genomics-enabled breeding (Figure 1.1). This framework structured and connected the chapters within this thesis.

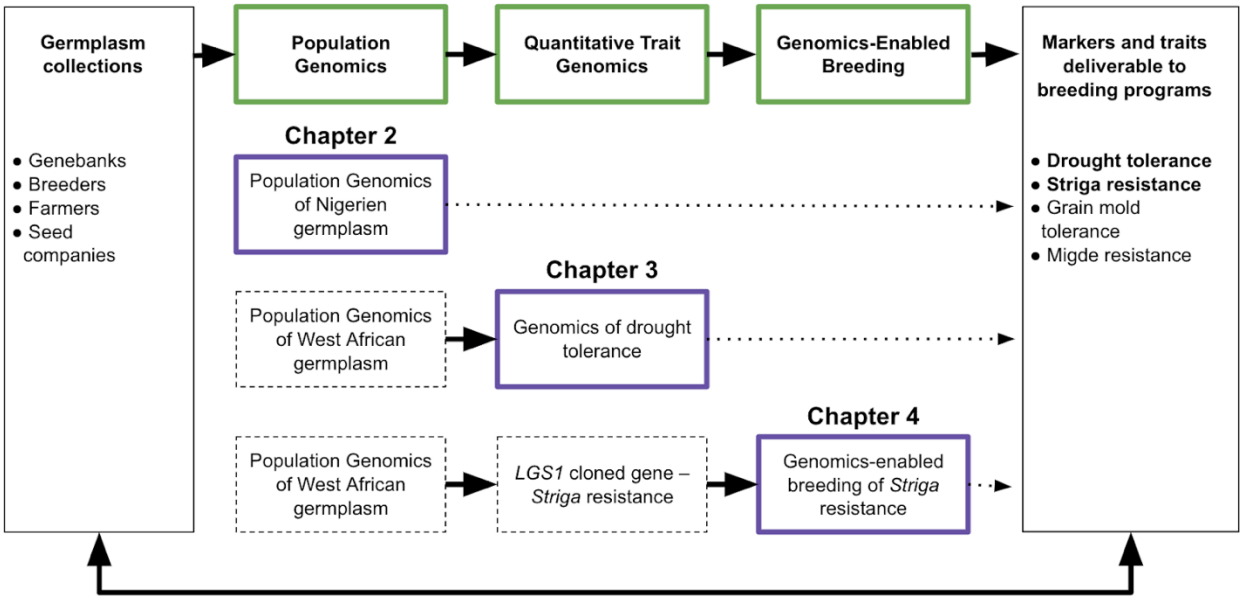


Figure 1.1. Conceptual roadmap for genomics-enabled breeding.

Green boxes represent the specific activities to deliver markers and traits to breeding programs. Purple boxes highlight chapters two, three, and four covered in this thesis. Dashed boxes summarize genomic resources leveraged in each chapter.

References

Abdoulaye, T., and J.H. Sanders. 2005. Stages and determinants of fertilizer use in semiarid African agriculture: the Niger experience. Agricultural Economics 32:167–179. doi:10.1111/j.0169-5150.2005.00011.x

Andrews, D.J. 1975. Sorghum Grain Hybrids in Nigeria. Experimental Agriculture 11:119–127. doi:10.1017/S0014479700006566

Assefa, Y., S. Staggenborg, and P.V.V. Prasad. 2010. Grain Sorghum Water Requirement and Responses to Drought Stress: A Review. Crop Management. doi:10.1094/CM-2010-1109-01-RV

Atlin, G.N., J.E. Cairns, and B. Das. 2017. Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Global Food Security 12:31–37. doi:10.1016/j.gfs.2017.01.008

Atlin, G.N., M. Cooper, and Å. Bjørnstad. 2001. A comparison of formal and participatory breeding approaches using selection theory. Euphytica 122:463–475. doi:10.1023/A:1017557307800

Badigannavar, A., N. Teme, A.C. de Oliveira, G. Li, M. Vaksmann, V.E. Viana, T.R. Ganapathi, and F. Sarsu. 2018. Physiological, genetic and molecular basis of drought resilience in sorghum [Sorghum bicolor (L.) Moench]. Ind J Plant Physiol. 23:670–688. doi:10.1007/s40502-018-0416-2

Bandillo, N.B., J.E. Anderson, M.B. Kantar, R.M. Stupar, J.E. Specht, G.L. Graef, and A.J. Lorenz. 2017. Dissecting the Genetic Basis of Local Adaptation in Soybean. Scientific Reports 7:17195. doi:10.1038/s41598-017-17342-w

Barnabás, B., K. Jäger, and A. Fehér. 2008. The effect of drought and heat stress on reproductive processes in cereals. Plant, Cell & Environment 31:11–38. doi:10.1111/j.1365-3040.2007.01727.x

Bernardo, R. 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Science 48:1649–1664. doi:10.2135/cropsci2008.03.0131

Beyene, A., S. Hussien, T. Pangirayi, and L. Mark. 2015. Physiological mechanisms of drought tolerance in sorghum, genetic basis and breeding methods: A review. African Journal of Agricultural Research 10:3029–3040. doi:10.5897/AJAR2015.9595

Bezançon, G., J.-L. Pham, M. Deu, Y. Vigouroux, F. Sagnard, C. Mariac, I. Kapran, A. Mamadou, B. Gérard, J. Ndjeunga, and J. Chantereau. 2009. Changes in the diversity and geographic distribution of cultivated millet (Pennisetum glaucum (L.) R. Br.) and sorghum (Sorghum bicolor (L.) Moench) varieties in Niger between 1976 and 2003. Genetic Resources and Crop Evolution 56:223–236. doi:10.1007/s10722-008-9357-3

Bhat, J.A., S. Ali, R.K. Salgotra, Z.A. Mir, S. Dutta, V. Jadon, A. Tyagi, M. Mushtaq, N. Jain, P.K. Singh, G.P. Singh, and K.V. Prabhu. 2016. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. Front Genet 7. doi:10.3389/fgene.2016.00221

Biasutti, M. 2019. Rainfall trends in the African Sahel: Characteristics, processes, and causes. Wiley Interdisciplinary Reviews: Climate Change 10:e591. doi:10.1002/wcc.591

Billot, C., P. Ramu, S. Bouchet, J. Chantereau, M. Deu, L. Gardes, J.-L. Noyer, J.-F. Rami, R. Rivallan, Y. Li, P. Lu, T. Wang, R.T. Folkertsma, E. Arnaud, H.D. Upadhyaya, J.-C. Glaszmann, and C.T. Hash. 2013. Massive Sorghum Collection Genotyped with SSR Markers to Enhance Use of Global Genetic Resources. PLOS ONE 8:e59714. doi:10.1371/journal.pone.0059714

Blum, A., J. Mayer, and G. Golan. 1989. Agronomic and physiological assessments of genotypic variation for drought resistance in sorghum. Aust. J. Agric. Res. 40:49–61. doi:10.1071/ar9890049

Borrell, A.K., J.E. Mullet, B. George-Jaeggli, E.J. van Oosterom, G.L. Hammer, P.E. Klein, and D.R. Jordan. 2014a. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. J. Exp. Bot. eru232. doi:10.1093/jxb/eru232

Borrell, A.K., E.J. Oosterom, J.E. Mullet, B. George-Jaeggli, D.R. Jordan, P.E. Klein, and G.L. Hammer. 2014b. Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. New Phytologist 203:817–830

Burow, G., R. Chopra, H. Hughes, Z. Xin, and J. Burke. 2019. Marker Assisted Selection in Sorghum Using KASP Assay for the Detection of Single Nucleotide Polymorphism/Insertion Deletion. Z.-Y. Zhao and J. Dahlberg, ed. Methods in Molecular Biology. Springer New York, New York, NY.

Ceccarelli, S., and S. Grando. 2020. Participatory plant breeding: Who did it, who does it and where?. Experimental Agriculture 56:1–11. doi:10.1017/S0014479719000127

Chambers, R. 1994. The origins and practice of participatory rural appraisal. World Development 22:953–969. doi:10.1016/0305-750X(94)90141-4

Christiansen, K. 2009. INTSORMIL Program Doubles Sorghum and Millet Yields in Mali, West Africa. INTSORMIL Impacts and Bulletins

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 169–196

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. doi:10.1098/rstb.2007.2170

Cornwall, A., and G. Pratt. 2011. The use and abuse of participatory rural appraisal: reflections from practice. Agric Hum Values 28:263–272. doi:10.1007/s10460-010-9262-1

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

Cuevas, H.E., L.K. Prom, and G. Rosa-Valentin. 2018. Population structure of the NPGS Senegalese sorghum collection and its evaluation to identify new disease resistant genes. PLOS ONE 13:e0191877. doi:10.1371/journal.pone.0191877

Deu, M., F. Rattunde, and J. Chantereau. 2006. A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49:168–180. doi:10.1139/g05-092

Deu, M., F. Sagnard, J. Chantereau, C. Calatayud, D. Hérault, C. Mariac, J.-L. Pham, Y. Vigouroux, I. Kapran, P.S. Traore, A. Mamadou, B. Gerard, J. Ndjeunga, and G. Bezançon. 2008. Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. Theor. Appl. Genet. 116:903–913. doi:10.1007/s00122-008-0721-7

Doggett, H. 1988. Sorghum. 2nd Edition.

Ejeta, G. 2007. Breeding for Resistance in Sorghum: Exploitation of an Intricate Host–Parasite Biology. Crop Science 47:S-216-S-227. doi:10.2135/cropsci2007.04.0011IPBS

Ejeta, G., and J. Gressel. 2007. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt. World Scientific Publishing Co Pte Ltd, Singapore, SINGAPORE.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6:e19379. doi:10.1371/journal.pone.0019379

Enenkel, M., L. See, R. Bonifacio, V. Boken, N. Chaney, P. Vinck, L. You, E. Dutra, and M. Anderson. 2015. Drought and food security – Improving decision-support via new technologies and innovative collaboration. Global Food Security 4:51–55. doi:10.1016/j.gfs.2014.08.005

FAO. 2020. FAOSTAT. http://www.fao.org/faostat/en/#home (accessed February 21, 2020).

Faye, J.M., F. Maina, Z. Hu, D. Fonceka, N. Cisse, and G.P. Morris. 2019. Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. Ecology and Evolution 9:6038–6051. doi:10.1002/ece3.5187

Glantz, M.H. 1987. Drought in Africa. Scientific American 256:34–41

Hammer, G.L., G. McLean, S. Chapman, B. Zheng, A. Doherty, M.T. Harrison, E. van Oosterom, and D. Jordan. 2014. Crop design for specific adaptation in variable dryland production environments. Crop Pasture Sci. 65:614–626. doi:10.1071/CP14088

Harlan, J.R., and J.M.J. de Wet. 1972. A Simplified Classification of Cultivated Sorghum1. Crop Science 12:172. doi:10.2135/cropsci1972.0011183X001200020005x

Harris, K., P. Subudhi, A. Borrell, D. Jordan, D. Rosenow, H. Nguyen, P. Klein, R. Klein, and J. Mullet. 2006. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. Journal of Experimental Botany 58:327–338. doi:10.1093/jxb/erl225

Haussmann, B.I.G., D.E. Hess, B.V.S. Reddy, H.G. Welz, and H.H. Geiger. 2000a. Analysis of resistance to <Emphasis Type="BoldItalic">Striga hermonthica</Emphasis> in diallel crosses of sorghum. Euphytica 116:33–40. doi:10.1023/A:1004046001009

Haussmann, B.I.G., D.E. Hess, H.-G. Welz, and H.H. Geiger. 2000b. Improved methodologies for breeding striga-resistant sorghums. Field Crops Research 66:195–211. doi:10.1016/S0378-4290(00)00076-9

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q.-F. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967. doi:10.1038/ng.695

IPCC. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Jackson, S.A., A. Iwata, S.-H. Lee, J. Schmutz, and R. Shoemaker. 2011. Sequencing crop genomes: approaches and applications. New Phytologist 191:915–925. doi:10.1111/j.1469-8137.2011.03804.x

J.-C. MAUBOUSSIN, I. GUEYE, and M. N'DIAYE*. 1977. AGRONOMIE TROPICALE XXXll

Jordan, D.R., C.H. Hunt, A.W. Cruickshank, A.K. Borrell, and R.G. Henzell. 2012. The Relationship Between the Stay-Green Trait and Grain Yield in Elite Sorghum Hybrids Grown in a Range of Environments. Crop Science 52:1153–1161. doi:10.2135/cropsci2011.06.0326

Kante, M., H.F.W. Rattunde, W.L. Leiser, B. Nebié, B. Diallo, A. Diallo, A.O. Touré, E. Weltzien, and B.I.G. Haussmann. 2017. Can Tall Guinea-Race Sorghum Hybrids Deliver

Yield Advantage to Smallholder Farmers in West and Central Africa?. Crop Science 57:833–842. doi:10.2135/cropsci2016.09.0765

Kapran, I., C. Grenier, and G. Ejeta. 2007. Introgression of genes for Striga resistance into African landraces of sorghum. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt 129–141

Kimber, C.T., J.A. Dahlberg, and S. Kresovich. 2013. The Gene Pool of Sorghum bicolor and Its Improvement. A.H. Paterson, ed. Plant Genetics and Genomics: Crops and Models. No. 11. Springer New York.

Kountche, B.A., S. Al-Babili, and B.I.G. Haussmann. 2017. Striga: A Persistent Problem on Millets. Elsevier.

Kouressy, M., M. Dingkuhn, M. Vaksmann, and A.B. Heinemann. 2008. Adaptation to diverse semi-arid environments of sorghum genotypes having different plant type and sensitivity to photoperiod. Agricultural and Forest Meteorology 148:357–371. doi:10.1016/j.agrformet.2007.09.009

Lasky, J.R., H.D. Upadhyaya, P. Ramu, S. Deshpande, C.T. Hash, J. Bonnette, T.E. Juenger, K. Hyma, C. Acharya, S.E. Mitchell, E.S. Buckler, Z. Brenton, S. Kresovich, and G.P. Morris. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. Science Advances 1:e1400218. doi:10.1126/sciadv.1400218

Leiser, W.L., H.F.W. Rattunde, E. Weltzien, N. Cisse, M. Abdou, A. Diallo, A.O. Tourè, J.V. Magalhaes, and B.I. Haussmann. 2014. Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African Sorghum. BMC Plant Biology 14:206. doi:10.1186/s12870-014-0206-6

Lobell, D.B., M.B. Burke, C. Tebaldi, M.D. Mastrandrea, W.P. Falcon, and R.L. Naylor. 2008. Prioritizing Climate Change Adaptation Needs for Food Security in 2030. Science 319:607–610. doi:10.1126/science.1152339

Lynam, J., E. Gilbert, E. Howard, and F. Bliss. 2011. Evolving a plant breeding and seed system in sub-Saharan Africa in an era of donor dependence : A report for the Global Partnership Initiative for Plant Breeding Capacity Building

Mace, E., D. Innes, C. Hunt, X. Wang, Y. Tao, J. Baxter, M. Hassall, A. Hathorn, and D. Jordan. 2018. The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. Theor Appl Genet. doi:10.1007/s00122-018-3212-5

McCormick, R.F., S.K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, M. Amirebrahimi, B.D. Weers, B. McKinley, A. Mattison, D.T. Morishige, J. Grimwood, J. Schmutz, and J.E. Mullet. 2018. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal 93:338–354. doi:10.1111/tpj.13781

Meyer, R.S., J.Y. Choi, M. Sanches, A. Plessis, J.M. Flowers, J. Amas, K. Dorph, A. Barretto, B. Gross, D.Q. Fuller, I.K. Bimpong, M.-N. Ndjiondjop, K.M. Hazzouri, G.B. Gregorio, and M.D. Purugganan. 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat Genet 48:1083–1088. doi:10.1038/ng.3633

Milner, S.G., M. Jost, S. Taketa, E.R. Mazón, A. Himmelbach, M. Oppermann, S. Weise, H. Knüpffer, M. Basterrechea, P. König, D. Schüler, R. Sharma, R.K. Pasam, T. Rutten, G. Guo, D. Xu, J. Zhang, G. Herren, T. Müller, S.G. Krattinger, B. Keller, Y. Jiang, M.Y. González, Y. Zhao, A. Habekuß, S. Färber, F. Ordon, M. Lange, A. Börner, A. Graner, J.C. Reif, U. Scholz, M. Mascher, and N. Stein. 2018. Genebank genomics highlights the diversity of a global barley collection. Nature Genetics 1. doi:10.1038/s41588-018-0266-x

Ministère de l'agriculture du Niger. 2012. Catalogue national des espèces et variétés végétales du Niger

Moore, P.H., and R. Ming eds. 2008. Genomics of Tropical Crop Plants. Plant Genetics and Genomics: Crops and Models. Springer-Verlag, New York.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. PNAS 110:453–458. doi:10.1073/pnas.1215985110

Mundia, C.W., S. Secchi, K. Akamani, and G. Wang. 2019. A Regional Comparison of Factors Affecting Global Sorghum Production: The Case of North America, Asia and Africa's Sahel. Sustainability 11:1–18

National Research Council. 1996. Lost Crops of Africa: Volume I: Grains

Nations, F. and A.O. of the U. 1989. Striga, Improved Management in Africa: Proceedings of the FAO/OAU All-Africa Government Consultation on Striga Control, Maroua, Cameroon, 20-24 October 1988. Food & Agriculture Org.

Ndjeunga, J., K. Mausch, and F. Simtowe. 2015. Assessing the effectiveness of agricultural R&D for groundnut, pearl millet, pigeonpea and sorghum in West and Central Africa and East and Southern Africa. T.S. Walker and J. Alwang, ed. CABI, Wallingford.

Olatoye, M.O., Z. Hu, F. Maina, and G.P. Morris. 2018. Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. G3: Genes, Genomes, Genetics 8:3269–3281. doi:10.1534/g3.118.200551

Olembo, K.N., F. M'mboyi, S. Kiplagat, J.K. Sitieney, and F.K. Oyugi. 2010. Sorghum Breeding in Sub-Saharan Africa: The Success Stories. African Biotechnology Stakeholders Forum

Ouedraogo, N., J. Sanou, V. Gracen, and P. Tongoona. 2017. Incorporation of stay-green Quantitative Trait Loci (QTL) in elite sorghum (Sorghum bicolor L. Moench) variety through marker-assisted selection at early generation. Journal of Applied Biosciences 111:10867–10876. doi:10.4314/jab.v111i1.3

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D.G. Peterson, Mehboob-ur-Rahman, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, and D.S. Rokhsar. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. doi:10.1038/nature07723

Purdue University. 1992. NIGER APPLIED AGRICULTURAL RESEARCH PROJECT. USAID.

Rattunde, H.F.W., E. Weltzien, B. Diallo, A.G. Diallo, M. Sidibe, A.O. Touré, A. Rathore, R.R. Das, W.L. Leiser, and A. Touré. 2013. Yield of Photoperiod-sensitive Sorghum Hybrids Based on Guinea-race Germplasm under Farmers' Field Conditions in Mali. Crop Science 53:2454–2461. doi:10.2135/cropsci2013.03.0182

Romero Navarro, J.A., M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, A. Terron, H.V. Delgado, V. Vidal, A. Ortega, A.E. Banda, N.O.G. Montiel, I. Ortiz-Monasterio, F.S. Vicente, A.G. Espinoza, G. Atlin, P. Wenzl, S. Hearne, and E.S. Buckler. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480. doi:10.1038/ng.3784

Scholes, J.D., and M.C. Press. 2008. Striga infestation of cereal crops – an unsolved problem in resource limited agriculture. Current Opinion in Plant Biology 11:180–186. doi:10.1016/j.pbi.2008.02.004

Singh, A.K., K.S. Varaprasad, and K. Venkateswaran. 2012. Conservation Costs of Plant Genetic Resources for Food and Agriculture: Seed Genebanks. Agric Res 1:223–239. doi:10.1007/s40003-012-0029-3

Smith, C.W., and R.A. Frederiksen. 2000. Sorghum, Origin, History, Technology and Production. Wiley Series in Crop Science. SB191.S7 S668 2000. Texas A & M University.

Soler, C., A.-A. Saidou, T. Vi Cao Hamadou, M. Pautasso, J. Wencelius, and H.H.I. Joly. 2013. Correspondence between genetic structure and farmers' taxonomy – a case study from dry-season sorghum landraces in northern Cameroon. Plant Genetic Resources 11:36–49. doi:10.1017/S1479262112000342

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

Sultan, B., D. Defrance, and T. Iizumi. 2019. Evidence of crop production losses in West Africa due to historical global warming in two crop models. Sci Rep 9:1–15. doi:10.1038/s41598-019-49167-0

Teshome, A., L. Fahrig, J.K. Torrance, J.D. Lambert, T.J. Arnason, and B.R. Baum. 1999. Maintenance of sorghum (sorghum bicolor, poaceae) landrace diversity by farmers' selection in Ethiopia. Econ Bot 53:79–88. doi:10.1007/BF02860796

Thomson, M.J. 2014. High-Throughput SNP Genotyping to Accelerate Crop Improvement. Plant Breeding and Biotechnology 2:195–212. doi:10.9787/PBB.2014.2.3.195

Tuberosa, R., and S. Salvi. 2006. Genomics-based approaches to improve drought tolerance of crops. Trends in Plant Science 11:405–412. doi:10.1016/j.tplants.2006.06.003

Tuinstra, M.R., E.M. Grote, P.B. Goldsbrough, and G. Ejeta. 1997. Genetic analysis of post-flowering drought tolerance and components of grain development in Sorghum bicolor (L.) Moench. Molecular Breeding 3:439–448. doi:10.1023/A:1009673126345

Upadhyaya, H.D., K. Narsimha Reddy, M. Vetriventhan, M. Irshad Ahmed, G. Murali Krishna, M. Thimma Reddy, and S.K. Singh. 2017. Sorghum germplasm from West and Central Africa maintained in the ICRISAT genebank: Status, gaps, and diversity. The Crop Journal 5:518–532. doi:10.1016/j.cj.2017.07.002

Vadez, V., S.P. Deshpande, J. Kholova, G.L. Hammer, A.K. Borrell, H.S. Talwar, and C.T. Hash. 2011. Stay-green QTLs effects on water extraction, transpiration efficiency and seed yield depend on recipient parent background. Functional Plant Biology 38:553–566

Varshney, R.K., A. Graner, and M.E. Sorrells. 2005. Genomics-assisted breeding for crop improvement. Trends in Plant Science 10:621–630. doi:10.1016/j.tplants.2005.10.004

Vigouroux, Y., C. Mariac, S.D. Mita, J.-L. Pham, B. Gérard, I. Kapran, F. Sagnard, M. Deu, J. Chantereau, A. Ali, J. Ndjeunga, V. Luong, A.-C. Thuillet, A.-A. Saïdou, and G. Bezançon. 2011. Selection for Earlier Flowering Crop Associated with Climatic Variations in the Sahel. PLOS ONE 6:e19563. doi:10.1371/journal.pone.0019563

Visendi, P., J. Batley, and D. Edwards. 2013. Next Generation Characterisation of Cereal Genomes for Marker Discovery. Biology (Basel) 2:1357–1377. doi:10.3390/biology2041357

Walker, T.S., and J. Alwang. 2015. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. CABI.

Westwood, J.H. 2013. The Physiology of the Established Parasite–Host Association. Springer, Berlin, Heidelberg.

Yoder, J.I., and J.D. Scholes. 2010. Host plant resistance to parasitic weeds; recent progress and bottlenecks. Current Opinion in Plant Biology 13:478–484. doi:10.1016/j.pbi.2010.04.011

Yohe, J., K. Christiansen, and J. Frederick. 2005. INTSORMIL 2005 ANNUAL REPORT. INTSORMIL Impacts and Bulletins

Yoshida Satoko, and Shirasu Ken. 2009. Multiple layers of incompatibility to the parasitic witchweed, Striga hermonthica. New Phytologist 183:180–189. doi:10.1111/j.1469-8137.2009.02840.x

Zonneveld, M. van, I. Dawson, E. Thomas, X. Scheldeman, J. van Etten, J. Loo, and J.I. Hormaza. 2014. Application of Molecular Markers in Spatial Analysis to Optimize In Situ Conservation of Plant Genetic Resources. R. Tuberosa, A. Graner, and E. Frison, ed. Springer Netherlands.

Chapter 2 - Population genomics of sorghum (*Sorghum bicolor*) across diverse agroclimatic zones of Niger

This chapter is published in Genome (Maina et al., 2018).

Abstract

Improving adaptation of staple crops in developing countries is important to ensure food security. In the West African country of Niger, the staple crop sorghum (*Sorghum bicolor*) is cultivated across diverse agroclimatic zones, but the genetic basis of local adaptation has not been described. The objectives of this study were to characterize the genomic diversity of sorghum from Niger and to identify genomic regions conferring local adaptation to agroclimatic zones and farmer preferences. We analyzed 516 Nigerien accessions for which local variety name, botanical race, and geographic origin were known. We discovered 144,299 single nucleotide polymorphisms (SNPs) using genotyping-by-sequencing (GBS). We performed discriminant analysis of principal components (DAPC), which identified six genetic groups, and performed a genome scan for loci with high discriminant loadings. The highest discriminant coefficients were on chromosome 9, near the putative ortholog of maize flowering time adaptation gene *Vgt1*. Next, we characterized differentiation among local varieties and used a genome scan of pairwise F_{ST} values to identify SNPs associated with specific local varieties. Comparison of varieties named for light- versus dark-grain identified differentiation near *Tannin1*, the major gene responsible for grain tannins. These findings could facilitate genomics-assisted breeding of locally-adapted and farmer-preferred sorghum varieties for Niger.

Keywords: Sorghum, Agroclimatic zones, Genomics, Local adaptation, Genomics-enabled breeding, Sub Saharan Africa.

Introduction

Sorghum (*Sorghum bicolor* L. Moench) is a major staple crop that is adapted to multiple agroclimatic zones of the world (Smith and Frederiksen, 2000). In smallholder production systems of Africa, sorghum is used as food, forage, and building material, and cultivated under constraints such as drought, insects, weeds, and bird predation (National Research Council, 1996; Smith and Frederiksen, 2000). In the West African country of Niger, sorghum is cultivated across agroclimatic zones defined by the precipitation gradient (Saharan, Sahelian, Sahelo-soudanian, and Soudanian) (Mohamed et al., 2002; Deu et al., 2010). Four botanical types (durra, caudatum, guinea, and bicolor), classified based on the morphology of the inflorescence and grain, are found distributed across the agroclimatic zones of Niger (Deu et al., 2010). Durra, caudatum, and their intermediates are grown in the Sahelian and Sahelo-soudanian zones (400 mm to 600 mm) while guinea is predominantly grown in the Soudanian zone (800 mm). Sorghum is grown by smallholder farmers of several ethnic groups including Hausa (predominantly in central Niger), Zarma (predominantly in western Niger), and Kanuri (predominantly in eastern Niger).

The sorghum breeding program at the National Institute of Agronomic Research of Niger (INRAN) has focused on improving local varieties by introgressing new traits conferring resistance to biotic (*Striga*) and abiotic (drought, soil fertility) stressors (Ejeta and Gressel, 2007). However, the level of farmer's adoption is limited (5–10%) due to the lack of varieties that combine improved yield and tolerance traits with locally-preferred traits related to maturity and grain pigmentation (Bezançon et al., 2009; Walker and Alwang, 2015). Marker-assisted selection (MAS) can accelerate the development of new varieties with improved yield and adaptation (Collard and Mackill, 2008). Genotyping-by-sequencing (GBS) can accelerate the

development of high-density markers for genetic diversity studies and MAS (Elshire et al., 2011). In sorghum, global and regional diversity have been studied using high-density GBS markers (Morris et al., 2013a; Leiser et al., 2014; Lasky et al., 2015). However, most studies of local diversity in African sorghum were performed using low-density markers (Barro-Kondombo et al., 2010; Deu et al., 2010; Ng'uni et al., 2011). Strong population structure has been observed across agroclimatic zones in Niger (Deu et al., 2010). High-density GBS markers could help identify loci underlying local adaptation and farmer preference and facilitate development of genomic tools for MAS.

Genome scans of diverse crop germplasm can identify loci under natural and human selection (Lasky et al., 2015; Romero Navarro et al., 2017). In Nigerien sorghum, we hypothesize that adaptation to agroclimatic zones has acted on loci controlling maturity and grain pigmentation. To characterize population structure and identify genomic regions associated with local adaptation and farmer preferences, we characterized 516 accessions of Nigerien sorghum germplasm at 144,299 SNPs. We performed the genome scans to identify genome regions associated with local adaptation and farmer preferences that can be used in breeding for climate-resilient, locally-preferred varieties.

Methods

Genetic materials

The collection is composed of 520 Nigerien accessions which were collected in 1976 (A. Borgel and J. Sequier, 1977). This collection was obtained from the United States National Plant Germplasm System (NPGS) along with passport data for each accession including local variety name, botanical type, latitude, and longitude coordinates. Precipitation maps were created using average annual precipitation from 1960–1990 obtained from WorldClim 1.4 using *raster*

package in R (Hijmans et al., 2005; R Core Team, 2013; Hijmans, 2016) (Figure 2.1). Two weeks after planting in the greenhouse, 516 accessions successfully germinated (File S1). Fresh leaf tissue (50 mg) was collected for each accession and transferred into a 96-deep well plate. Each plate contains 95 samples, with an empty well as the control. Tissue was lyophilized (Labconco Freeze Dryer) for two days and then grounded using 96-well plate plant tissue grinder (Retsch Mixer Mill). Genomic DNA was extracted using the BioSprint 96 DNA Plant Kit (QIAGEN), quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit, and normalized to 10ng/ul.

Genotyping-by-sequencing (GBS)

GBS method (Elshire et al., 2011) was used for this study. DNA libraries were first digested using *Ape*KI restriction enzyme (NEB R0643L) and ligated with the barcode and common adapters using T4 DNA ligase (NEB M0202L). Ligated libraries were pooled (96-plex libraries) before PCR amplification. Libraries were purified with QIAquick PCR purification kit and a Bioanalyzer (Agilent Technologies 2100) was used to determine the library size distribution. The 384-plex library was obtained after pooling four 96-plex libraries. Sequencing was performed using single end 100-cycle sequencing with Illumina HiSeq2500 (University of Kansas Medical Center). TASSEL3 GBS pipeline (Glaubitz et al., 2014) was used to call SNP genotypes. The tags were aligned with the sorghum reference genome version v3.0 (www.phytozome.jgi.doe.gov) using Burrows-Wheeler Aligner (Li and Durbin, 2010), and SNPs were filtered at minor allele frequency (MAF) < 0.01. Missing genotypes were imputed with Beagle (Browning and Browning, 2009) using default settings. For comparison with the Nigerien germplasm, published raw GBS sequence data for the global Sorghum Association Panel (SAP)

(n = 356) were obtained (Morris et al., 2013a). We then combined both raw sequences from fastq format and re-called the SNPs.

Genomic analysis

To determine genetic relatedness among accessions, neighbor-joining analysis was performed using *ape* package in R (Paradis et al., 2004; R Core Team, 2013). Genetic groups were inferred using the discriminant analysis of principal component (DAPC) package *Adegenet* in R (Jombart and Ahmed, 2011; R Core Team, 2013) using the Bayesian Information Criterion (BIC), criterion="goodfit", n.iter=1e5, and n.start=1000. To characterize SNP variation, we used VCFtools (Danecek et al., 2011). We estimated minor allele frequency, nucleotide diversity, Tajima's *D*, and *F*-statistics (Weir and Cockerham, 1984) between botanical types, and local varieties.

For comparison with genome-wide scans, we develop a list of *a priori* candidates genes for some traits (maturity, plant height, grain tannins, and stay-green) that have been implicated in agroclimatic adaptation in sorghum and other cereals (File S2) (Wu et al., 2012; Borrell et al., 2014; Lasky et al., 2015; Romero Navarro et al., 2017). Among the candidate genes, we refer to those identified based on synteny as orthologs (Schnable and Freeling, 2011) while we refer to those based on high pairwise similarity in Phytozome as putative orthologs (www.phytozome.jgi.doe.gov). To identify SNP markers associated with environmental variables, we performed genome-wide association studies (GWAS) using the GAPIT regular mixed linear model (MLM) method (Lipka et al., 2012). Annual precipitation and annual mean temperature for each accession were obtained from WorldClim 1.4 using *raster* package in R (Hijmans et al., 2005; R Core Team, 2013).

Results

A genome-wide map of SNP variation for Nigerien sorghum

We discovered 158,019 SNPs in the Nigerien sorghum germplasm. After removing 13,720 monomorphic SNPs (8.7%), we retained 144,299 SNPs. The average Tajima's *D* across all the germplasm is 0.36 and the nucleotide diversity (π) is 0.00010 (averaged over 1 kb windows) after removing the singletons. By contrast, the average Tajima's *D* of the Nigerien diversity alone is 0.51, and the nucleotide diversity (π) is 0.00046. To better understand the genetic structure of Nigerien germplasm relative to global sorghum diversity, diversity analysis, and neighbor-joining analysis were conducted with 473,279 SNPs for the Nigerien germplasm and global diversity panel (SAP) (Casa et al., 2008; Morris et al., 2013a).

Nigerien germplasm forms several clusters relative to global diversity (Figure 2.2A). The SAP accessions El Mota and Mota Maradi (local varieties from Niger in SAP) grouped with the Nigerien landraces germplasm. By contrast Sepon82, SRN39, and MR732 (non-Nigerien varieties/lines released in Niger) grouped with east African accessions. Figure 2.2B shows the same neighbor-joining tree color-coded by botanical type. The caudatum type was composed of two clusters with one cluster mainly consisting of SAP accessions and the other with Nigerien germplasm. Neighbor-joining analysis of Nigerien germplasm without the SAP shows two durra-caudatum clusters, two caudatum clusters, and one guinea cluster (Figure A.1).

Population structure

To characterize the genetic structure, we performed DAPC which identified six genetic groups (I–VI) (Figure 2.3A and 3B). The genetic groups explained 99% of the SNP variation in five principal components (PCs) (Figure A.2). The first and the second components, which capture 44% and 18% of variation respectively, separate groups I–V from group VI (Figure

2.3A). The third, fourth, and fifth components explained 15%, 14%, and 9% of the total variation, respectively, and separate groups I–V from each other (Figure 2.3B).

We hypothesized that genetic groups reflect population structure across the agroclimatic zones. We tested the relationship between the genetic groups and annual precipitation using analysis of variance (ANOVA) and found a significantly higher annual precipitation for group I and VI versus groups II-V (P-value < 0.05) (Figure 2.3C). Consistent with this finding, we observed strong geographic clustering when the accessions were mapped by group (Figure 2.3D). Genome scan of loadings on the first principal component identifies multiple genome regions that are highly differentiated (top 0.1% of SNPs) on each chromosome (Figure 2.3E; File S3). Some of these loading peaks colocalize (within 200 kb) with *a priori* candidate genes for maturity (*PhyB*, *GI*, *abph1,* and *Vgt1*) (Zhang and Yuan, 2014).

To further characterize the genetic groups, we summarized kernel color, botanical type, and other traits for each group (Table 2.1). Group I (95 accessions) is mostly represented by guinea-caudatum and caudatum-bicolor types. The majority of the accessions in this group have a white kernel color with a predominance of Farfara (Hausa name), Amo Koire (Zarma name), and Gaberi Boul (Kanuri name), whose names indicate the white kernel color. Group II (74 accessions) is predominantly caudatum-bicolor accessions. The main kernel color in this group is brown, and Mota is the most frequent local variety. Group III (87 accessions) is predominantly durra-caudatum accessions with a mix of brown and white kernel color. Most frequent local variety in group III is Amo Koire (27 accessions of 87 accessions). Group IV (158 accessions) is predominantly durra-caudatum accessions. Babadia (25 accessions) (Hausa name) and Mota (14 accessions) are the frequent local varieties in group IV. Group V (77 accessions) is mostly represented by caudatum accessions from central Niger with a mix of brown and white kernel

color. Jenjari (Hausa name) is the most common in group V. Group VI (25 accessions), which is the most differentiated from other groups, has most of the accessions from southwestern Niger (21 of 25). Sokombe (Zarma name, guinea botanical type) is the major local variety found in group IV.

Pairwise F_{ST} estimates among botanical types and local varieties

To better understand the genetic differentiation among botanical types, we performed F_{ST} genome scan for each botanical type and evaluated the Tajima's *D* based on 100 kb window size for the botanical types (Figure 2.4). We considered the six most common botanical types and intermediates (Figure 2.4A). Durra, caudatum, and durra-caudatum accessions show low F_{ST} values across the genome with some distinct F_{ST} peaks. In contrast, guinea, guinea-caudatum, and bicolor have high F_{ST} throughout the genome (Figure 2.4B). The overall Tajima's *D* averaged over the 100 kb windows is 0.83.

Next, we cataloged the eleven most common local varieties (with >10 accessions) (Figure 2.5A) and characterized genetic differentiation (pairwise F_{ST}) among these varieties (Figure 2.5B). Sokombe, a guinea variety, is highly differentiated from all others. In contrast, the caudatum variety Mota is less differentiated from others, including Babadia, Bazanga, Bogoba, Fara Dawa, and Lalla. Based on these results, we investigated specific loci that may be involved in this differentiation, focusing on the two most common local varieties, Mota and Jenjari. Accessions of Mota and Jenjari both originate from central Niger, where annual precipitation is 400–600 mm (Figure 2.5C).

We identified regions with high F_{ST} values (top 0.1% of SNPs) highlighting the differentiation between Mota versus all other local varieties, Jenjari versus all other local varieties, and Mota versus Jenjari (Figure 2.5D). We plotted the Tajima's *D* (100 kb windows)

for the local varieties and identified genomic regions with high and low values (Figure 2.5D). The F_{ST} estimates between Mota and other frequently observed local varieties are relatively less differentiated, but specific loci were found above the threshold with less differentiation across the genome. In contrast, Jenjari appears highly differentiated with a threshold F_{ST} value of 0.75. Mota and Jenjari have substantial differentiation across the genome, including near the major grain pigmentation gene *Tannin1* (Wu et al., 2012) with Tajima's *D* of 1.9 near the genomic region (Figure 2.5E).

Genome-wide association studies for environmental variables

To further identify the genomic regions underlying local climate adaptation in the Nigerien germplasm, we performed a genome-wide association scan for annual mean temperature and annual precipitation (Figure 2.6). For annual mean temperature we found significant associations at two loci on chromosome 1 (S1_7769638, S1_7769764, S1_59520997, S1_59654095, S1_59654013) and one locus on chromosome 3 (S3_67229110) (Figure 2.6A). For annual precipitation, there were several association peaks, but none exceeded the Bonferroni significance threshold (Figure 2.6B).

Discussion

The structure of genomic diversity in Nigerien sorghum

The comparison of the Nigerien collection to the global sorghum association panel SAP (Figure 2.2A) helps to characterize the diversity of released varieties in Niger relative to local landraces and global sorghum diversity. As expected, the landrace-derived Nigerien varieties in the SAP (El Mota and Mota Maradi) clustered with Nigerien caudatum landraces. MR732 (parent of the NAD1 hybrid), Sepon82, and SRN39 clustered with other germplasm from ICRISAT, along with East African caudatum accessions, which were major contributors to the

ICRISAT breeding program (Vaidya et al., 1988; Bantilan et al., 2004). Together, these results demonstrate that the released sorghum varieties in Niger reflect much, but not all, of the genetic diversity of landraces in Niger. Future studies that include other released varieties (e.g. SSD-35, MDK, and S39) (Ministère de l'agriculture du Niger, 2012) and breeding lines will be needed to evaluate the existing diversity in the breeding program and prioritize efforts to increase genetic diversity (Jordan et al., 2011).

The neighbor-joining tree for the Nigerien germplasm (Figure A.1) shows clusters separated by botanical types and geography. This is consistent with previous studies that have shown that sorghum populations in Niger, and in Africa more broadly, are structured according to farmer preferences and agroclimatic zones (Barnaud et al., 2007; Deu et al., 2010; Labeyrie et al., 2014). The DAPC analysis (Figure 2.3) separates guinea (from the southwestern part of Niger, which receives the greatest annual precipitation) from the other groups. Surprisingly, we also observed two distinct groups of caudatum (groups I and V) and two distinct groups of durra-caudatum (groups III and IV). Both caudatum groups are found in central Niger, in apparent sympatry, yet are genetically distinct. The genetic differentiation may be due to the lack of gene flow between the groups, where group V, mostly composed of Jenjari variety (Figure 2.5A, Table 2.1), is cultivated in distant fields compared to group I due to its tannin content which prevents bird predation in unattended fields (Doggett, 1988). In addition, group V accessions (the dark-kernel caudatums) are more commonly found in high rainfall areas than group I accessions (light-kernel caudatums), consistent with natural selection favoring tannins in humid areas (Lasky et al., 2015).

Similarly, ethnic groups and geographical origin may influence population structure (Labeyrie et al., 2014; Westengen et al., 2014). The two durra-caudatum groups are

geographically distant and have local varieties of Hausa and Zarma names. Group III is predominantly found in the western part of the country, with Amo Koire as local variety whereas Group IV is found in the central part with Babadia as local variety. In contrast, some local varieties are found in multiple groups suggesting that the population structure is affected by other factors such as seed exchange and food preferences (Deu et al., 2008, 2010).

Genomic regions underlying local adaptation and farmer preferences

To identify genomic regions associated with local adaptation and farmer preferences, we performed a variety of genome scans. The SNP S9_2371665 (C/G) on chromosome 9 has the highest DAPC loading coefficient among all the SNPs (Figure 2.3E). The minor allele frequency is 0.38 for the G allele compared to the C allele (0.62). Interestingly, this SNP is located ~200 kb from the putative ortholog of the maize *Vgt1* gene (Ducrocq et al., 2008), a flowering time gene associated with maize adaptation. Another SNP, S6_51192692 (MAF = 40%) among the top 0.1% on chromosome 6, was found close to the Sobic.006G151800 (~61 kb), the ortholog of the maize *abph1* gene for floral regulation (Jackson and Hake, 1999; Giulini et al., 2004). This suggests that *Vgt1* and *abph1* are promising candidates for future studies of late-maturity adaptation in guinea sorghum of the Soudanian zone (Deu et al., 2008).

Early maturity is often favored in the Sahelian due to the short rainy season. For instance, Mota (which means “car” in multiple languages in Niger) is early maturing relative to other local landraces (Ejeta and Gressel, 2007). The SNP on chromosome 9 (S9_57517028, F_{ST} = 0.43, MAF = 13%), identified from the F_{ST} genome scan between Mota and other local varieties (Figure 2.5D), is colocalized with *SbFL9.1*, a major QTL underlying flowering time in global sorghum germplasm (Thurber et al., 2013; Bouchet et al., 2017).

Kernel color is another main characteristic that guides smallholder farmer's selection (Mekbib, 2007). For instance, white grain sorghums have local varieties in different languages of Fara Dawa (Hausa), Farfara (Hausa), Gaberi Boul (Kanuri), and Amo Koire (Zarma) whereas red grain sorghums have local varieties of Jenjari (Hausa), Gaberi Kime (Kanuri), and Ja Dawa (Hausa) names. Kernel color classification from the gene bank demonstrates (Figure 2.5A) that varieties named for red grain (Jenjari, for example) are predominantly brown or red kernel (i.e. dark-grained). The major gene responsible for grain tannin content variation is *Tannin1 (*Sobic.004G280800*)* on chromosome 4 (Wu et al., 2012; Morris et al., 2013b; Rhodes et al., 2014). For the genome scan between sympatric red/brown grain (Jenjari) and Mota varieties (Figure 2.5D), a high F_{ST} SNP on chromosome 4 (S4_62359202, F_{ST} = 0.47, MAF = 34%) is close to *Tannin1* gene (~43 kb). Similarly, the F_{ST} genome scan between Jenjari and Fara Dawa (white-grained) showed highly-differentiated SNPs near the *Tannin1* gene *S4_62389180* (F_{ST} = 0.47, MAF=13%), ~73 kb from *Tannin1*) (Figure A.3). Together, this suggests a history of balancing selection on *Tannin1* in Jenjari due to the farmer practice of planting dark-grained Jenjari in distant fields, where high tannin reduces bird predation. In this case, there is feedback of natural selection (from bird predation) and human selection (on grain color) acting on *Tannin1*.

To identify loci that may underlie adaptation to temperature and precipitation we also performed GWAS on these environmental variables. We do not find any genes in our *a priori* candidate list that are very near (< 150 kb) to the significant SNPs. For annual mean temperature, the peak around 7.7 Mb on chromosome 1 (S1_7769638 and S1_7769764) is ~400 kb from an *a priori* candidate gene, the sorghum ortholog (Sobic.001G140200) of maize floral regulator *knotted1* (*KN1*) (Zhang and Yuan, 2014) (Figure 2.6A). While there are no significant

associations for annual precipitation, this is not surprising since there is low power to detect climate association with mixed models when population structure is confounded with climate variables (Lasky et al., 2015). Further phenotyping and trait mapping studies will be needed to confirm if some loci with signatures of selection control flowering time and kernel color in Nigerien germplasm.

Implications for genomics-enabled breeding for smallholder farmers

Genomic regions that are highly differentiated among local varieties and botanical types may be useful in marker-assisted selection for local adaptation. For instance, when a local variety is crossed with a non-adapted trait-donor line, markers could be used to screen progeny lines carrying a larger number of alleles for local adaptation and/or farmer preference. Local varieties have been used as recurrent parents in INRAN breeding program (Kapran et al., 2007). For example, Mota Maradi, a selection of Mota developed by INRAN, is the second most widely cultivated improved variety in Niger (3.6% of production area) (Walker and Alwang, 2015). The F_{ST} estimates between Mota and other local varieties (Figure 2.5D) reveal regions on the genome that are differentiated on chromosomes 3, 5, 7, and 9. When improving Mota, those regions can be selected in order to keep their preferred traits while introgressing new alleles from the donor line.

For some botanical types it is more challenging to prioritize genomic regions for selection. For instance, the F_{ST} for guinea and guinea-caudatum shows high differentiation across almost the entire genome (Figure 2.4B), consistent with the population differentiation of guinea sorghums relative to all other botanical types (Figures 2.2B, 2.3A). Thus, it might be easier to select markers across the genome with durra, caudatum and their intermediates where few genome regions are differentiated (Figure 2.4B). Although the F_{ST} genome scan for guinea types

did not reveal specific loci under selection, the DAPC loadings (Figure 2.3E), Tajima's *D* (Figure 2.4), or environmental association scans (Figure 2.6) may identify loci that control guinea traits, which could be useful for improving sorghum in humid regions. Follow-up studies that combine population analyses with quantitative trait analyses should provide a more complete understanding of sorghum adaptation in Niger and facilitate breeding of locally-adapted and farmer-preferred varieties.

Data availability

Raw sequencing data are available in the NCBI Sequence Read Archive under project accession SRP132525. Imputed hapmap file available in Dryad Digital Repository (doi:10.5061/dryad.5n2bs6r). Please contact the corresponding author for availability.

Acknowledgments

This paper is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A-13-00047. The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. We thank two anonymous reviewers for suggestions that improved the manuscript.

References

A. Borgel, and J. Sequier. 1977. Prospection des mils pénicillaires et sorghos en Afrique de l'Ouest. Campagne 1976 - Niger

Bantilan, M., U. Deb, C. Gowda, B. Reddy, A. Obilana, and R. Evenson eds. 2004. Sorghum Genetic Enhancement: Research Process, Dissemination and Impacts. International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502324, Andhra Pradesh, India.

Barnaud, A., M. Deu, E. Garine, D. McKey, and H.I. Joly. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor Appl Genet 114:237–248. doi:10.1007/s00122-006-0426-8

Barro-Kondombo, C., F. Sagnard, J. Chantereau, M. Deu, K. vom Brocke, P. Durand, E. Gozé, and J.D. Zongo. 2010. Genetic structure among sorghum landraces as revealed by morphological variation and microsatellite markers in three agroclimatic regions of Burkina Faso. Theor Appl Genet 120:1511–1523. doi:10.1007/s00122-010-1272-2

Bezançon, G., J.-L. Pham, M. Deu, Y. Vigouroux, F. Sagnard, C. Mariac, I. Kapran, A. Mamadou, B. Gérard, J. Ndjeunga, and J. Chantereau. 2009. Changes in the diversity and geographic distribution of cultivated millet (Pennisetum glaucum (L.) R. Br.) and sorghum (Sorghum bicolor (L.) Moench) varieties in Niger between 1976 and 2003. Genetic Resources and Crop Evolution 56:223–236. doi:10.1007/s10722-008-9357-3

Borrell, A.K., J.E. Mullet, B. George-Jaeggli, E.J. van Oosterom, G.L. Hammer, P.E. Klein, and D.R. Jordan. 2014. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. J. Exp. Bot. eru232. doi:10.1093/jxb/eru232

Bouchet, S., M.O. Olatoye, S.R. Marla, R. Perumal, T. Tesso, J. Yu, M. Tuinstra, and G.P. Morris. 2017. Increased Power To Dissect Adaptive Traits in Global Sorghum Diversity

Using a Nested Association Mapping Population. Genetics 206:573–585. doi:10.1534/genetics.116.198499

Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am J Hum Genet 84:210–223. doi:10.1016/j.ajhg.2009.01.005

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community Resources and Strategies for Association Mapping in Sorghum. Crop Science 48:30–40. doi:10.2135/cropsci2007.02.0080

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. doi:10.1098/rstb.2007.2170

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The Variant Call Format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Deu, M., F. Sagnard, J. Chantereau, C. Calatayud, D. Hérault, C. Mariac, J.-L. Pham, Y. Vigouroux, I. Kapran, P.S. Traore, A. Mamadou, B. Gerard, J. Ndjeunga, and G. Bezançon. 2008. Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. Theor. Appl. Genet. 116:903–913. doi:10.1007/s00122-008-0721-7

Deu, M., F. Sagnard, J. Chantereau, C. Calatayud, Y. Vigouroux, J.L. Pham, C. Mariac, I. Kapran, A. Mamadou, B. Gérard, J. Ndjeunga, and G. Bezançon. 2010. Spatio-temporal dynamics of genetic diversity in Sorghum bicolor in Niger. Theor. Appl. Genet. 120:1301–1313. doi:10.1007/s00122-009-1257-1

Doggett, H. 1988. Sorghum. 2nd Edition.

Ducrocq, S., D. Madur, J.-B. Veyrieras, L. Camus-Kulandaivelu, M. Kloiber-Maitz, T. Presterl, M. Ouzunova, D. Manicacci, and A. Charcosset. 2008. Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information. Genetics 178:2433–2437. doi:10.1534/genetics.107.084830

Ejeta, G., and J. Gressel. 2007. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt. World Scientific Publishing Co Pte Ltd, Singapore, SINGAPORE.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6:e19379. doi:10.1371/journal.pone.0019379

Giulini, A., J. Wang, and D. Jackson. 2004. Control of phyllotaxy by the cytokinin-inducible response regulator homologue ABPHYL1. Nature 430:1031–1034. doi:10.1038/nature02778

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One 9. doi:10.1371/journal.pone.0090346

Hijmans, R.J. 2016. Raster: Geographic Analysis and Modeling with Raster Data

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965–1978. doi:10.1002/joc.1276

Jackson, D., and S. Hake. 1999. Control of phyllotaxy in maize by the abphyl1 gene. Development 126:315–323

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071. doi:10.1093/bioinformatics/btr521

Jordan, D.R., E.S. Mace, A.W. Cruickshank, C.H. Hunt, and R.G. Henzell. 2011. Exploring and Exploiting Genetic Variation from Unadapted Sorghum Germplasm in a Breeding Program. Crop Science 51:1444–1457. doi:10.2135/cropsci2010.06.0326

Kapran, I., C. Grenier, and G. Ejeta. 2007. Introgression of genes for Striga resistance into African landraces of sorghum. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt 129–141

Labeyrie, V., M. Deu, A. Barnaud, C. Calatayud, M. Buiron, P. Wambugu, S. Manel, J.-C. Glaszmann, and C. Leclerc. 2014. Influence of Ethnolinguistic Diversity on the Sorghum Genetic Patterns in Subsistence Farming Systems in Eastern Kenya. PLOS ONE 9:e92178. doi:10.1371/journal.pone.0092178

Lasky, J.R., H.D. Upadhyaya, P. Ramu, S. Deshpande, C.T. Hash, J. Bonnette, T.E. Juenger, K. Hyma, C. Acharya, S.E. Mitchell, E.S. Buckler, Z. Brenton, S. Kresovich, and G.P. Morris. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. Science Advances 1:e1400218. doi:10.1126/sciadv.1400218

Leiser, W.L., H.F.W. Rattunde, E. Weltzien, N. Cisse, M. Abdou, A. Diallo, A.O. Tourè, J.V. Magalhaes, and B.I. Haussmann. 2014. Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African Sorghum. BMC Plant Biology 14:206. doi:10.1186/s12870-014-0206-6

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. doi:10.1093/bioinformatics/btp698

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Maina, F., S. Bouchet, S.R. Marla, Z. Hu, J. Wang, A. Mamadou, M. Abdou, A.-A. Saïdou, and G.P. Morris. 2018. Population genomics of sorghum (Sorghum bicolor) across diverse agroclimatic zones of Niger. Genome 1–10. doi:10.1139/gen-2017-0131

Mekbib, F. 2007. Infra-specific folk taxonomy in sorghum (Sorghum bicolor (L.) Moench) in Ethiopia: folk nomenclature, classification, and criteria. J Ethnobiol Ethnomed 3:38. doi:10.1186/1746-4269-3-38

Ministère de l'agriculture du Niger. 2012. Catalogue national des espèces et variétes végétales du Niger

Mohamed, A.B., N. van Duivenbooden, and S. Abdoussallam. 2002. Impact of Climate Change on Agricultural Production in the Sahel – Part 1. Methodological Approach and Case Study for Millet in Niger. Climatic Change 54:327–348. doi:10.1023/A:1016189605188

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013a. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. PNAS 110:453–458. doi:10.1073/pnas.1215985110

Morris, G.P., D.H. Rhodes, Z. Brenton, P. Ramu, V.M. Thayil, S. Deshpande, C.T. Hash, C. Acharya, S.E. Mitchell, E.S. Buckler, J. Yu, and S. Kresovich. 2013b. Dissecting Genome-Wide Association Signals for Loss-of-Function Phenotypes in Sorghum Flavonoid Pigmentation Traits. G3 3:2085–2094. doi:10.1534/g3.113.008417

National Research Council. 1996. Lost Crops of Africa: Volume I: Grains

Ng'uni, D., M. Geleta, and T. Bryngelsson. 2011. Genetic diversity in sorghum (Sorghum bicolor (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). Hereditas 148:52–62. doi:10.1111/j.1601-5223.2011.02208.x

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290. doi:10.1093/bioinformatics/btg412

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rhodes, D.H., L. Hoffmann, W.L. Rooney, P. Ramu, G.P. Morris, and S. Kresovich. 2014. Genome-Wide Association Study of Grain Polyphenol Concentrations in Global Sorghum [Sorghum bicolor (L.) Moench] Germplasm. J. Agric. Food Chem. 62:10916–10927. doi:10.1021/jf503651t

Romero Navarro, J.A., M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, A. Terron, H.V. Delgado, V. Vidal, A. Ortega, A.E. Banda, N.O.G. Montiel, I. Ortiz-Monasterio, F.S. Vicente, A.G. Espinoza, G. Atlin, P. Wenzl, S. Hearne, and E.S. Buckler. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480. doi:10.1038/ng.3784

Schnable, J.C., and M. Freeling. 2011. Genes Identified by Visible Mutant Phenotypes Show Increased Bias toward One of Two Subgenomes of Maize. PLOS ONE 6:e17855. doi:10.1371/journal.pone.0017855

Smith, C.W., and R.A. Frederiksen. 2000. Sorghum, Origin, History, Technology and Production. Wiley Series in Crop Science. SB191.S7 S668 2000. Texas A & M University.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biology 14:R68. doi:10.1186/gb-2013-14-6-r68

Vaidya, P.K., B. Raghavender, and S.Z. Mukuru. 1988. . Striga research in sorghum at ICRISAT Center. Pages 1–24 Workshop on Combating Striga in Africa - Opportunitines for Research Collabortion, IITA, Ibadan, Nigeria.

Walker, T.S., and J. Alwang. 2015. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. CABI.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358–1370. doi:10.2307/2408641

Westengen, O.T., M.A. Okongo, L. Onek, T. Berg, H. Upadhyaya, S. Birkeland, S.D.K. Khalsa, K.H. Ring, N.C. Stenseth, and A.K. Brysting. 2014. Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. PNAS 111:14100–14105. doi:10.1073/pnas.1401646111

Wu, Y., X. Li, W. Xiang, C. Zhu, Z. Lin, Y. Wu, J. Li, S. Pandravada, D.D. Ridder, G. Bai, M.L. Wang, H.N. Trick, S.R. Bean, M.R. Tuinstra, T.T. Tesso, and J. Yu. 2012. Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1. PNAS 109:10281–10286. doi:10.1073/pnas.1201700109

Zhang, D., and Z. Yuan. 2014. Molecular Control of Grass Inflorescence Development. Annual Review of Plant Biology 65:553–578. doi:10.1146/annurev-arplant-050213-040104

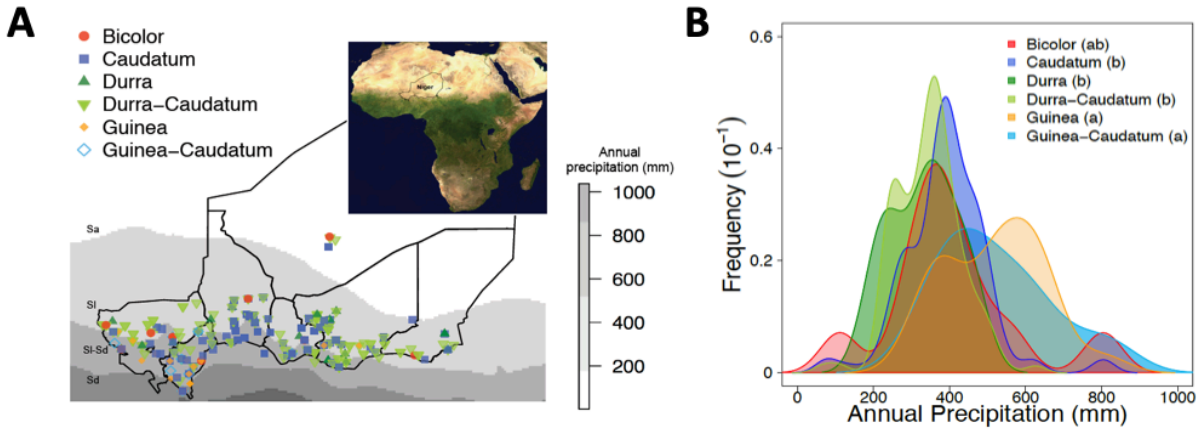


Figure 2.1. Distribution of Nigerien germplasm across precipitation gradient.

(A) Niger map with precipitation gradient colored in grey. The points represent the georeferenced accessions and each color represents the botanical type of the accession. The precipitation gradient, decreasing from south to north, reflects the agroclimatic zones. The highest precipitation is in the southwestern part of the country. Agroclimatic zones key: Sa, Saharian; Sl, Sahelian; Sl-Sd, Sahelo-Soudanian; and Sd, Soudanian. (B) Density plot of the botanical types according to annual precipitation for each accession. Different lowercase letters (a, b) represent significant differences in annual precipitation (at the site of collection) among accessions from each botanical type.

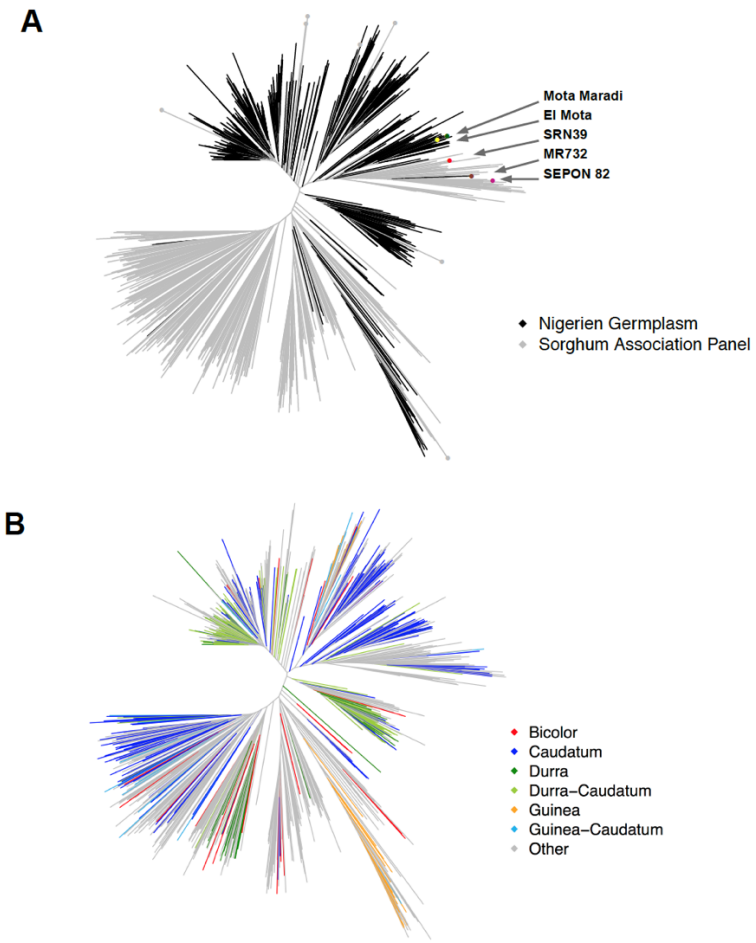


Figure 2.2. Comparison of Nigerien and global diversity (SAP) accessions.

(A) Neighbor-joining tree of 872 accessions including both Nigerien and SAP accessions. The grey lines represent the SAP (356 accessions) and the black lines represent the Nigerien germplasm (516 accessions). The grey arrows represent the Nigerien accessions in SAP: Mota Maradi (PI656050), El Mota (PI656035), SRN39 (PI656027), MR732 (PI656051) and Sepon 82 (PI656024). The grey points represent the SAP accessions (PI656066, PI597972, PI597973, PI655981, PI533882, PI656077, PI656115) that are found within Nigerien clusters. (B) The same neighbor-joining tree of 872 accessions (including both Nigerien and SAP accessions) color-coded by botanical type.

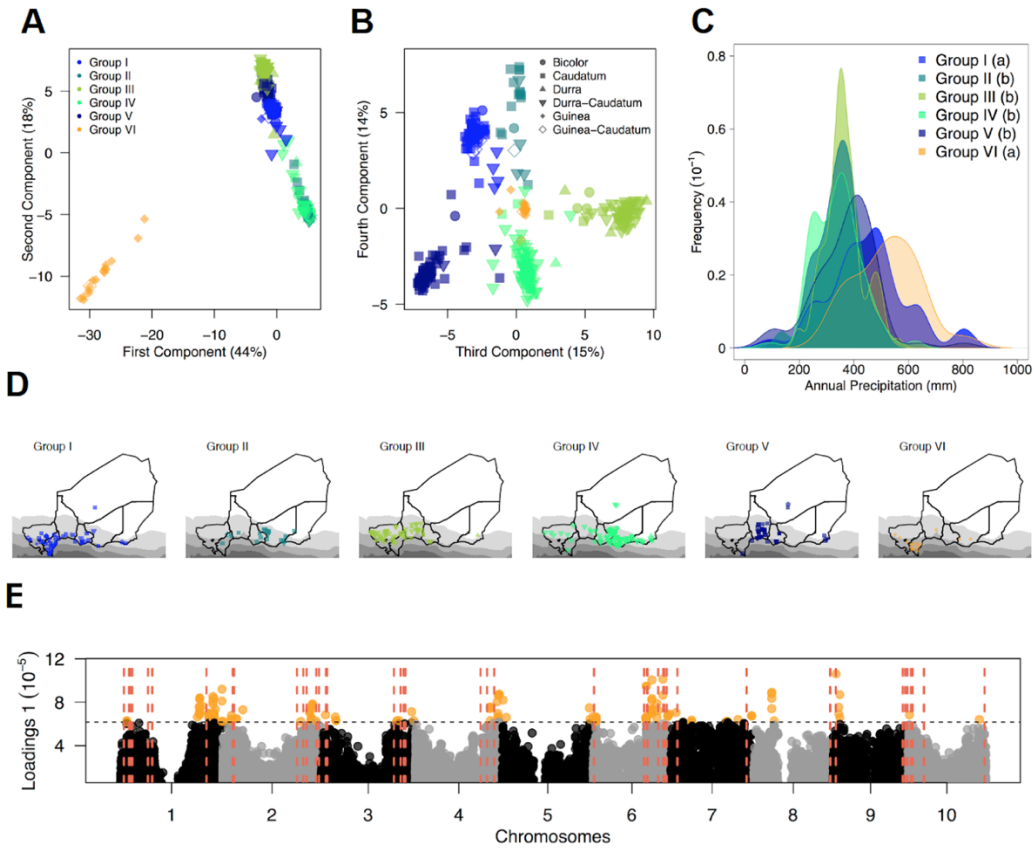


Figure 2.3. Population structure of the Nigerien germplasm using Discriminant Analysis of Principal Components.

Scatter plot of the principal components, with first and second components (A) and third and fourth components (B) that identify six genetic groups. (C) Density plot of the genetic groups according to annual precipitation for each accession. Different lowercase letters (a, b) represent significant differences among groups. (D) Maps of Niger with the collection location of accessions for each genetic group. For A-C, botanical type for each accession is noted by shape and genetic group is noted by color. (E) Genome scan of the loadings from the first principal component. The *x*-axis represents the position of each SNP on the chromosomes and the *y*-axis

the loadings values. The colored points represented highly differentiated genome regions (top 0.1% of SNPs). The dashed vertical lines represent *a priori* candidate genes.

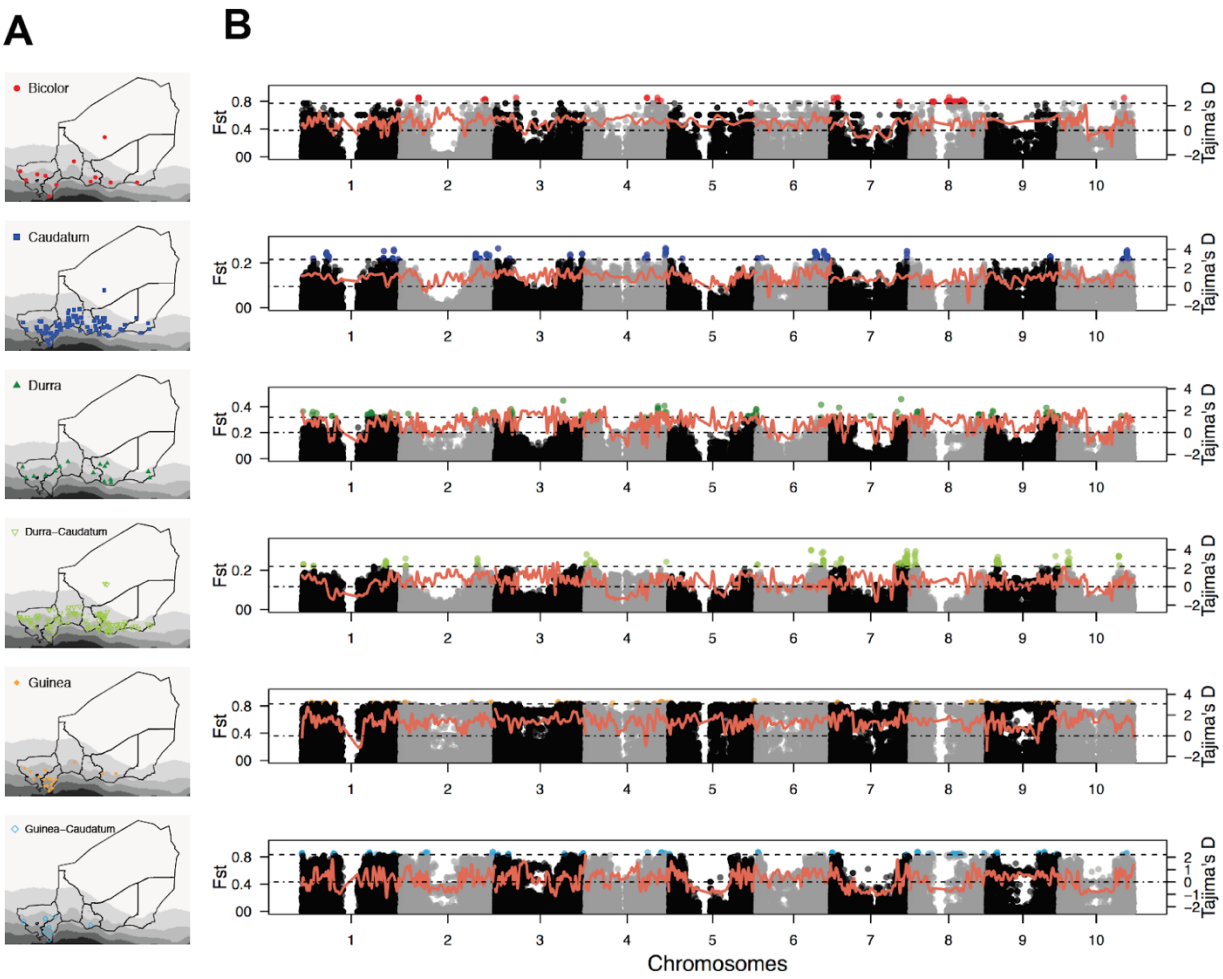
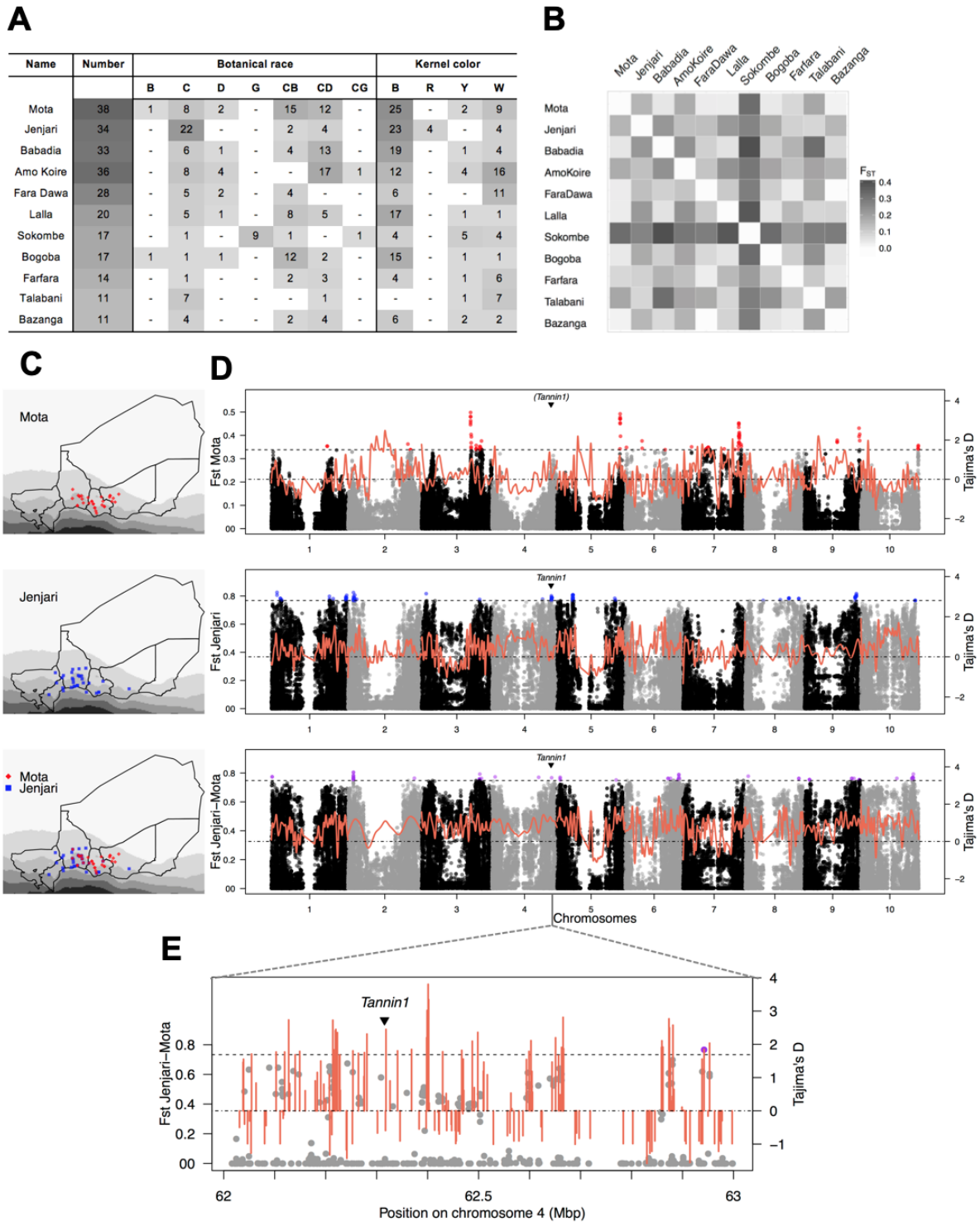


Figure 2.4. Genome scan for differentiation among botanical types.

(A) Niger map with the georeferenced accessions for each botanical type. (B) F_{ST} and Tajima's *D* genome scans for the botanical types. The *x*-axis represents the chromosomes and the first y-axis the F_{ST} values. Each point represents the SNP marker. The top 0.1% was used to set the threshold. The colors represent the SNPs above the threshold comparing one botanical type compared to the five others. F_{ST} was calculated using Weir and Cockerham's method. The second *y*-axis represents the Tajima's *D* average of 100 kb window size.



Name	Number	Botanical race							Kernel color			
		B	C	D	G	CB	CD	CG	B	R	Y	W
Mota	38	1	8	2	-	15	12	-	25	-	2	9
Jenjari	34	-	22	-	-	2	4	-	23	4	-	4
Babadia	33	-	6	1	-	4	13	-	19	-	1	4
Amo Koire	36	-	8	4	-	-	17	1	12	-	4	16
Fara Dawa	28	-	5	2	-	4	-	-	6	-	-	11
Lalla	20	-	5	1	-	8	5	-	17	-	1	1
Sokombe	17	-	1	-	9	1	-	1	4	-	5	4
Bogoba	17	1	1	1	-	12	2	-	15	-	1	1
Farfara	14	-	1	-	-	2	3	-	4	-	1	6
Talabani	11	-	7	-	-	-	1	-	-	-	1	7
Bazanga	11	-	4	-	-	2	4	-	6	-	2	2

Figure 2.5. Genomic Differentiation among the varieties with different local names.

(A) Summary table of the eleven most common local varieties with their number of accessions with local name, botanical types and kernel colors number. The color gradient shows the scale of the occurrence for each column. (B) Heatmap of genetic differentiation (F_{ST}) values among local varieties. The F_{ST} varies from less differentiated (white) to highly differentiated (blue). Sokombe (guinea type) appeared the most differentiated among the local varieties. (C) Niger map with the georeferenced accessions for Mota and Jenjari. (D) F_{ST} and Tajima's *D* genome scans between Mota between the ten local varieties, Jenjari between the ten local varieties, and Mota between Jenjari. (E) F_{ST} and Tajima's *D* genome scans for the *Tannin 1* region between and Mota between Jenjari. The *x*-axis represents the chromosomes and the first y-axis the F_{ST} values. Each point represents the SNP marker. The top 0.1% was used to set the threshold. The colors represent the SNPs above the threshold that differentiated the local varieties. F_{ST} was calculated using Weir and Cockerham's method. The second *y*-axis represents the Tajima's *D* average of 100 kb window size. Botanical type key: B, bicolor; C, caudatum; D, durra; G, guinea; C-B, caudatum-bicolor; C-D, caudatum-durra; C-G, caudatum-guinea. Kernel color key: B, brown; R, red; Y, yellow; W, white.

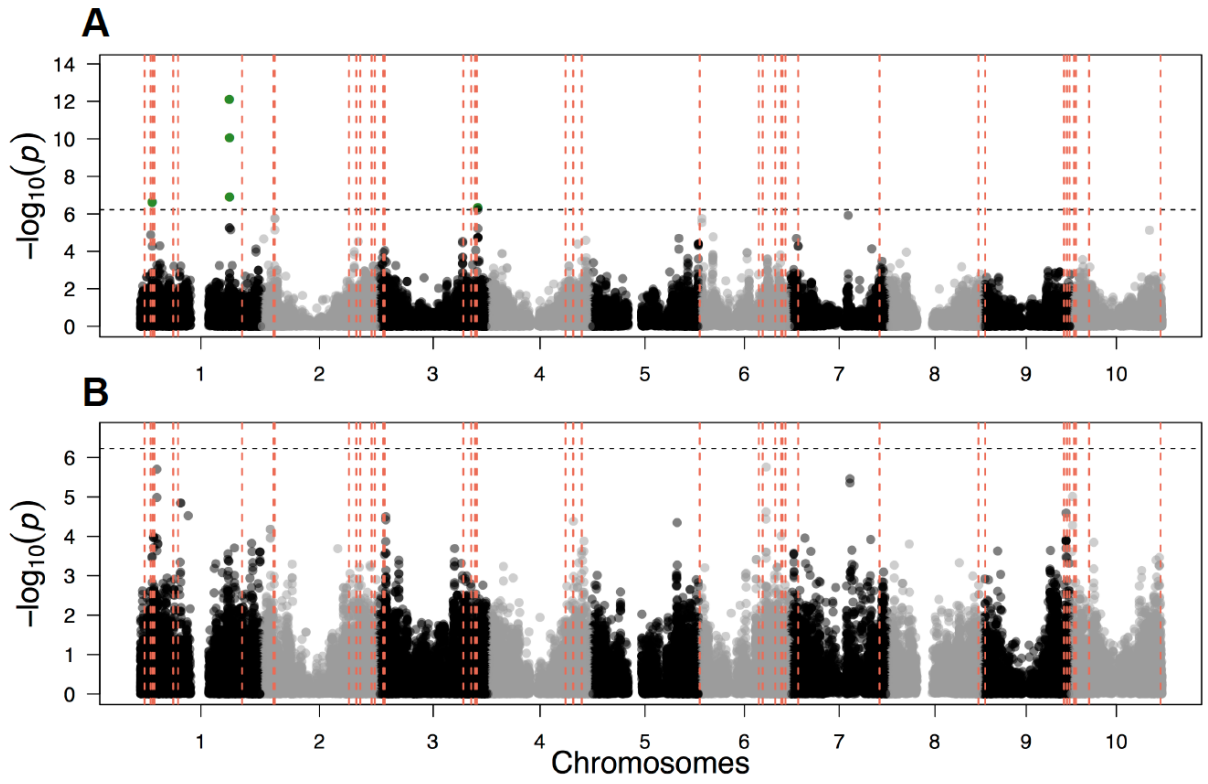


Figure 2.6. Genome-wide association studies of environmental variables.

(A) Manhattan plot for annual mean temperature. (B) Manhattan plot for annual precipitation. The *x*-axis represents the chromosomes and the *y*-axis the $-\log_{10}$(*P*-values) for the marker-trait association. Each point represents the SNP marker. The threshold is set based on the Bonferroni correction of the *P*-values. The dashed lines represent *a priori* candidate genes.

Table 2.1. Summary of the genetic groups inferred by the Discriminant Analysis of Principal Components

Group	Number	Botanical race							Kernel Color				Main Local names				
		B	C	D	G	CB	CD	CG	B	R	Y	W	M	J	B	A	S
I	95	4	3 2 1	-	7	12	4	8	1 5 6	2	1 7	4 6	2 1	2	1	7	3
II	74	1	5 1	-	-	44	4	-	2 3		2	2 2	6	1	3	- 2	-
III	87	4	2 1	9 1	-	5	37	-	6 7	2	9 1	6 4	1 1	7	- 2	7	1
IV	158	2	8 4	1		33	74	-	9 3	3	0	6 2	4	2	5	2	-
V	77	1	9	-	-	3	4	-	0	2	3 1	8	1	4	-	-	- 1
VI	25	-	-	-	6	-	-	1	5	-	2	2	-	-	-	-	3

Botanical type key: B, bicolor; C, caudatum; D, durra; G, guinea; CB, caudatum-bicolor; CD, caudatum-durra; CG, caudatum-guinea. Kernel color key: B, brown; R, red; Y, yellow; W, white. Main local names key: M, Mota; J, Jenjari; B, Babadia; A, Amo Koire; S, Sokombe.

Chapter 3 - Ecophysiological genomics of drought tolerance in West African Sorghum

Abstract

In semi-arid regions of sub-Saharan Africa, drought frequently reduces crop productivity and threatens food security. Sorghum (*Sorghum bicolor* L. Moench), a staple crop in these regions, is known to be among the most adapted cereal crops to drought, but little is known on the genetic basis of its drought tolerance. Also, sorghum varieties show differential response to drought. The study aims to characterize and identify sources of drought tolerance in a West African sorghum association panel (WASAP) and to identify quantitative trait loci (QTL) associated with yield components in an induced-drought experiment. We hypothesize that West African sorghum harbor variation for drought tolerance controlled by few genes. A subset of 214 genotypes was phenotyped in a lysimeter system under well-watered (WW) and water-stressed (WS) treatments with intermittent water deficit applied, starting at flag leaf appearance. Variations for plant height and days to flowering across countries of origin and botanical types were observed. Water stress decreased grain weight per panicle. To identify QTLs associated with the measured traits, we conducted genome-wide association studies using 90,148 single nucleotide polymorphisms. Marker-trait associations were co-localized with known and novel loci associated with the variation in yield components under WW, WS, and responses to water deficit. These results provide resources for genomics-enabled breeding for drought tolerance in semi-arid regions.

Keywords: Adaptation, lysimeter, crop, drought, sorghum, water deficit

Introduction

Semi-arid regions have experienced prolonged periods of drought that have led to severe staple crop yield losses in smallholder farming systems, while shifts in precipitation patterns were observed in recent years (Glantz, 1987; Burke et al., 2009). Future climates will increase spatial and temporal variabilities of rainfall (Biasutti, 2019), especially in the Sahel region, where severe drought occurred in the 1970s and 1980s (Barbé and Lebel, 1997). There will be increased variability of drought scenarios throughout the growing season limiting crop production in smallholder farming systems from between 10 to 50% yield reduction in the Sahel (Traoré et al., 2011; Sultan et al., 2013). Thus, there is an urgent need to develop drought-tolerant crop varieties. However, screening for drought stress on-station is often challenging in small breeding programs leading to the inability to efficiently screening under water-limited conditions or in relevant stress conditions that replicate targeted environments (Messina et al., 2011; Fang and Xiong, 2015). To facilitate efficient selection, experimental approaches are needed for screening in specific water-limited conditions.

Complex traits remain a major constraint for crop improvement due to the difficulty to replicate stress conditions in fields and when screening in multi-environments. Recent advances in phenotyping and genotyping methods have provided for a more detailed analysis of complex traits under various scenarios (Hamblin et al., 2011; Morrell et al., 2012; Varshney et al., 2014; Pieruschka and Schurr, 2019). Furthermore, those methods are being used to assist crop development efforts worldwide (Morrell et al., 2012; Cobb et al., 2013). Rapid and precise phenotyping methods that provide a better understanding of crop varieties under water deficit may require screening in controlled environments that replicate targeted population environments (Cobb et al., 2013; Vadez et al., 2013). Given the use of high-throughput

genotyping methods, diversity panels have been sequenced, providing genomic resources for understanding the genetic diversity and providing tools for molecular breeding (Elshire et al., 2011; Morrell et al., 2012; Hathorn and Chapman, 2014). Leveraging those approaches may shed light on deciphering the genetic basis of complex traits (Vadez et al., 2015; Crain et al., 2017). Genome-wide association studies have been applied in analyzing the genetic basis of drought tolerance with enhanced use of markers for accelerated breeding (Collard and Mackill, 2008; Flint-Garcia, 2013; Varshney et al., 2014). For instance, genomic regions associated with drought tolerance may be useful in breeding through genomics-assisted breeding to achieve the goal of providing improved staple crops in semi-arid regions.

Sorghum (*Sorghum bicolor*) is adapted to different environmental conditions (semi-arid and sub-humid zones), including those on marginal lands (Doggett, 1988; Smith and Frederiksen, 2000). Sorghum will be a suitable crop in the future climates where increasing water scarcity is projected. In West Africa (WA), sorghum exhibits substantial diversity related to agroclimatic zones and botanical types. As new variations might have been selected during adaptation to environmental changes in the Sahel (Orr, 1998), sorghum may have been adapted to optimize yield components in water deficit conditions. Those include adaptive mechanisms such as early flowering, delayed leaf senescence, and less reduction of grain yield (Blum, 2004; Borrell et al., 2014). These traits might be controlled by a few genes. The characterization of germplasm collections under drought stress are limited in WA, thus challenging the selection of appropriate and targeted markers for genomics-enabled breeding. To understand the genetic basis of complex traits and the mechanisms underlying stress tolerance in WA sorghum, phenotyping in contrasting environments coupled with genomic studies are needed. In this study, we evaluated a subset of the West African Sorghum Association Panel (WASAP) in a managed

environment for well-watered and water-stressed treatment. Phenotypic data were obtained in the lysimeter system allowing the control of water deficit. The objectives are to (i) characterize the genetic variation of WASAP under pre-flowering water limitation, and (ii) to identify quantitative trait loci (QTLs) associated with yield components under pre-flowering water limitation.

Material and Methods

Plant material

The West African sorghum association panel (WASAP), consisting of landraces and breeding lines, was generated from four countries (Faye et al., *in prep*). Seeds were obtained from each sorghum breeding program in Mali, Niger, Senegal, and Togo. We investigated the subset of WASAP composed of 219 genotypes (WASAP_Lysi; Table 3.1, Table B.1). Nigerien local genotypes (n = 13, Table B.1) used in previous experiments in the lysimeter system (Hamidou et al., *in prep*), mostly durra botanical type used by smallholder farmers across Niger, were included. In addition, United States breeding lines Tx7000 (pre-flowering drought-tolerant) and BTx642 (pre-flowering drought susceptible) were included (Tuinstra et al., 1996; Borrell et al., 2014).

Genomic data

Genotypic data used was generated using the genotyping-by-sequencing (GBS) method with the *ApeKI* restriction enzyme (Elshire et al., 2011). To characterize the genetic diversity of WASAP_Lysi relative to the global sorghum diversity (Sorghum Association Panel, SAP) and the West African sorghum USDA-GRIN (WAGRIN) (Figure 3.1), we analyzed their sequence reads obtained from previous studies (Morris et al., 2013; Lasky et al., 2015; Maina et al., 2018; Olatoye et al., 2018; Faye et al., 2019). The reference genome BTx623 version 3.1 (Paterson,

2013; McCormick et al., 2018) was used to align sequencing data using Burrows–Wheeler Alignment (Li and Durbin, 2010), and the SNPs were discovered using TASSEL 5 GBS pipeline (Bradbury et al., 2007). Missing SNPs were imputed in Beagle v1.4 (Browning and Browning, 2009). High-quality GBS data of 219 WASAP_Lysi genotypes, 1527 West African genotypes, and 342 SAP genotypes were further analyzed.

Genomic analysis

To characterize the WASAP_Lysi genetic diversity relative to the SAP and the WAGRIN, principal component analysis (PCA) was performed in R using the *SNPRelate* package (Zheng et al., 2012). First, the variance components of each genotype in the WASAP_Lysi using the training set from SAP with ~90,000 biallelic SNPs present in both collections (WASAP_Lysi and SAP) was predicted. Next, the variance components of WASAP_Lysi genotypes were predicted using WAGRIN as a training set with ~138,000 biallelic SNPs present in both collections (WASAP_Lysi and WAGRIN).

Lysimeter phenotyping

A lysimeter experiment was conducted at the ICRISAT Sahelian Center, Niger (Sadoré, 13.15˚N, 2.18˚E) from the month of June to the month of November for two years (2017 and 2018). The rainy season in Niger is within the months from June to October. The system, equipped with rainout shelter, consists of evaluating the physiological characteristics of crops under managed experiments in PVC tubes (Vadez et al., 2008, 2013). Tubes of 1.30 m height and 0.25 m diameter were placed in trenches (Halilou et al., 2015) and filled with soil from a farm in Sadoré station (Figure 3.2A). Each tube was fertilized with ~ 6g of 15–15–15 (N-P_2O_5-K_2O) following a recommended standard fertilization (Ministère de l'agriculture du Niger, 2012). To limit soil evaporation, the soil surface was covered with transparent plastic bags and

~350 g of polyethylene beads. Tubes were weighed weekly using a scale (Mettler Toledo) with a maximum capacity of 200 kg (20 g accuracy) to estimate water transpired by the plant after 80% field capacity.

Experimental conditions for well-watered and water-stressed treatments

The experiment was conducted in a split-unit design with two treatments, well-watered (WW) and water-stressed (WS), and with three replications for each treatment. The experimental unit was each lysimeter tube. Three seeds in each tube were planted, and two plants were left in the tube a week after emergence. Two weeks after planting, one plant was removed from the tube and used to estimate initial biomass, and one individual plant in each tube was grown until physiological maturity. The two genotypes with known drought stress response, Tx7000 and BTx642 were highly replicated in both treatments (n = 14). In the well-watered treatment (WW), genotypes were irrigated at 80% field capacity until physiological maturity. Thus, the amount of water after weighing the tubes was estimated and added to the tubes to keep soil at 80% field capacity. In the water-stressed treatment (WS), genotypes were irrigated as in the WW until flag leaf appeared for each genotype in the lysimeter tube before intermittent drought stress was applied. This corresponds to a few days before flowering time (Figure 3.2B). The water deficit treatment consisted of skipping irrigation upon the observation of wilting point on the last three leaves before relieving the water stress with one liter of water. To maintain the difference in available water for each genotype between treatments, the same quantity of water was added in WW treatment. Total water transpired was estimated in both treatments as the total water added from week four until physiological maturity of each lysimeter tube.

Trait notation

Plant height was measured from the base of the stem to the tip of the last fully emerged leaf each week. Days to flowering was recorded at the date of anther appearance on each plant while days to maturity was recorded as final physiological maturity for the entire panicle. Yield components include final vegetative biomass (vegetative biomass in g/plant), panicle length (cm, from the first node of the rachis to the tip of the panicle), and grain weight (g/panicle). Harvest index was calculated as:

$$\text{HI (\%)} = \frac{GW}{\text{PW} + \text{FVB}}$$

Where GW as grain weight per plant (g/plant); PW as panicle weight (g/plant); FVB as final vegetative biomass (g/plant).

Statistical analysis of phenotypes

Analyses were performed in the R program (R Core Team, 2013). To estimate the variance components across years within a treatment, *lme4* package was used (Bates et al., 2015) [1].

lmer(Trait= 1 + Y + (1|G) + (1|G :Y :R) +(1|G :R)+ (1|G :Y)) [1],

where G is the genotype, Y is the year, R is the replications. Year is treated as a fixed effect. Best linear unbiased predictions (BLUPS) were then estimated for genome-wide association studies.

Broad-sense heritability across years within the same treatment was estimated as:

$$H^2 = \frac{{\sigma^2}_G}{{\sigma^2}_G + \left(\frac{{\sigma^2}_{G\times Y}}{R}\right) + \left(\frac{{\sigma^2}_E}{RY}\right)},$$

Where ${\sigma^2}_G$ is the genotypic variance, ${\sigma^2}_E$ the residual variance, Y is the number of years, R is the number of repetitions, and ${\sigma^2}_{G\times Y}$ is the genotype-by-year interaction.

ANOVA was performed to assess whether the sources of variation were significant for the evaluated traits between treatments for each country and botanical type on a mean basis.

Drought stress response was calculated as the difference between well-water and water-stressed treatments on a mean basis for the two years. The broad-sense heritability was then estimated using the estimate of genetic variance $\sigma^2{}_G$ and the residual variance $\sigma^2{}_E$.

Genome-wide association studies

To identify genomic regions associated with phenotypic traits measured in this experiment, we used 73,730 genome-wide SNP markers in WASAP_Lysi after filtering for minor allele frequency (MAF = 0.05). To account for population structure (Q), we performed principal component analysis (PCA) of the collection on Tassel 5 version 5.2.4 (Glaubitz et al., 2014) while kinship between genotype (K) inferred in GAPIT (Lipka et al., 2012). Next, we conducted GWAS for yield components in both treatments separately for WW, WS, and the difference between WW and WS. BLUPS were estimated with the linear model above mentioned and carried out GWAS using the general linear model (GLM) and mixed linear model (MLM) and multilocus mixed linear model (MLMM) (Lipka et al., 2012; Segura et al., 2012) for plant height, days to flowering, vegetative biomass, panicle length, and grain weight for well-watered and water-stressed conditions across two years. The difference between WW and WS phenotypic values was estimated for each genotype as in phenotypes for the MLMM model. We used the Bonferroni correction threshold ($\alpha = 0.05$) to identify significant markers associated with the phenotypes. To identify outlier SNPs (marker-trait associations, MTAs) associated with each trait, the top 0.01% lowest P-values in the MLMM model were selected.

To determine whether MTAs co-localized with previously identified and putative genes, a review of *a priori* candidate genes was conducted. A list of *a priori* candidate genes including

their genomic positions (n = 67) in sorghum and orthologs of rice and maize for plant height, maturity, grain yield, biomass, and drought tolerance (Table A.1) was generated based on the literature review. The sorghum QTL Atlas was used to determine co-localizations of MTAs with previously identified QTLs in multiple studies (Mace et al., 2018).

Results

Genetic structure of WASAP_Lysi relative to SAP and WAGRIN

To characterize the genetic diversity of WASAP_Lysi with respect to the global sorghum diversity and the complete WASAP, we analyzed 90,148 genome-wide SNPs to characterize the genetic diversity of WASAP_Lysi with respect to the global sorghum diversity and the complete WASAP. Using SAP as the global diversity germplasm consisting of all the botanical types of sorghum, the first principal and second principal component explained 8.7% and 6.3% of the total variation, respectively (Figure 3.1A). When the WASAP_Lysi was projected on the principal components of SAP (Figure 3.1A), the WASAP_Lysi genotypes clustered with the caudatum, durra, and guinea in SAP on the first two PCs (Figure 3.1B). As expected, none of the genotypes were found in the kafir group of SAP as this type is mostly found in southern Africa.

To characterize the genetic structure of WASAP_Lysi with respect to West African sorghum, we conducted PCA on 138,000 genome-wide SNPs. To ensure that the variation in this subset covers most of the diversity existing in West Africa, we compared the WASAP_Lysi genotypes to West African genotypes available in the global collection of USDA-GRIN (WAGRIN; Figure 3.1C). The first principal and second principal components of WAGRIN explain 13.2% and 12.2% of the overall variation, respectively. The first principal component separates this guinea cluster from most of the West African genotypes. The majority of the genotypes of WAGRIN were found along with the second principal component where all

WASAP_Lysi genotypes were clustered. We observed a distinct cluster composed of genotypes from Senegal and Togo (guinea types) mainly that were not sampled in the WASAP_Lysi (Figure 3.1D).

Since the majority of WASAP_Lysi do not have botanical types assigned phenotypically, we used sequencing data to classify genotypes into major types with respect to the original classification of each genotype. Principal components of SAP were used to predict the botanical types of WASAP_Lysi as most of them were found in three major types of SAP. Following our observations in the SAP, PC1, PC2, and PC3 distinguish major groups of caudatum, durra-caudatum, and guinea respectively (Figure 3.1A-B). Moreover, we infer botanical types from admixture (K = 3) using membership probability > 0.6 to identify the major groups even though the lowest cross-validation was found at K = 6. We also took into account the original botanical types available to infer the botanical types. Then, each WASAP_Lysi genotype has the assigned botanical types which are further used for phenotypic analyses (Figure B.5).

Water extraction and harvest indices variability

Under the hypothesis that late pre-flowering water deficit reduces water use in sorghum, we predicted that total water transpiration would be significantly higher in WW than WS treatments. Analysis on water transpiration showed significant genotypic variation. Tx7000 had less water transpired in both treatments compared to BTx642 (Figure 3.3A-C). With regards to days after planting, cumulative water transpired weekly in WS treatment was greater than the WW treatment (Figure 3.3B). The maximum water use in both genotypes was reached two weeks after flowering. Reduced water availability in WS treatment delayed significantly Tx7000 but not BTx642 (Figure 3.3B). For the line Tx7000, the average water transpired in WS was

significantly higher than the WW treatment (p-value > 0.05) (Figure 3.3C). In contrast, there was no significant difference in total water transpired for BTx642 between treatments (Figure 3.3C).

Next, we hypothesized that botanical types are sources of water-use variation in sorghum. We predicted a significant reduction of total water transpired (from week four until physiological maturity) in WS treatment compared to WW treatment in each botanical type group. Total water transpired was significantly reduced within each botanical type (Figure 3.3E). On average, durra-caudum (~ 20 liters) and guinea (~ 20 liters) types have less water transpired in WW treatment than the caudatum type (~ 19 liters). In WS treatment, the average water transpired was 5 to 7 liters in each of the botanical types.

Under the hypothesis that water-deficit reduces yield components in sorghum, we predicted a significant decrease in harvest index under WS treatment compared to WW treatment. Preliminary analysis to determine the impact of water deficit on harvest index showed that there is no considerable reduction of harvest index for Tx7000 under WS compared to WW while the harvest index for BTx642 is significantly reduced (p-value < 0.001) under WS compared to WW (Figure 3.3D). Harvest indices were significantly reduced in all the botanical types evaluated (Figure 3.3E) (p-values < 0.001 for caudatum type and < 0.001 for durra-caudatum and guinea types). We observed that Caudatum types have higher values in both treatments compared to durra-caudatum and guinea types.

Phenotypic and genetic variability

The phenotypic evaluation was conducted under the assumption that managed drought stress in the lysimeter can be used for drought stress experiments. To test this hypothesis, we conducted the experiment including BTx642 and Tx7000 in the lysimeter for both well-watered and water-stressed conditions. The first prediction was for the two experiments conducted in

2017 and 2018, there was no significant variation of grain weight per panicle within treatments and repetitions across years. Treatment-by-year interaction was not significant; similar observations were made for the genotypes replicated within treatments (Table 3.2). Genotype and genotype-by-treatment interactions were significant ($p = 0.001$ and $p = 0.01$, respectively) whereas treatment-by-year interaction was not significant ($p = 0.89$) (Table 3.2).

The second prediction was a significantly lower reduction in grain weight per panicle for pre-flowering drought-tolerant, Tx7000 versus pre-flowering drought-susceptible, BTx642. We used 2018 data since more number of replicates were analyzed ($n = 14$ per genotype, per treatment). Plant height and vegetative biomass did not show significant differences between treatments (Figure 3.4). Days to flowering difference was marginally significant ($p < 0.1$) in Tx7000 with an average delay of 3 days in WS compared to WW, but not significant in BTx642 ($p = 0.2$) in which plants tend to flower earlier in WW than WS. As expected, pre-flowering drought-tolerant Tx7000 has less reduction in grain weight compared to pre-flowering drought susceptible BTx642 ($p < 10\text{-}3$) (Figure 3.4). We observed no significant reduction in grain weight for Tx7000 compared to BTx642.

In the WASAP, we observed a variation for the traits measured within WW and WS treatments (Figure B.3). Between treatments, the variation did not reflect a significant difference for plant height (Figure 3.4). Days to flowering differed significantly based on country of origin. Togolese and Senegalese genotypes have an average of 114 and 99, respectively, while an average of 84 and 87 days was recorded for Malian and Nigerien genotypes, respectively (Table B.2). To investigate whether water deficit reduces yield components in the West African germplasm, we estimated BLUP values of the traits for each genotype within treatment in the collection. We analyzed the genotypes per group (country of origin and botanical types) to test

their performance in each treatment under the hypothesis that response to drought differs based on origin and types. Based on the hypothesis that the performance of the genotypes depends on the adaptation to a specific agroclimatic zone, we predicted that there would be a significant decrease in yield components under water deficit in Soudanian (Senegal and Togo) genotypes compared to Sahelian (Mali and Niger) genotypes. A significant decrease in grain weight per panicle was observed under WS for all the genotypes in all countries of origin groups (Figure 3.4, Table B.2). The proportion of phenotypic variance explained by country of origin in WW represents 33%, 19%, 8%, and 7% for days to flowering, vegetative biomass, panicle length, and grain weight of the total variance, respectively.

Next, we compared botanical types in terms of their performance in WS relative to WW under the hypothesis that caudatum and durra-caudatum are more tolerant to water deficit than guinea. There was no significant difference for plant height and days to flowering within botanical type, while variations were observed across botanical types (Table B.2). For instance, taller plants (>270 cm) with longer days to flowering (> 100 days) belong to the guinea group. In contrast, intermediate height (200–250 cm) and days to flowering (< 100 days) were mostly found in caudatum and durra-caudatum groups (Figure 3.4). There were significant reductions in vegetative biomass for all botanical types between WW and WS. We also observed a significant reduction of grain weight per panicle in all three groups in WS (Table B.3). The proportion of phenotypic variance explained by botanical type in WW represents 12%, 14%, 16%, and 30% for days to flowering, vegetative biomass, panicle length, and grain weight of the total variance, respectively.

The contribution of genotypes as a source of variation was significant for all traits measured in WW and WS treatments. In all traits, estimated broad-sense heritability ranged from

0.75 to 0.96, with the highest heritabilities for days to flowering in WW (Table 3.3). Similar observations were made in WS in which the contribution of the genotypes is significant for all traits. For the drought stress response, plant height heritability was 0.05, while vegetative biomass, panicle length, grain weight, and days to flowering have relatively higher heritability (Table 3.3).

GWAS for plant height and days to flowering under well-watered treatment using GLM and MLM models

Under WW, significant marker-trait association (MTA) for the measured traits are predicted. We calculated mean values from replicated genotypes within a year for final plant height and days to flowering since the collection was composed of diverse germplasm. Using the GLM model, MTAs, above the Bonferroni threshold ($\alpha = 0.05$), for plant height in 2017 were identified on chromosomes 4, 6, 7, and 8 with 11% and 25% of the total phenotypic variance. Among the significant associations ($\alpha = 0.05$), S7_59412395 (MAF = 0.05) explained 19% of the total variation and co-localized with the classical dwarfing gene *Dw3* (~ 0.4 Mb) (Figure 3.5). In 2018 data, 19 significant MTAs were located on chromosomes 6, 7, and 10 at the Bonferroni threshold (Figure 3.5). The total amount of phenotypic variation explained ranged from 12% to 23%. In the MLM model, we observed ten MTAs on chromosomes 7 and 10 in both years, contributing between 12% to 19% of the total phenotypic variation. For the flowering time, three associated markers above the significant threshold were identified on chromosome 6 (S6_50696803; MAF = 0.13, S6_50716126; MAF = 0.16, and S6_50716244; MAF = 0.16) for 2017 data in the GLM model whereas no significant markers were identified in 2018 data above the Bonferroni threshold (Figure 3.6). Those MTAs did not co-localize with our *a priori*

candidate genes for flowering time and maturity. In the MLM model, no MTAs were identified above the Bonferroni threshold in both years (Figure 3.6).

Marker-trait association (MTA) under well-watered and water-stressed treatments

To further identify genomic regions associated with phenotypic variation under WW conditions across years, we conducted association analysis using MLMM for plant height, days to flowering, vegetative biomass, panicle length, and grain weight on BLUP values. Besides three significant associations for plant height and days to flowering identified, there were no associations found for vegetative biomass, panicle length, and grain weight per panicle under WW conditions above the conservative threshold (Figures 3.7A, 3.8A, 3.9A, 3.10A, 3.11A). We then investigated the top 0.01% lowest p-values of associated markers for each trait (Table A.1). Among the associated markers with vegetative biomass for the top 0.01% lowest p-values, S2_72712959 (MAF = 0.19) is co-localized with QTDBM2.4 (total dry biomass, 70–76 Mb) (Mocoeur et al., 2015), and S7_59459123 (MAF = 0.06) colocalized with *Dw3* gene (~290 kb). For panicle length, S2_73868064 (MAF = 0.3) is 65 kb away from the maize *ramosa3* ortholog of Sobic.002G381600 for inflorescence morphology (Satoh-Nagasawa et al., 2006). S6_47319583 (MAF = 0.14) was found ~185 kb from leading SNPs for panicle erectness in the diversity panel (Hu et al., 2019) and within the QTL region QDTFL6.50 (0.5 Mb region) associated with flowering time (Mace et al., 2013).

In WS for all the traits analyzed (plant height, days to flowering, vegetative biomass, panicle length, and grain weight) (Figures 3.7B, 3.8B, 3.9B, 3.10B, 3.11B), only three MTAs were observed on chromosomes 7 (S7_50055849; MAF = 0.13, S7_59459123; MAF = 0.06) and 9 (S9_56534065; MAF = 0.07). The latter MTA overlaps with QHGHT9.45 (56.1–57.8Mb) for plant height (Figure 3.7B) (Felderhoff et al., 2012). Two MTAs were found on chromosomes 3

(S3_19589652; MAF = 0.13) and 4 (S4_14172212; MAF = 0.11) for days to flowering under WS conditions (Figure 3.8B). No significant associations were found above the threshold for vegetative biomass and panicle length under water deficit conditions (Figures 3.9B, 3.10B). Three associations were found on chromosomes 6 (S6_18075344; MAF = 0.17 and S6_46923493; MAF = 0.06) and 7 (S7_50055849; MAF = 0.14) for grain weight per panicle above the conservative threshold (Figure 3.11B).

Marker-trait association (MTA) for responses to water stress

To identify genomic regions associated with responses to water stress using MLMM, we used BLUPs values for each trait to calculate the difference between WW and WS GWAS (Figures 3.7C, 3.8C, 3.9C, 3.10C, 3.11C). MTA at S4_149131 (MAF = 0.04) for plant height was identified above the conservative threshold (Figure 3.7C). The associated marker was found above the conservative threshold on chromosome 5 (S5_52090779; MAF = 0.07) for flowering time (Figure 3.8C). This marker is colocalized with QDTFL5.8 (9 Mb region) for average days to flowering (Mace et al., 2013). Among the top 0.01% associated markers for responses to water stress (Table B.3), S5_64932515 (MAF = 0.05, Figure 3.8C) in days to flowering co-localized with QDTFL5.10 (0.8 Mb) controlling days to flowering (Srinivas et al., 2009). We found three associated markers chromosome 2 S2_57620549, S2_57664163, and S2_57663973 (MAF= 0.4, 0.36, and 0.37 respectively) spanning ~43 kb region within *Stg3a* for the difference of WW and WS in vegetative biomass (Figure 3.9C, Table B.4). Those three MTAs are in linkage disequilibrium with each other (Figure B. 4). MTA S5_16199662 (MAF = 0.09) for grain weight co-localized within *Stg4* loci (Figure 3.11C). Shared SNPs on chromosomes 4, 5, and 7 were observed for multiple traits (Figure 3.12, Figure 3.13). For instance, S7_59459123 (MAF = 0.06)

is significant for plant height in both water conditions, while S7_50055849 (MAF = 0.14) is significant for grain weight and vegetative biomass under WS (Figure 3.13).

Discussion

Evidence of quality managed drought experiment in the lysimeter

High-quality phenotypic data is essential for dissecting complex traits (Mackay et al., 2009; Cobb et al., 2013). Under stress conditions, phenotyping requires adequate screening with known genotypes for their responses to validate the experiment. In our study, we first analyzed known genotypes to test whether they perform similarly, as confirmed in previous studies. We observed genotype-by-environment interaction between checks (Table 3.2); Tx7000 showed a lower reduction in grain weight per panicle than BTx642 under WS (Figure 3.4), as found in earlier studies (Tuinstra et al., 1996; Harris, 2007; Borrell et al., 2014). Surprisingly, BTx642 and Tx7000 did not perform well in this experiment as some WASAP_Lysi genotypes in well water conditions (Figures B.1, B.2). Moreover, the plant heights of Tx7000 and BTx642 are shorter (< 90 cm) than the local genotypes that are taller and preferred for their multiple purposes in smallholder farming systems of West Africa (National Research Council, 1996). We used only two genotypes, as sources for known drought responses in this study; investigating more genotypes with multiple replicates will be useful when using large scale trials for drought experiments.

To ensure accurate phenotypic data, an appropriate experimental design with precise and accurate data collection is required (Tuberosa, 2012; Pieruschka and Schurr, 2019). We observed significant variations of the phenotypic data across years in this study (Figure B.1). Note, the number of replicated genotypes was increased in 2018 to provide more accurate data than the 2017 experiment, specifically the checks (Figures B.2, B.3). This observation suggests that in

managed experiments, often in a single location, a sufficient number of experimental units and blocks would be needed to reduce errors for non-adapted genotypes in the collection (Nettleton, 2006). Moreover, phenotyping errors may also be due to the measurement of some traits as they were recorded manually. For precise and accurate data, using cost-effective phenotyping approaches would be needed (Vadez et al., 2015) to increase phenotyping accuracy.

Responses to water stress in West African sorghum captured by panicle length and grain weight

West African sorghum harbors extensive phenotypic diversity under rainfed conditions and drought scenarios (Haussmann et al., 1998; Bhosale et al., 2011; Upadhyaya et al., 2017a; b). The phenotypic diversity existing in West African sorghum under WW and WS conditions leads us to hypothesize that there are different responses of water deficit in WA sorghum. The observed variation of the WASAP in the lysimeter system in WW reflects the underlying genetic variation (Tables B.2, B.3), even though we analyzed ~219 genotypes in the experiment. The heritabilities were slightly higher in WS than WW for all traits except days to flowering (Table 3.3). Heritabilities for panicle length were ~ 0.9 in both treatments and lower in response to drought (WW-WS, $H^2 = 0.06$), suggesting the water deficit did not impact panicle exertion as discussed in previous studies (Blum et al., 1989; Assefa et al., 2010).

Water deficit conditions impact flowering by either shortening or delaying days to flowering (Blum, 2004; Barnabás et al., 2008). In contrast, for our study, there was no significant variation in days to flowering between WW and WS (Figure 3.4). Our observation implies that the water stress imposition at flag-leaf did not delay flowering in most of the genotypes since the panicle is entirely developed, and anthesis starts a few days after flag-leaf appearance (Smith and

Frederiksen, 2000; Blum, 2004). However, panicle exertion and grain filling might be the most affected by the pre-flowering water deficit (Rosenow et al., 1983; Assefa et al., 2010).

Grain yield under stressed conditions is an essential component for drought tolerance in crops (Blum, 2004; Sinclair, 2011). Our analysis of botanical types reveals significant grain weight reduction in all major types confirming the effect of water deficit (Figure 3.4, Figure B.1). The reduction of grain weight in the study was significant in botanical types and country of origin, suggesting that a moderate water deficit can trigger sorghum development specifically during the reproductive stage (Blum, 2004). We also observed similar grain weight in WW for durra and guinea types, not reflecting previous observations that durra types have higher seed weight than guinea types (Wang et al., 2019). One reason for these contrasting observations is based on the fact that we conducted the study in lysimeter tubes restricting plant development, especially for durra types. Probably, durra types were assayed in a stressed environment (i.e. PVC tubes), limiting roots expansion to capture more available water in the soil. In this perspective, further research focusing on the evaluation of durra and guinea botanical types for grain weight and root architecture in the lysimeter and fields might shed light on sorghum responses to new environments.

Representative genetic diversity is captured in WASAP for dissecting complex traits

Sorghum exhibits large phenotypic diversity in West Africa (Bhosale et al., 2011; Upadhyaya et al., 2017b). The observation of a distinct cluster composed of mostly Senegalese landraces from GRIN (Figure 3.1C-D) may suggest that the selection of WASAP collection covers specific regions. It might be possible due to the narrowed diversity in breeding programs or loss of genetic diversity in the current collection as accessions from GRIN were collected in

the 1970s (Walker and Alwang, 2015; Upadhyaya et al., 2017b). Variations across precipitation gradients have been observed in different countries in West Africa from USDA-GRIN (Maina et al., 2018; Olatoye et al., 2018; Faye et al., 2019) may reveal genomic regions of interest under diverse agroclimatic zones and in common garden.

MTAs show evidence of phenotypic variations under managed environments

Association studies rely on accurate phenotyping, genotyping, and statistical models to increase the detection of true associations (Long and Langley, 1999; Ball, 2013; Wang et al., 2014). Although this collection did not capture a wide range of major climatic zones in West Africa, we found associated markers underlying yield components in water deficit conditions. Compared to the GLM and MLM, using multiple loci simultaneously, MLMM allows a greater identification of significant markers and reduced false discovery (Segura et al., 2012). Above the conservative threshold ($\alpha = 0.05$), we found 13 MTAs across traits and treatments, only a few of those MTAs co-localized with known QTLs while novel associations were found in the study. These results indicate that accounting for population structure is not sufficient for genetically diverse germplasm. It might be useful to account for plant height, flowering time, additional botanical types, and the origin as covariates in association studies when using a large number of genotypes.

With the assumption that the WASAP is composed of landraces and breeding lines from different countries, we predicted that MTAs co-localized with *Ma6* previously found in West Africa (Olatoye et al., 2018; Faye et al., 2019). But we did not find MTAs co-localized with *a priori* flowering time genes. Note, the WASAP was first evaluated in common garden fields in Bambey (Senegal) (Sine et al., *in prep*), and genotypes with less incidence of photoperiod sensitivity were selected for the drought experiment thus there are not enough genotypes

segregating for *Ma6* or *Ma1*. This might be one of the reasons flowering time associated markers were not observed in both WW and WS among *a priori* candidate genes. Instead, the majority of Malian and Togolese genotypes have longer days to flowering that is associated with S4_64398335 (Table 3.4). The C allele has a minor allele frequency at S4_64398335 for longer days to flowering under WW. Also, Togolese genotypes (n=18) with the G allele at S4_14172212 have longer days to flowering under WS (Table 3.4). Such MTAs might be under consideration when improving varieties from the Soudanian agroclimatic zones especially for guinea botanical types and farmer preferred traits (Kante et al., 2017).

MTAs identified in regard to drought responses across traits are candidates for water deficit adaptation. Among the top 0.01% of the lowest P-values, MTAs were co-localized with previously identified QTLs (Mace et al., 2018). Panicle length associated markers were co-localized with three QTLs suggesting that the panicle exertion was not only affected by botanical types but also water deficit in the managed experiment. We found stay-green regions associated with grain weight in this study (Table B.3), while previous associations were found under the post-flowering water deficit (Jordan et al., 2012; Mace et al., 2012). The shared MTA S7_50055849 showed variation in vegetative biomass and grain weight under water deficit (Figure 3.13) could be explained by a possible source to sink remobilization to complete grain filling (Blum, 1998; Tuberosa, 2012; Ongom et al., 2016).

For the difference between WW and WS of grain weight, the variation was larger for the genotypes with the G allele at S6_18075344 (Figure 3.12). We found that at this MTA, Tx7000 has the A allele while BTx642 has the G allele. Malian (n=7) and Nigerien (n=12) genotypes have similar alleles as Tx7000. To test that the MTA interacts with grain yield in drought conditions, a significant reduction in grain weight is predicted for the contrasting genotypes at

the G allele compared to the A allele under a managed environment and field conditions. In contrast, the second associated marker on chromosome 6 (S6_46923493) (Figure 3.12), Nigerien genotypes (n= 4) carry the minor allele A. One hypothesis is that the QTL controls grain yield under pre-flowering drought. Further analysis will confirm whether the SNPs tag the variation of grain yield in West African germplasm.

MTAs of WW and WS differences of vegetative biomass (S2_57620549, S2_57664163, and S2_57663973) and grain weight (S5_16199662) (Figures 3.9 and 3.11) that co-localized with *Stg3a* and *Stg4* respectively, are not in favorable coupling in this study. The number of genotypes evaluated was not sufficient to dissect perfectly the contribution of S5_16199662 (a rare allele) on grain weight under WS treatment. Even though MTAs in this study did not co-localized with additional stay-green loci (*Stg1* and *Stg2*), those loci might be present in the West African germplasm that was not captured in the subset. Signatures of selection might have acted on major stay green loci individually in West African sorghum (Tuinstra et al., 1996; Harris, 2007). Along with those associations further hypothesis testing might underlie the lack of statistical power associated with our data.

The statistical power to detect QTL may be influenced by factors, such as the number of genotypes, experimental design, environmental effects, and the population structure (Long and Langley, 1999). Note, in our experiment, a representative subset of the WASAP was evaluated based on our hypothesis that much of the diversity is captured to detect variations. Multiple phenotyping may enhance the phenotypic variance and reduce errors (Atwell et al., 2010; Korte and Farlow, 2013). A larger number of genotypes and phenotyping data would increase the statistical power of detecting drought tolerance MTAs used in marker-assisted selection.

Potential for breeding for drought tolerance in semi-arid regions

In sub-Saharan Africa, water deficit occurs throughout the growing season, from seedling to terminal stages of sorghum development (Mohamed et al., 2002; Biasutti, 2019). Depending on the spatial distribution of drought severity, sorghum is more or less affected in some regions. The West African sorghum association panel is a combined effort germplasm exchange between countries relevant for crop improvement with a considerable number of genotypes that efficiently represent the existing genetic diversity in West Africa. Most of the genotypes in the WASAP derived from to breeder's working collections in each country. In our study, we focused on investigating pre-flowering drought stress specifically at flag leaf appearance. Our experiment provides a primary starting point of future managed drought experiments with controls to mimic pre-flowering drought stress for ecophysiological genomic studies and crop improvement. Nevertheless, in the breeding perspective, further studies would be needed to test the hypothesis under which the managed experiment reflects the targeted population environments.

West African soils are characterized by low fertility, while farmers use low inputs during growing seasons (Leiser et al., 2014). Although we used NPK application in our experiment, we were not able to compare the sorghum performance without fertilizers as in some farmer's fields with low input or under a sustainable intensification system and multiple environments. We applied the microdose technique, a recommended fertilization method for farmers consisting of the application of inorganic fertilizer (6g/hill) at planting (Ministère de l'agriculture du Niger, 2012). Increasing the quantity of fertilizer might result in higher yields, thus a different interpretation of the findings since we aim to replicate as closely as possible farmer's fields with low nutrient. In the breeding perspective, developing varieties under recommended nutrient input would be preferable when conducting experiments on stations and farmer's fields. In addition to

participatory selection, breeding programs should orient product developments towards exploring genotype by environment by management interactions.

Considering the limited knowledge and use of genomic tools to accelerate variety development in small breeding programs (Walker and Alwang, 2015), our results provide a step forward in understanding the genetic basis of adaptation under water deficit in West African sorghum for breeding programs. Diagnostic markers, when validated, might be used for marker selection across West Africa, such as stay-green markers available in high-throughput genotyping services. Once confirmed in multiple environments, MTAs identified in this study might be good candidates for developing cost-effective markers for drought-tolerant sorghum with respect to the local environment and farmer's preferences. Some phenotyping approaches have major questions remaining to be answered in the breeding perspective. For instance, managed experiments can lead to a better understanding of the physiology of the plants. However, performance may differ when those plants are in field conditions (e.g. TPE) where some parameters are not controlled (Brachi et al., 2010). More research is needed to test whether there is a high correlation between the lysimeter and TPEs.

References

Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D.G. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson, and M. Nordborg. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:627–631. doi:10.1038/nature08800

Ball, RoderickD. 2013. Designing a GWAS: Power, Sample Size, and Data Structure. C. Gondro, J. van der Werf, B. Hayes, and J. van der Werf, ed. Methods in Molecular Biology. No. 1019. Humana Press.

Barbé, L.L., and T. Lebel. 1997. Rainfall climatology of the HAPEX-Sahel region during the years 1950–1990. Journal of Hydrology 188–189:43–73. doi:10.1016/S0022-1694(96)03154-X

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:1–48. doi:10.18637/jss.v067.i01

Biasutti, M. 2019. Rainfall trends in the African Sahel: Characteristics, processes, and causes. Wiley Interdisciplinary Reviews: Climate Change 10:e591. doi:10.1002/wcc.591

Billot, C., P. Ramu, S. Bouchet, J. Chantereau, M. Deu, L. Gardes, J.-L. Noyer, J.-F. Rami, R. Rivallan, Y. Li, P. Lu, T. Wang, R.T. Folkertsma, E. Arnaud, H.D. Upadhyaya, J.-C. Glaszmann, and C.T. Hash. 2013. Massive Sorghum Collection Genotyped with SSR Markers to Enhance Use of Global Genetic Resources. PLOS ONE 8:e59714. doi:10.1371/journal.pone.0059714

Blum, A. 1998. Improving wheat grain filling under stress by stem reserve mobilisation. Euphytica 100:77–83. doi:10.1023/A:1018303922482

Blum, A. 2004. Sorghum Physiology. H. Nguyen and A. Blum, ed. CRC Press.

Boopathi, N.M. 2013. Phenotyping. Springer, India.

Borrell, A.K., J.E. Mullet, B. George-Jaeggli, E.J. van Oosterom, G.L. Hammer, P.E. Klein, and D.R. Jordan. 2014. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. J. Exp. Bot. eru232. doi:10.1093/jxb/eru232

Brachi, B., N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, J. Bergelson, J. Cuguen, and F. Roux. 2010. Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLOS Genet 6:e1000940. doi:10.1371/journal.pgen.1000940

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am J Hum Genet 84:210–223. doi:10.1016/j.ajhg.2009.01.005

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community Resources and Strategies for Association Mapping in Sorghum. Crop Science 48:30–40. doi:10.2135/cropsci2007.02.0080

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126:867–887. doi:10.1007/s00122-013-2066-0

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. doi:10.1098/rstb.2007.2170

Crain, J., M. Reynolds, and J. Poland. 2017. Utilizing High-Throughput Phenotypic Data for Improved Phenotypic Selection of Stress-Adaptive Traits in Wheat. Crop Science 57:648–659. doi:10.2135/cropsci2016.02.0135

Doggett, H. 1988. Sorghum. 2nd Edition.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6:e19379. doi:10.1371/journal.pone.0019379

Fang, Y., and L. Xiong. 2015. General mechanisms of drought response and their application in drought resistance improvement in plants. Cell. Mol. Life Sci. 72:673–689. doi:10.1007/s00018-014-1767-0

Faye, J.M., F. Maina, Z. Hu, D. Fonceka, N. Cisse, and G.P. Morris. 2019. Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. Ecology and Evolution 9:6038–6051. doi:10.1002/ece3.5187

Felderhoff, T.J., S.C. Murray, P.E. Klein, A. Sharma, M.T. Hamblin, S. Kresovich, W. Vermerris, and W.L. Rooney. 2012. QTLs for Energy-related Traits in a Sweet × Grain Sorghum [Sorghum bicolor (L.) Moench] Mapping Population. Crop Science 52:2040–2049. doi:10.2135/cropsci2011.11.0618

Flint-Garcia, S.A. 2013. Genetics and Consequences of Crop Domestication. J. Agric. Food Chem. 61:8267–8276. doi:10.1021/jf305511d

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One 9. doi:10.1371/journal.pone.0090346

Halilou, O., F. Hamidou, B.K. Taya, S. Mahamane, and V. Vadez. 2015. Water use, transpiration efficiency and yield in cowpea (Vigna unguiculata) and peanut (Arachis hypogaea) across water regimes. Crop Pasture Sci.. doi:10.1071/CP14182

Hamblin, M.T., E.S. Buckler, and J.-L. Jannink. 2011. Population genetics of genomics-based crop improvement methods. Trends in Genetics 27:98–106. doi:10.1016/j.tig.2010.12.003

Harris, K.R. 2007. Genetic analysis of the Sorghum bicolor stay-green drought tolerance trait. Ph.D. Thesis. Texas A&M University, United States -- Texas

Hathorn, A., and S.C. Chapman. 2014. Historical and Prospective Applications of 'Quantitative Genomics' in Utilising Germplasm Resources. R. Tuberosa, A. Graner, and E. Frison, ed. Springer Netherlands.

Hu, Z., M.O. Olatoye, S. Marla, and G.P. Morris. 2019. An Integrated Genotyping-by-Sequencing Polymorphism Map for Over 10,000 Sorghum Genotypes. The Plant Genome 12. doi:10.3835/plantgenome2018.06.0044

Jordan, D.R., C.H. Hunt, A.W. Cruickshank, A.K. Borrell, and R.G. Henzell. 2012. The Relationship Between the Stay-Green Trait and Grain Yield in Elite Sorghum Hybrids Grown in a Range of Environments. Crop Science 52:1153–1161. doi:10.2135/cropsci2011.06.0326

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29. doi:10.1186/1746-4811-9-29

Lasky, J.R., H.D. Upadhyaya, P. Ramu, S. Deshpande, C.T. Hash, J. Bonnette, T.E. Juenger, K. Hyma, C. Acharya, S.E. Mitchell, E.S. Buckler, Z. Brenton, S. Kresovich, and G.P.

Morris. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. Science Advances 1:e1400218. doi:10.1126/sciadv.1400218

Leiser, W.L., H.F.W. Rattunde, E. Weltzien, N. Cisse, M. Abdou, A. Diallo, A.O. Tourè, J.V. Magalhaes, and B.I. Haussmann. 2014. Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African Sorghum. BMC Plant Biology 14:206. doi:10.1186/s12870-014-0206-6

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. doi:10.1093/bioinformatics/btp698

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Long, A.D., and C.H. Langley. 1999. The Power of Association Studies to Detect the Contribution of Candidate Genetic Loci to Variation in Complex Traits. Genome Res. 9:720–731. doi:10.1101/gr.9.8.720

Mace, E., D. Innes, C. Hunt, X. Wang, Y. Tao, J. Baxter, M. Hassall, A. Hathorn, and D. Jordan. 2018. The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. Theor Appl Genet. doi:10.1007/s00122-018-3212-5

Mace, E.S., C.H. Hunt, and D.R. Jordan. 2013. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet 126:1377–1395. doi:10.1007/s00122-013-2059-z

Mace, E.S., V. Singh, E.J.V. Oosterom, G.L. Hammer, C.H. Hunt, and D.R. Jordan. 2012. QTL for nodal root angle in sorghum (Sorghum bicolor L. Moench) co-locate with QTL for traits associated with drought adaptation. Theor Appl Genet 124:97–109. doi:10.1007/s00122-011-1690-9

Mackay, T.F.C., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577. doi:10.1038/nrg2612

Maina, F., S. Bouchet, S.R. Marla, Z. Hu, J. Wang, A. Mamadou, M. Abdou, A.-A. Saïdou, and G.P. Morris. 2018. Population genomics of sorghum (Sorghum bicolor) across diverse agroclimatic zones of Niger. Genome 1–10. doi:10.1139/gen-2017-0131

McCormick, R.F., S.K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, M. Amirebrahimi, B.D. Weers, B. McKinley, A. Mattison, D.T. Morishige, J. Grimwood, J. Schmutz, and J.E. Mullet. 2018. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal 93:338–354. doi:10.1111/tpj.13781

Messina, C.D., D. Podlich, Z. Dong, M. Samples, and M. Cooper. 2011. Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62:855–868. doi:10.1093/jxb/erq329

Ministère de l'agriculture du Niger. 2012. Catalogue national des espèces et variétés végétales du Niger

Mocoeur, A., Y.-M. Zhang, Z.-Q. Liu, X. Shen, L.-M. Zhang, S.K. Rasmussen, and H.-C. Jing. 2015. Stability and genetic control of morphological, biomass and biofuel traits under temperate maritime and continental conditions in sweet sorghum (Sorghum bicolour). Theor Appl Genet 128:1685–1701. doi:10.1007/s00122-015-2538-5

Mohamed, A.B., N. van Duivenbooden, and S. Abdoussallam. 2002. Impact of Climate Change on Agricultural Production in the Sahel – Part 1. Methodological Approach and Case Study for Millet in Niger. Climatic Change 54:327–348. doi:10.1023/A:1016189605188

Morrell, P.L., E.S. Buckler, and J. Ross-Ibarra. 2012. Crop genomics: advances and applications. Nat Rev Genet 13:85–96. doi:10.1038/nrg3097

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. PNAS 110:453–458. doi:10.1073/pnas.1215985110

National Research Council. 1996. Lost Crops of Africa: Volume I: Grains

Nettleton, D. 2006. A Discussion of Statistical Methods for Design and Analysis of Microarray Experiments for Plant Scientists. Plant Cell 18:2112–2121. doi:10.1105/tpc.106.041616

Olatoye, M.O., Z. Hu, F. Maina, and G.P. Morris. 2018. Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. G3: Genes, Genomes, Genetics 8:3269–3281. doi:10.1534/g3.118.200551

Ongom, P.O., J.J. Volenec, and G. Ejeta. 2016. Selection for drought tolerance in sorghum using desiccants to simulate post-anthesis drought stress. Field Crops Research 198:312–321. doi:10.1016/j.fcr.2016.03.015

Paterson, A.H. 2013. The Sorghum Genome Sequence: A Core Resource for Saccharinae Genomics. A.H. Paterson, ed. Plant Genetics and Genomics: Crops and Models. No. 11. Springer New York.

Pieruschka, R., and U. Schurr. 2019. Plant Phenotyping: Past, Present, and Future. https://spj.sciencemag.org/plantphenomics/2019/7507131/ (accessed November 26, 2019).

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Satoh-Nagasawa, N., N. Nagasawa, S. Malcomber, H. Sakai, and D. Jackson. 2006. A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441:227–230. doi:10.1038/nature04725

Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825–830. doi:10.1038/ng.2314

Smith, C.W., and R.A. Frederiksen. 2000. Sorghum, Origin, History, Technology and Production. Wiley Series in Crop Science. SB191.S7 S668 2000. Texas A & M University.

Sultan, B., P. Roudier, P. Quirion, A. Alhassane, B. Muller, M. Dingkuhn, P. Ciais, M. Guimberteau, S. Traore, and C. Baron. 2013. Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa. Environ. Res. Lett. 8:014040. doi:10.1088/1748-9326/8/1/014040

Traoré, S.B., A. Alhassane, B. Muller, M. Kouressy, L. Somé, B. Sultan, P. Oettli, A.C. Siéné Laopé, S. Sangaré, M. Vaksmann, M. Diop, M. Dingkhun, and C. Baron. 2011. Characterizing and modeling the diversity of cropping situations under climatic constraints in West Africa. Atmosph. Sci. Lett. 12:89–95. doi:10.1002/asl.295

Tuberosa, R. 2012. Phenotyping for drought tolerance of crops in the genomics era. Front. Physiol. 3. doi:10.3389/fphys.2012.00347

Tuinstra, M.R., E.M. Grote, P.B. Goldsbrough, and G. Ejeta. 1996. Identification of Quantitative Trait Loci Associated with Pre-Flowering Drought Tolerance in Sorghum. Crop Science 36:1337. doi:10.2135/cropsci1996.0011183X003600050043x

Upadhyaya, H.D., K. Narsimha Reddy, M. Vetriventhan, M. Irshad Ahmed, G. Murali Krishna, M. Thimma Reddy, and S.K. Singh. 2017. Sorghum germplasm from West and Central

Africa maintained in the ICRISAT genebank: Status, gaps, and diversity. The Crop Journal 5:518–532. doi:10.1016/j.cj.2017.07.002

Vadez, V., J. Kholová, G. Hummel, U. Zhokhavets, S.K. Gupta, and C.T. Hash. 2015. LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. J Exp Bot 66:5581–5593. doi:10.1093/jxb/erv251

Vadez, V., J. Kholová, R.S. Yadav, and C.T. Hash. 2013. Small temporal differences in water uptake among varieties of pearl millet (<Emphasis Type="Italic">Pennisetum glaucum</Emphasis> (L.) R. Br.) are critical for grain yield under terminal drought. Plant Soil 371:447–462. doi:10.1007/s11104-013-1706-0

Vadez, V., S. Rao, J. Kholova, L. Krishnamurthy, J. Kashiwagi, P. Ratnakumar, K. Sharma, P. Bhatnagar-Mathur, and P S. Basu. 2008. Roots research for legume tolerance to drought: Quo vadis?. Journal of Food Legumes 21:77–85

Varshney, R.K., R. Terauchi, and S.R. McCouch. 2014. Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. PLOS Biology 12:e1001883. doi:10.1371/journal.pbio.1001883

Walker, T.S., and J. Alwang. 2015. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. CABI.

Wang, J., Z. Hu, H.D. Upadhyaya, and G.P. Morris. 2019. Genomic signatures of seed mass adaptation to global precipitation gradients in sorghum. Heredity 1–14. doi:10.1038/s41437-019-0249-4

Wang, Q., F. Tian, Y. Pan, E.S. Buckler, and Z. Zhang. 2014. A SUPER Powerful Method for Genome Wide Association Study. PLOS ONE 9:e107684. doi:10.1371/journal.pone.0107684

Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, and B.S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. doi:10.1093/bioinformatics/bts606

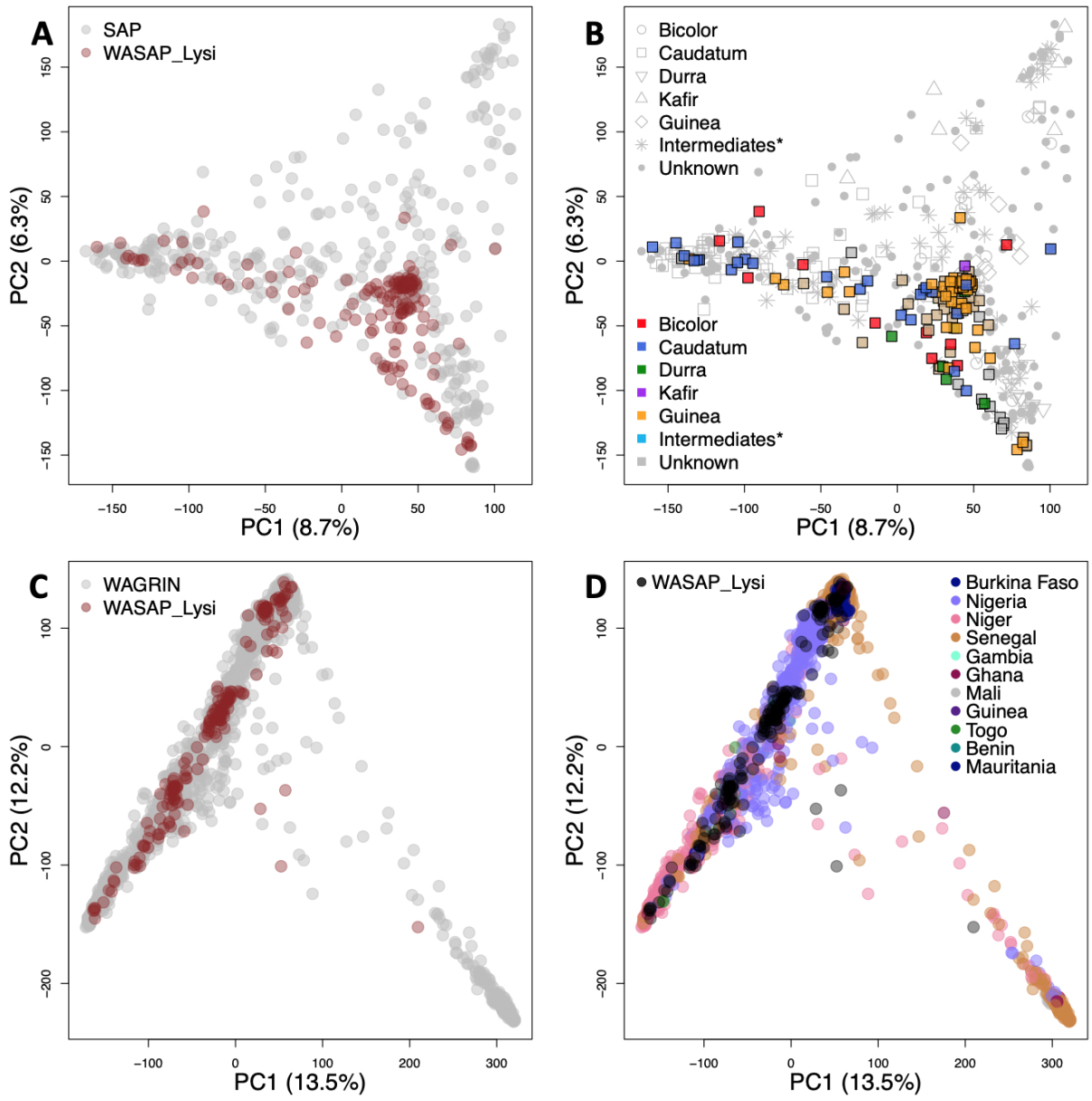


Figure 3.1. Principal component analysis (PCA) of genome-wide polymorphism.

(A) PCA of WASAP_Lysi relative to the Sorghum Association Panel (SAP). (B) PCA of WASAP_Lysi (in brown) relative to the Sorghum Association Panel (SAP) with botanical types information; black squared contours represent the WASAP_Lysi genotypes. (C) PCA of WASAP_Lysi relative to the West African genotypes of USDA-GRIN. (D) PCA of WASAP_Lysi relative to the West African genotype color-coded by country of origin.

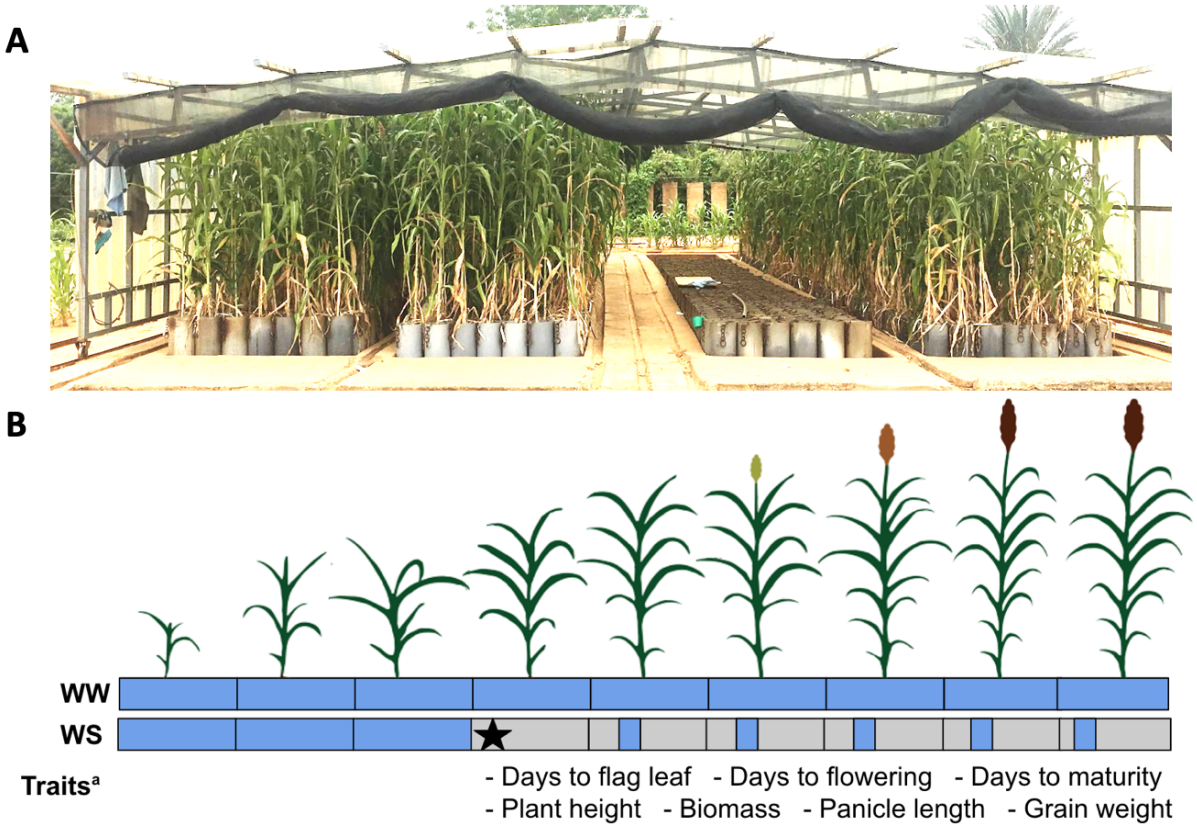


Figure 3.2. Experimental design for phenotypic evaluation in the managed environment.

(A) Lysimeter facility at ICRISAT Sadoré-Niger. (B) A schematic experimental approach in the study for well-watered (WW) and water-stressed (WS) treatments. Intermittent WS treatment was imposed from flag leaf appearance for each individual at different times (represented by the star) until physiological maturity.

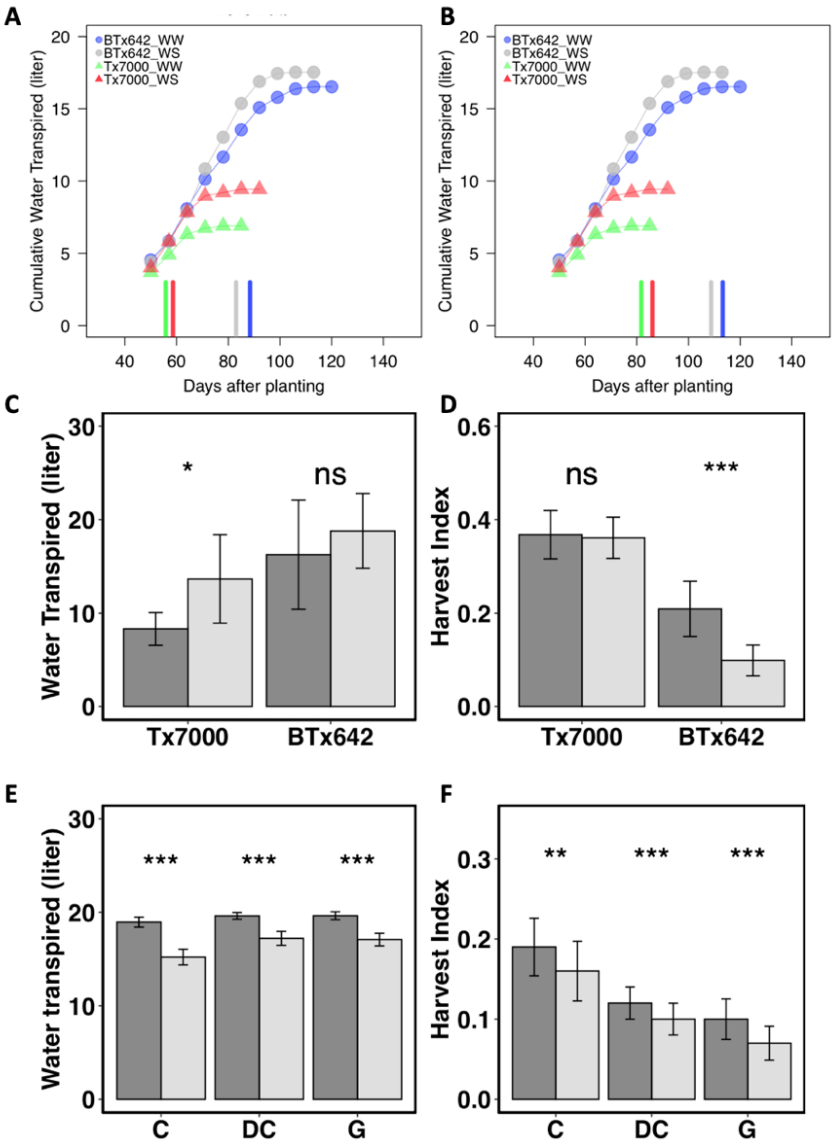


Figure 3.3. Performance of genotypes for water use and harvest indices.

Cumulative water transpired weekly until physiological maturity in well-watered (WW) and water-stressed treatments (WS) for Tx7000 and BTx642 in 2018. Vertical bars represent average days to flowering (A) and average days to maturity (B). Average values of water transpired (C) and harvest indices (D) for Tx7000 and BTx642 under pre-flowering drought stress in 2018 (n = 14). BLUP values of water transpired (E) and harvest indices (F) for genotypes across years grouped by botanical types. Error bars represent the 95% confidence intervals of the mean. Dark bars represent the average performance in well-watered treatment and lightbars represent the

average performance in water-stressed treatment. C, DC, and G are botanical types caudatum, durra-caudatum, and guinea respectively. *, **, *** Significance at $p < 0.05$, 0.01, and 0.001, respectively; ns: not significantly different, $p > 0.05$.

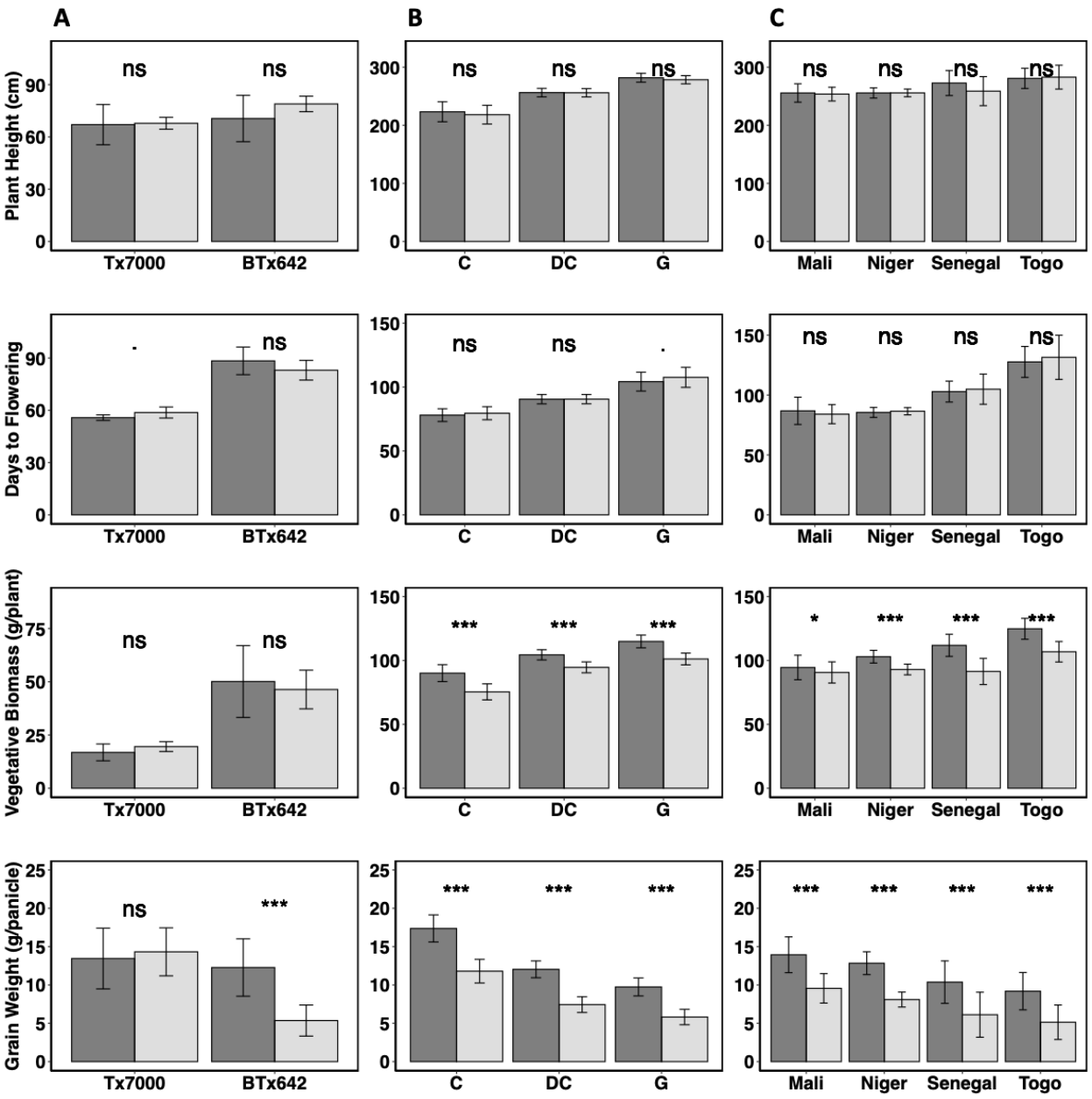


Figure 3.4. Phenotypic performance of genotypes for plant height, days to flowering, vegetative biomass, and grain weight under well-watered and water-stressed.

Average values of BTx642 and Tx7000 under pre-flowering drought stress in 2018 (n = 14) (A). BLUP values of the genotypes across year grouped by country of origin (B) and by genetic classification of botanical types (C).

Error bars represent the 95% confidence intervals of the mean. Light bars represent the average performance in water-stressed treatment and dark bars represent the average performance in well-watered treatment. C, DC, and G are botanical types caudatum, durra-caudatum, and guinea respectively. *, **, *** Significance at $p < 0.05$, 0.01, and 0.001, respectively; ns: not significant.

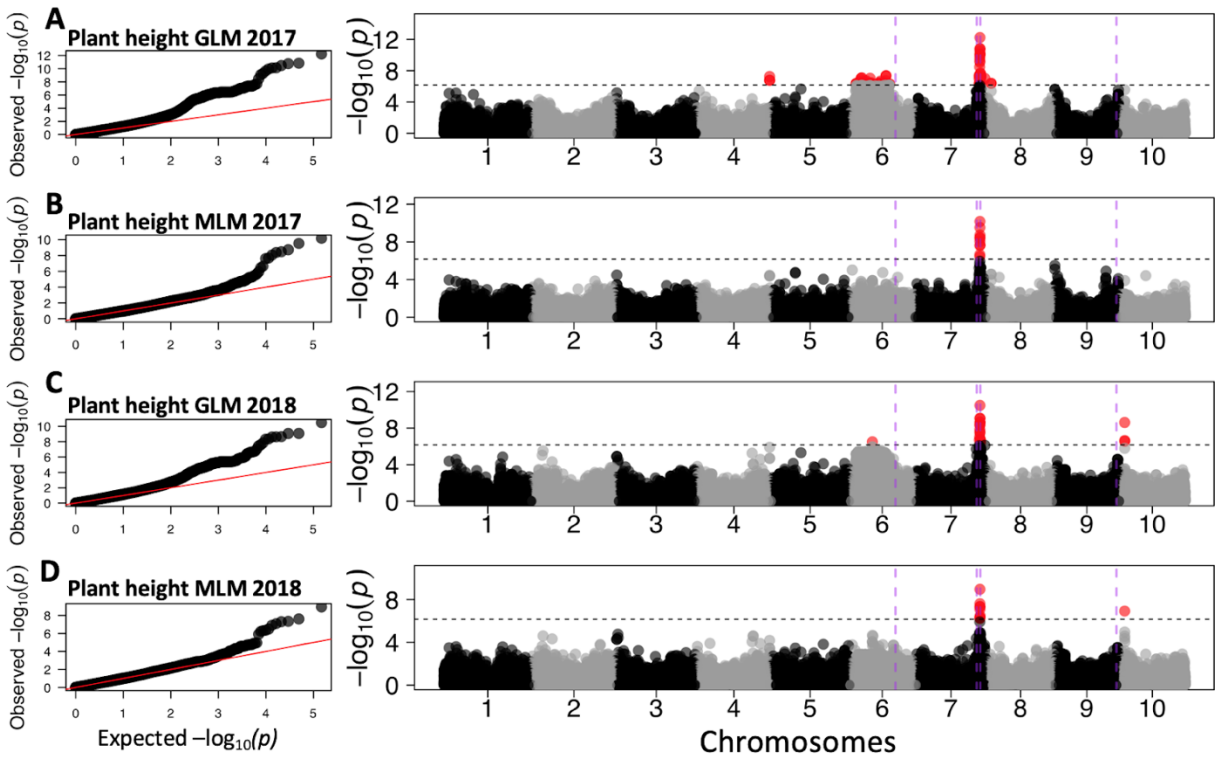


Figure 3.5. Genome-wide association studies of plant height in well-watered treatment.

(A-B) Quantile-quantile and Manhattan plots in 2017 using the general linear model (GLM) (A) and mixed linear model (MLM) (B). (C-D) Quantile-quantile and Manhattan plots in 2018 using general linear model (GLM) (C) and mixed linear model (MLM) (D). Vertical dashed lines (purple) represent *a priori* candidate genes for plant height *Dw2* (chromosome 6), *Qpht7.1* (chromosome 7), *Dw3* (chromosome 7), and *Dw1* (chromosome 9). Horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$).

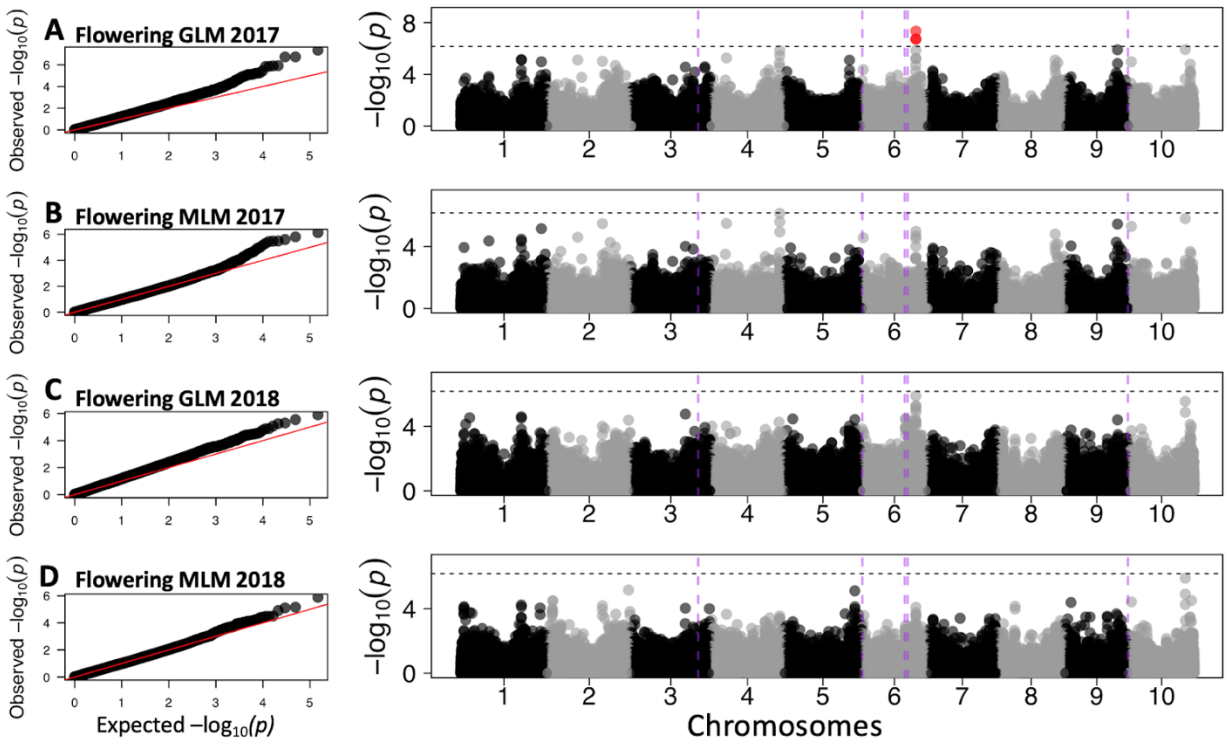


Figure 3.6. Genome-wide association studies of days to flowering in well-watered treatment.

(A-B) Quantile-quantile and Manhattan plots in 2017 using the general linear model (GLM) (A) and mixed linear model (MLM) (B). (C-D) Quantile-quantile and Manhattan plots in 2018 using general linear model (GLM) (C) and mixed linear model (MLM) (D). Vertical dashed lines (purple) represent *a priori* candidate genes for maturity *SbCN12* (chromosome 3), *Ma6* (chromosome 6), *SbCN4* (chromosome 6), *Ma1* (chromosome 6), *SbFL9.1* (chromosome 9). Horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$).

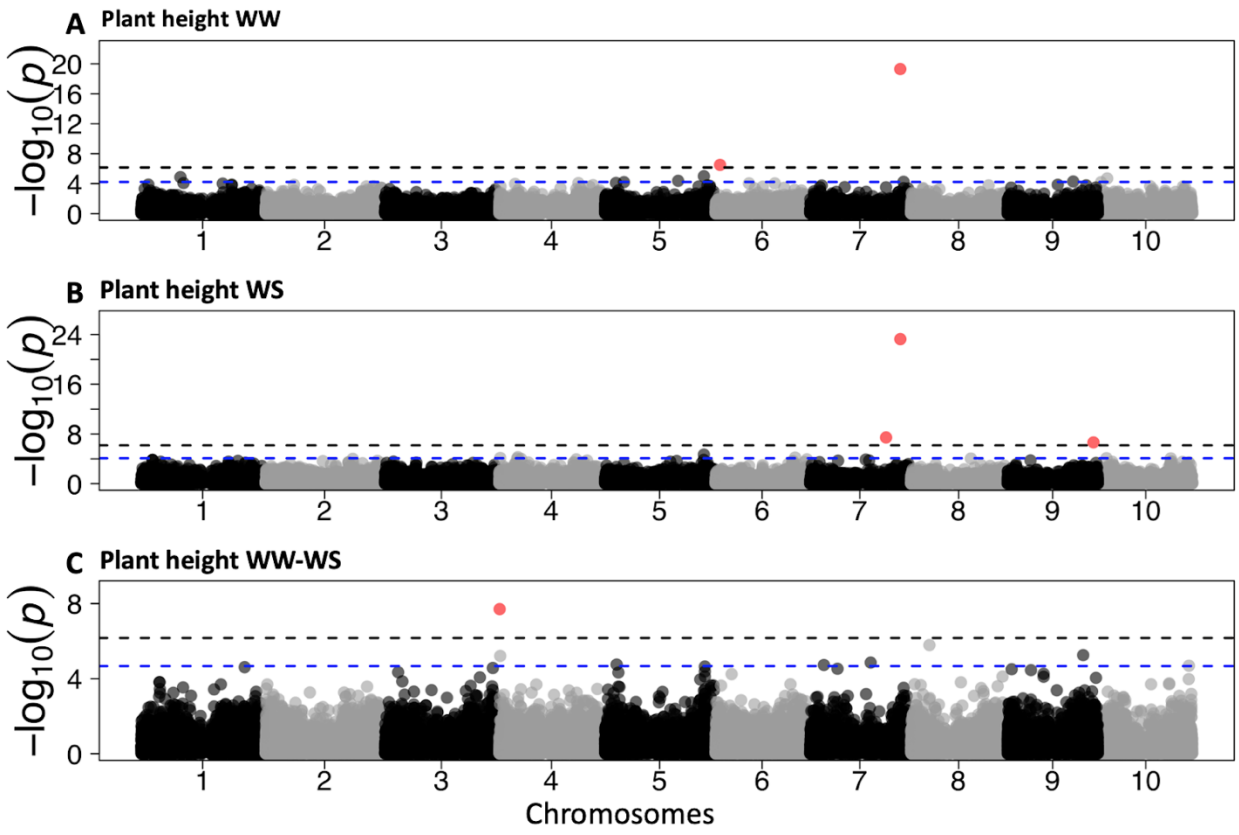


Figure 3.7. Genome-wide association study using multilocus mixed linear model (MLMM) for plant height.

In well-watered treatment (A), in water-stressed treatment (B), and differences in well-watered and water-stressed treatments (C). Black horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$). Blue horizontal dashed lines represent 0.01% lowest p-values.

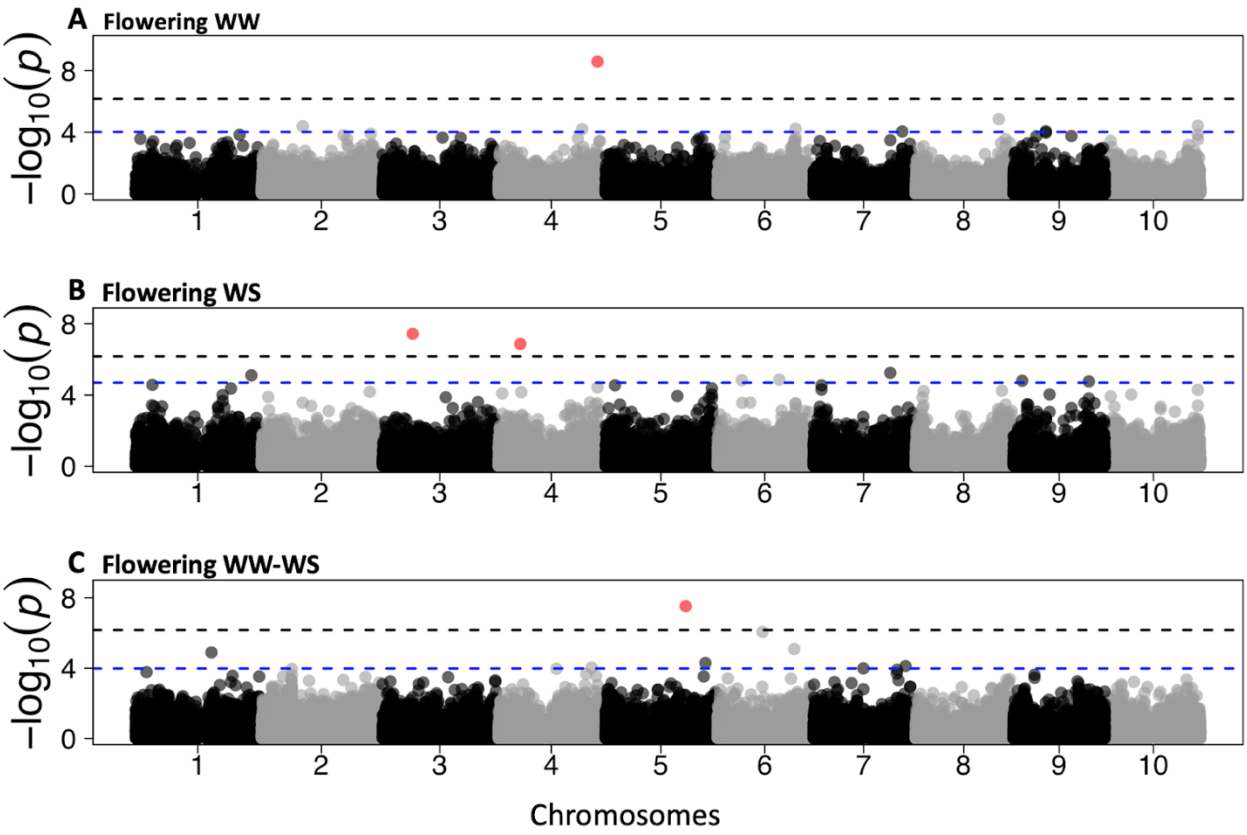


Figure 3.8. Genome-wide association study using multilocus mixed linear model (MLMM) for days to flowering.

In well-watered treatment (A), in water-stressed treatment (B), and differences in well-watered and water-stressed treatments (C). Black horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$). Blue horizontal dashed lines represent 0.01% lowest p-value

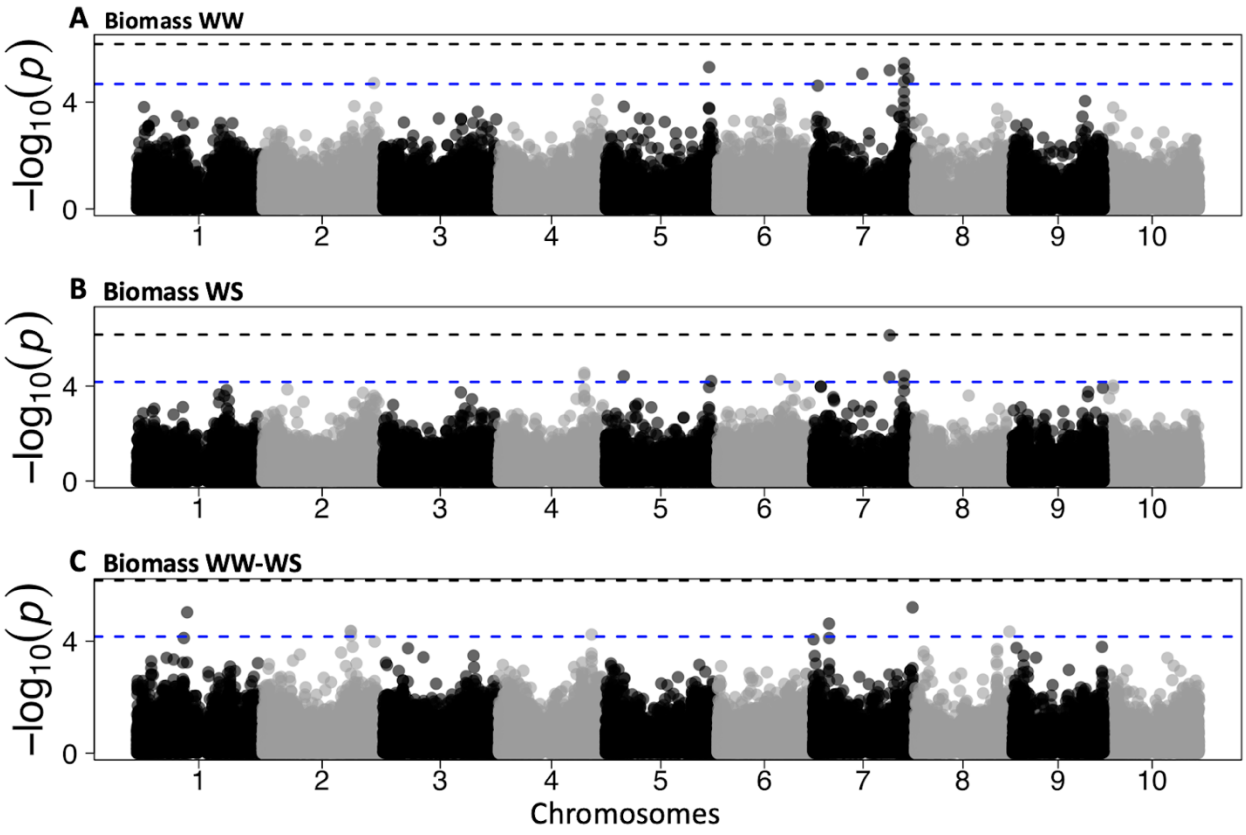


Figure 3.9. Genome-wide association study using multilocus mixed linear model (MLMM) for vegetative biomass.

In well-watered treatment (A), in water-stressed treatment (B), and differences in well-watered and water-stressed treatments (C). Black horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$). Blue horizontal dashed lines represent 0.01% lowest p-values.

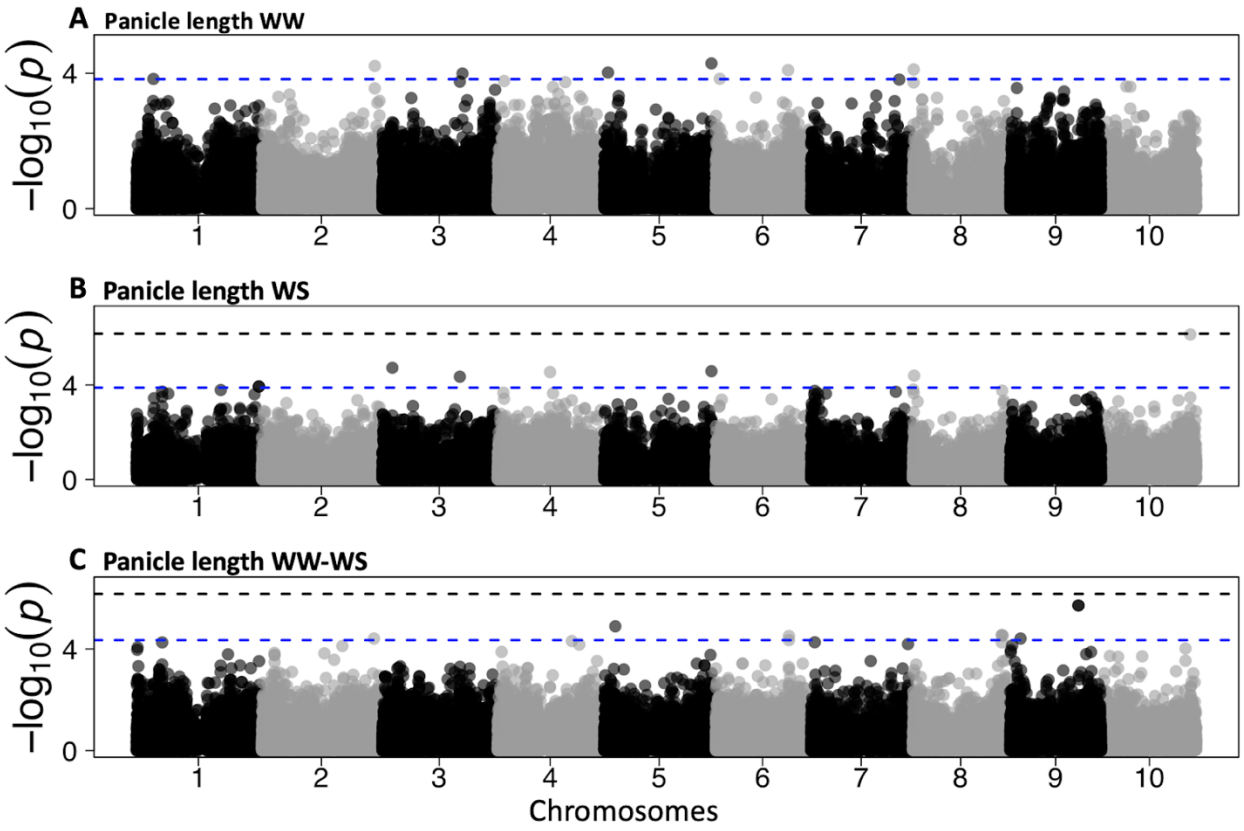


Figure 3.10. Genome-wide association study using multilocus mixed linear model (MLMM) for panicle length.

In well-watered treatment (A), in water-stressed treatment (B), and differences in well-watered and water-stressed treatments (C). Black horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$). Blue horizontal dashed lines represent 0.01% lowest p-values.

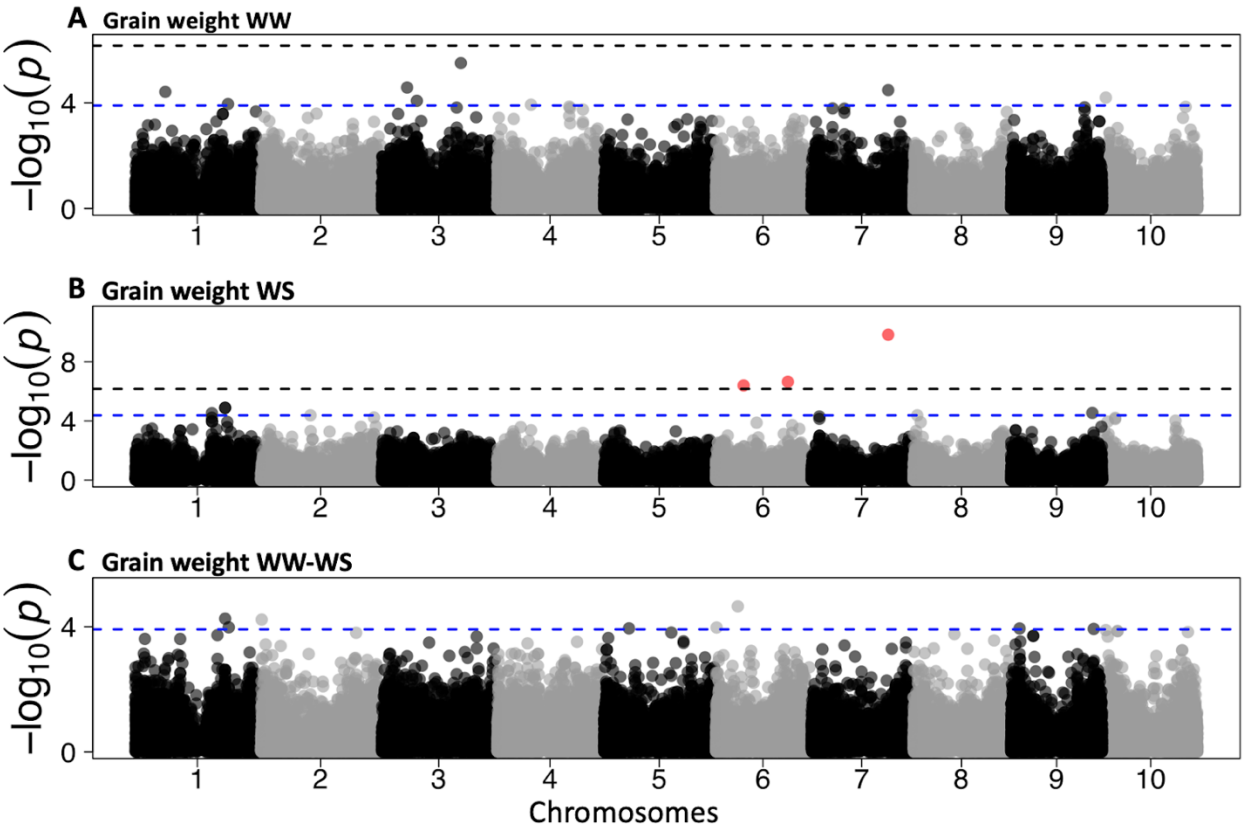


Figure 3.11. Genome-wide association study using multilocus mixed linear model (MLMM) for grain weight.

In well-watered treatment (A), in water-stressed treatment (B), and differences in well-watered and water-stressed treatments (C). Black horizontal dashed lines represent the threshold at Bonferroni correction ($\alpha = 0.05$). Blue horizontal dashed lines represent 0.01% lowest p-values.

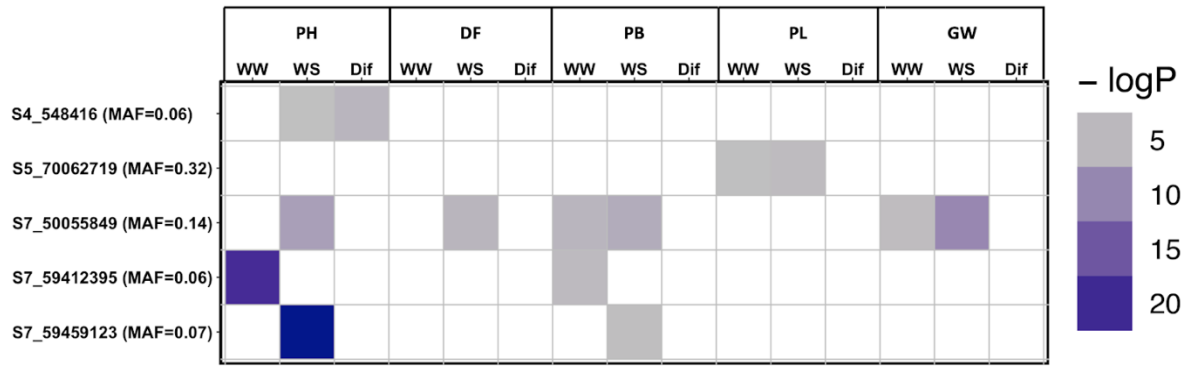


Figure 3.12. Heatmap of the -log(P-values) of shared associated SNPs in the study.

Traits: DF_WS: BLUPs for days to flowering in WS; Dif_WW_WS_PH: plant height difference between WW and WS; GW_WW: BLUPs for grain weight in WW; PB_WS: BLUPs for vegetative biomass in WS; PB_WW: BLUPs for vegetative biomass in WS; PH_WS: BLUPs for plant height in WS; PL_WS: BLUPs for panicle length in WS; PL_WW: BLUPs for panicle length in WW.

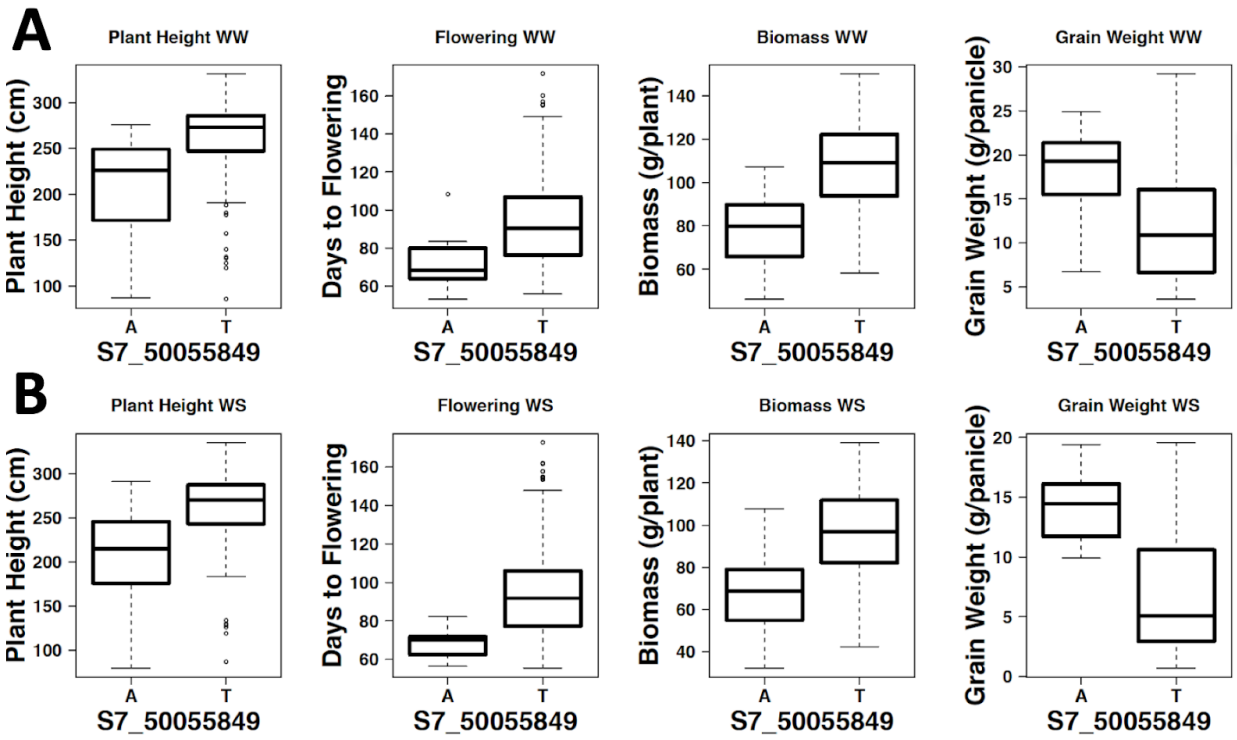


Figure 3.13. Variation of shared associated SNP S7_50055849.

For plant height, flowering, vegetative biomass, and grain weight in well-watered (A) and water-stressed (B) treatments. WW, well-watered, WS, water-stressed. A and T represent the two alleles at S7_50055849.

Table 3.1. Country of origin and botanical type of the panel (WASAP_Lysi).

Country	Number	Botanical Type[a]								
		B	C	D	G	GC	DC	DG	DB	U
Mali	31	3	-	-	3	-	-	1	1	23
Niger	116	12	-	5	4	21	54	1	3	12
Senegal	37	-	-	2	29	5	-	1	-	-
Togo	28	-	-	-	-	3	-	1	-	24
Other	7	-	-	-	-	4	-	-	-	3

[a]Botanical type key: B, bicolor; C, caudatum; D, durra; G, guinea; GC, guinea-caudatum; DC, durra-caudatum; DG, durra-guinea; DB, durra-bicolor; U, unknown. Other represents international breeding lines (e.g. BTx642, Tx7000, Macia, Sureno).

Table 3.2. Summary of ANOVA for pre-flowering drought tolerant (Tx7000) and susceptible (BTx642) genotypes.

Source of variation	Df	Mean Square	P-Value
Treatment	1	129	0.07 (.)
Genotype	1	573	3e-4 (***)
Year	1	0.2	0.94 (ns)
Repetition	1	0	0.99 (ns)
Genotype x Treatment	1	230	0.019 (*)
Treatment x Year	1	0.8	0.89 (ns)
Genotype x Treatment x Repetition	1	24.51	0.43 (ns)
Residuals	65	40	

Df: Degree of freedom; ns: not significant

*, **, *** Significance at $p < 0.05$, 0.01, and 0.001, respectively.

Table 3.3. Estimates of variance components of BLUPs and heritability of five traits in well-watered, water-stressed, and their differences across years (2017-2018).

Trait [a]	Well-watered (WW)				Water-stressed (WS)				WW-WS	
	Mean/sd	G	G x Y	H^2	Mean/sd	G	G x Y	H^2	G	H^2
PH	255/46.2	47.6 (***)	25 (ns)	0.96	252/45.3	48.7 (***)	7.5 (***)	0.96	0 (ns)	0.05
FB	105/22.4	26.9 (***)	16 (***)	0.77	91.7/22.7	26.1 (***)	14.1 (***)	0.83	12.3 (ns)	0.86
DF	91/23.3	24 (***)	1	0.96	92/24.2	25 (***)	3 (.)	0.97	40 (ns)	0.09
PL	24.07/5	30.2 (***)	0	0.93	24.7/5.4	34.7 (***)	2.5 (*)	0.91	7.8 (ns)	0.06
GW	12.6/6.1	54.8 (***)	16.6 (*)	0.75	8.14/5.4	40.4 (***)	7 (*)	0.79	6.7 (ns)	0.50

[a]Trait key: PH: Plant height; PB: Vegetative biomass; DF: Days to flowering; PL: Panicle Length; GW: Grain weight.

*, **, *** Significance at $p < 0.05$, 0.01, and 0.001, respectively; ns: not significant.

Mean: average value of the trait; sd: standard deviation of the trait; G: Genotype; G x Y: Genotype-by-Year interaction; H^2: Broad sense heritability.

Table 3.4. Summary of significant SNPs associated in the MLMM model with the traits above the Bonferroni threshold (α = 0.05).

Traits	Treatment[a]	SNP ID	-log(P-value)	MAF
Flowering	WW	S4_64398335	8.6	0.07
Flowering	Difference (WW-WS)	S5_52090779	7.5	0.07
Flowering	WS	S3_19589652	7.4	0.13
Flowering	WS	S4_14172212	6.9	0.08
Grain Weight	WS	S7_50055849	9.8	0.13
Grain Weight	WS	S6_46923493	6.7	0.05
Grain Weight	WS	S6_18075344	6.4	0.18
Plant Height	WS	S7_59459123	23.3	0.06
Plant Height	WW	S7_59412395	19.3	0.05
Plant Height	Difference (WW-WS)	S4_149131	7.7	0.04
Plant Height	WS	S7_50055849	7.4	0.13
Plant Height	WS	S9_56534065	6.6	0.07
Plant Height	WW	S6_2900751	6.5	0.09

[a]WW: well-watered; WS: water-stressed.

Chapter 4 - Genomics-enabled breeding of *Striga* resistant sorghum varieties for West Africa

Abstract

Striga hermonthica is a devastating obligate parasite of staple crops in West Africa (WA). This parasite requires signals from the host for its germination. One of the mechanisms of *Striga* resistance is low germination stimulant production by the host due to loss of function of the *LGS1* gene. The development of resistant varieties through phenotypic selection is challenged by *Striga*'s spatial variability. Diagnostic markers targeting *LGS1* loss of function alleles could accelerate the development of *Striga*-resistant sorghum. The deletion of the cloned *LGS1* is hypothesized to confer resistance in WA *Striga* using breeder-friendly markers. The objectives of this study are to (i) identify single nucleotide polymorphisms (SNPs) that are linked and in linkage disequilibrium (LD) with *LGS1* deletion in diverse sorghum germplasm, (ii) develop Kompetitive Allele-Specific PCR (KASP) markers that accurately genotype these SNPs, and (iii) test the markers for *Striga* resistance in sorghum biparental population. Genotyping-by-sequencing of WA germplasm was analyzed to identify SNPs in LD with deletion alleles at *LGS1*. Eight SNPs that are linked and in LD with *LGS1* deletion were converted into KASP markers. Putative resistance and susceptible classes in inbred lines and F_3 progeny of SRN39 (resistant) and Mota Maradi (susceptible) were differentiated through KASP genotyping. Root exudates from progenies with putative deletion of *LGS1* based on KASP assay were tested for laboratory germination assay of *Striga* seeds. Our findings suggest that these new KASP markers will be suitable for marker-assisted selection for *Striga* resistance in multiple genetic backgrounds of WA germplasm.

Introduction

Multiple abiotic and biotic stressors threaten agricultural productivity in semi-arid regions. Among the biotic limiting factors, *Striga* (witchweed) damages most sorghum production in sub-Saharan Africa. *Striga hermonthica*, an obligate hemiparasite, infests sorghum, millet, and corn, causing up to 50% yield losses in sorghum in this region (Haussmann et al., 2000; Ejeta, 2007). The mechanism of *Striga* infestation is complex in regions affected by drought and low soil fertility (Scholes and Press, 2008; Yoder and Scholes, 2010). *Striga* seeds require metabolites to induce germination in the soil before establishing haustorial attachment for nutrient uptake from the host (Joel, 2013; Kountche et al., 2017). Strigolactones, metabolites produced by the host plant, induce *Striga* germination, which will attach to the host through haustorial initiation and attachment (Ejeta, 2007; Scholes and Press, 2008; Bandaranayake and Yoder, 2013). Strigolactones are secondary metabolites that promote the formation of lateral roots and root hairs, along with mycorrhizal association with the plant host (López-Ráez et al., 2017). *Striga* resistance due to germination stimulants has been identified in SRN39 through the high level of strigolactones compound orobanchol (which does not induce germination) rather than 5-deoxystrigol (Gobena et al., 2017).

In semi-arid regions of West Africa, developing *Striga* resistant varieties is among the major priorities of breeding programs. These product developments were challenging with limited resources and capacity in phenotyping for resistance to local environments (Olembo et al., 2010; Atlin et al., 2017). Traditional breeding methods were successful, but often phenotypic evaluation of the obligate parasite is complex. Phenotyping on-station does not always represent the targeted population environment (Atlin et al., 2017), and there is spatial variability of *Striga* seeds in farmer's fields, leading to differential responses (Kapran et al., 2007). *Striga* resistant

varieties (SRN39, N13, and Framida) that have been developed are not adapted to several environments in Africa, characterized by diverse agroclimatic zones, limiting their adoption by farmers (Haussmann et al., 2000; Ejeta, 2007; Kapran et al., 2007; Walker and Alwang, 2015; Yohannes et al., 2015). Future product developments might focus on locally adapted sorghum varieties by introgressing only the trait conferring *Striga* resistance.

Advances in genomics have increased our understanding of staple crops and their interaction with biotic stressors (Collard and Mackill, 2008). Genotyping approaches have allowed the utilization of molecular markers to establish sorghum improvement for *Striga* resistance in developing countries (Haussmann et al., 2004; Kapran et al., 2007; Deshpande et al., 2013; Gobena et al., 2017). For instance, marker-assisted selection through SSR markers was used to develop *Striga* resistant varieties in Eritrea with N13 as the donor parent; however, only a few background markers were recovered (Yohannes et al., 2015). Furthermore, studies have identified genes responsible for low germination stimulants (*LGS1*) in sorghum due to loss of function of *LGS1* in the resistant varieties (Gobena et al., 2017) detectable with PCR based markers. Although PCR based markers are used in multiple breeding programs, their implementation in small breeding programs is often challenging with limited access to genotyping and cost-prohibitive laboratories.

Outsourced genotyping methods are now convenient to rapidly evaluate large datasets at reduced cost and integrate molecular breeding approaches using SNP genotyping (Thomson, 2014). Thus, to efficiently develop farmer-preferred varieties, it would be valuable to take advantage of resources and genomic advancements already publicly available. For the development of *Striga* resistant varieties, genetic variants linked and in linkage disequilibrium with the loss of function of *LGS1* could be converted into Kompetitive Allele-Specific PCR

(KASP) markers for outsourced cost-effective genotyping. KASP genotyping uses two allele-specific primers designed with respect to the polymorphic allele at one locus, and one common primer (He et al., 2014). Along with the sequencing data of West African sorghum germplasm, those genomic resources will be useful in identifying markers and the assessment of genomics-enabled breeding for *Striga* resistance.

This study focused on the identification of genomic variants associated with *Striga* resistance and the development of KASP markers to enable breeding in West Africa with rapid and cost-effective high throughput genotyping. To date, there are no diagnostic KASP markers for *Striga* resistance delivered to breeders to precisely introgress the resistance into locally adapted sorghum varieties in West Africa. We hypothesized that the deletion of *LGS1* locus confers *Striga* resistance via a low germination stimulant. The objectives of this study are to (i) identify markers associated with the deletion of *LGS1* in diverse sorghum germplasm, (ii) develop KASP markers that tag the deletion of *LGS1* using SNPs derived from sequencing data, and (iii) test the markers for *Striga* resistance in sorghum biparental population.

Material and Methods

Plant material

Genotyping-by-Sequencing (GBS) data from the sorghum association panel (Morris et al., 2013), West African sorghum germplasm from GRIN (Lasky et al., 2015; Maina et al., 2018; Olatoye et al., 2018; Faye et al., 2019), and the West African Sorghum Association Panel (Faye, Maina et al., *in prep*) were analyzed. Mapping populations were developed with SRN39 (PI656027) as the donor parent while recurrent parents used were Mota Maradi (PI656050),

MR732 (PI656051), and Sepon82 (PI656024) from the sorghum association panel, also released varieties in Niger.

PCR based assay

Genomic DNA of BTx623, Mota Maradi, El Mota, and SRN39 was extracted following the CTAB method and quantified using the nanodrop. After normalization to 10 ng/ul, 2ul of each sample was used for the PCR reaction. Two primer sequences were obtained from Gobena et al., (2017) (Figure 4.1). The first primer (G2133) has a predicted amplicon size of 351 bp on all genotypes assayed, while the second primer (PDstrigalgs5b) has a predicted amplicon size of 628 bp for only susceptible genotypes without the deletion of *LGS1*. The reaction volume was composed of 6.5 ul of master mix (NEB M0270L), 3.5 ul of water, and 1 ul of each of the forward and reverse primers. The PCR conditions were 94 °C for 15 seconds, 62 °C for 30 seconds, and 72 °C for 30 seconds each cycle with a total of 40 cycles. Amplicons were visualized by electrophoresis on 1.5% agarose gel.

Genomic analysis

The reference genome BTx623 version 3.1 (Paterson et al., 2009; McCormick et al., 2018) was used to align sequencing data using Burrows-Wheeler Alignment (Li and Durbin, 2010) and to discover single nucleotide polymorphisms (SNPs) in TASSEL 5 GBS pipeline (Bradbury et al., 2007). To test whether sequencing data detect misalignment of *LGS1* region in some genotypes due to the deletion, non-imputed data on chromosome 5, aligned with the reference genome BTx623, were analyzed. *LGS1* region on chromosome 5 (69.97 Mb – 70.01 Mb) was selected with flanking regions (Figure 4.1). The selected regions (*LGS1* region + flanking regions) were hereafter referred to as "extended *LGS1*". The predicted percentage of missing data for the resistant line, SRN39 will be 100% within the *LGS1* region. A frequency of

more than 80% missing data within the region was considered to account for polymorphism present in the flanking regions not related to the deletion. The extent of linkage disequilibrium between the deletion and flanking SNPs was analyzed using R^2 and *D'* to account for the allele frequency and co-inheritance between the SNPs and the deletion of *LGS1*. Pairwise estimates were calculated among the SNPs within the *LGS1* region in Tassel 5 (Glaubitz et al., 2014). Allele frequencies were calculated using VCFtools version 0.1.13 (Danecek et al., 2011).

KASP marker design and genotyping

SNPs in LD with the deletion were selected to develop KASP markers. Only SNPs that distinguish between the resistance line, SRN39, and the susceptible lines (e.g. BTx623) were selected to design KASP primers and SNP verification. A total of 21 SNPs were used to design primers with 50 bp flanking sequences around each SNP obtained from https://phytozome.jgi.doe.gov for *Sorghum bicolor* version 3 based on the physical position (Table 4.1, Table C.1) ranged between 11 to 823 kb upstream and downstream of the *LGS1* deletion region. Note, SNPs outside the region of interest (Figure 4.1) were included for the KASP marker development as they differ between BTx623 and SRN39.

After the SNP verification, selected markers were tested in inbred lines and progenies derived from resistant and susceptible lines to determine how the markers would classify individuals based on *a priori* predictions on the parents. Within a 96 well plate, eight biological replicates were used for BTx623 and SRN39 each. Other inbred lines and progenies were replicated four and two times in the plate. Artificial heterozygotes, a mix of leaf tissue between one susceptible line and one resistant line, were also included in the design.

Tissue collection was performed using the established protocol in Morris Lab. Genotypes were grown in the greenhouse for 14 days. Each genotype has at least three individuals referred

to as biological replicates. Leaf tissues were collected using a hole punch. Two leaf punches for each sample were placed into the respective well of a 96-deep-well plate following the plate design. For the artificial heterozygotes, one leaf punch of the two samples was used. The last two wells (H11 and H12) were left empty as KASP controls. The plate was then placed in a bag containing silica beds overnight. Dried leaf tissues were sent for DNA extraction and KASP genotyping at Intertek AgriTech (Alnarp, Sweden).

Another genotyping test, composed of inbred lines and mapping populations distributed in four plates, was carried out. In total, 376 samples (biological replicates and technical replicates) and eight blanks were assayed. The design is composed of highly replicated genotype (34 samples for BTx623 and SRN39), moderately replicated genotypes (14 samples for each genotype), low replicated genotypes (five and three samples for each genotype), and artificial heterozygotes (Table C.3). Each plate is composed of three blank wells. Genotypes were distributed across the four plates to reduce errors related to tissue collection. The design is randomized on each plate.

For the mapping populations of SRN39 x Mota Maradi, MR72, and Sepon82 (F2 progenies) were grown in the greenhouse in single pots for each biological replicate; leaf tissues were collected 14 days after planting for all biological replicates. The F3 progenies of SRN39 x Mota Maradi were grown in the field from June to October 2019 in Manhattan Kansas. A total of 50 seeds for each progeny is sown within the plot. In five random plots, two plants were selected for genotyping as biological replicates within the plot. For the remaining plots, a single plant was genotyped.

Striga germination stimulant assay

The germination assay was conducted in the quarantine lab in the department of biology at Pennsylvania State University following the protocol used in Bellis et al., (2020).

Striga seeds preconditioning

Striga hermonthica seeds, collected in Siby, Mali (12°23′ N; 8°20′ W) provided by the Malian breeding program (Bellis et al., 2020), were used for the germination assay. Seeds were first sterilized with 0.5% sodium hypochlorite for 30 seconds and rinsed three times with sterile distilled water. On average, 75 seeds were transferred into a 12 well-plate with 500 µl of deionized water (DI) in each well. Seeds were preconditioned in the dark for 10 days at 30℃ before root exudates application.

Sorghum growth and root exudates extraction

Progenies derived from SRN39 and Mota Maradi were used for the assay. Based on the KASP genotyping, five predicted resistant and five predicted susceptible lines were selected, and the parents were selected. Sorghum seeds were grown in sand/calcined clay mix (1:1 sand/calcined clay mix) for four weeks in the greenhouse. For each line, six biological replicates were grown. Plants were watered daily with fertigate (top-watering) with ¼ strength Miracle-Gro (Miracle-Gro® Water Soluble All Purpose Plant Food, The Scotts Company, LLC. Marysville, OH). Four weeks after planting, each plant (root and shoot) was removed from the pot carefully. Roots were washed with deionized (DI) water to remove remaining potting debris. The plants were then placed in a flask containing DI water using a 1:5 ratio of root:DI water, sealed with parafilm. Flasks were placed in darkness at room temperature. After 48 hours, the solution was centrifuged at 4˚C at 9,000 rpm for 10-15 minutes, and the supernatant was placed in ice to prevent exudate's degradation.

Sorghum root exudates application

After *Striga*'s preconditioning, a volume of 1.5 ml root exudates was applied to each well. Three technical replicates were used for each biological replicate. GR24 (at 0.1ppm), the synthetic strigolactone and DI water were used as positive and negative controls respectively. After 72 hours, germinated seeds were counted under a stereomicroscope.

Results

Genomic analysis

To validate the presence of a deletion on chromosome 5 in the resistant variety SRN39 that was used in further analysis, previously developed primers were assayed (Figure 4.2). The first set of primers (G1233) produced amplicons of ~ 350 bp for all the lines evaluated which is as predicted. In contrast, the second set of primer (PDstrigalgs5b) produced amplicons of ~ 600 bp for only BTx623, Mota Maradi, and El Mota. None of the four samples of SRN39 were amplified for the primer PDstrigalgs5b.

To test whether the deletion of *LGS1* is captured in the GBS data, the region of 570 kb which includes the *LGS1* deletion range and flanking regions on both sides were selected (Figure 4.1, Figure 4.3). We predicted that for the resistant lines the *LGS1* region will have 100% missing data, while the flanking region will have less than 80% missing data under the prediction that there are no mutations in the flanking regions. Among the tested genotypes, three genotypes that have deletion within the selected *LGS1* region, 54.K.94 (PI533752), SRN39 (PI656027), and Framiola (PI533976) were identified, known as *Striga* resistant genotypes. The minor allele frequency around the causative gene Sobic.005G213600 is within the range of 0.01 to 0.14 in the collections (Table C.2).

A few genotypes carrying the deletion of *LGS1* will be found in *Striga* prone regions of West Africa. Using georeferenced sorghum germplasm, accessions with missing SNPs similar to the resistant genotype SRN39 across West Africa (Figure 4.3A) were observed. The percentages of occurrence of the hypothetical presence of *LGS1* deletion are 7%, 7%, and 9% over the total number of sequenced genotypes for Niger, Nigeria, and Senegal, respectively. Based on the empirical distribution of the missing SNPs (Figure C.3), the false discovery rate is 35% at p-value $<10^{-3}$. For the genetic relatedness, accessions with predicted deletion of *LGS1* were not clustered in one group for both neighbor-joining trees for chromosome 5 and the *LGS1* regions selected (Figures 4.3B-C).

Identifying SNPs around *LGS1* region using GBS data

To identify SNPs linked and in LD with *LGS1* deletion on chromosome 5, upstream and downstream flanking regions (Figure 4.4) were further analyzed. We selected a total of eight SNPs that differ between known resistant from susceptible lines (Figure 4.4A, Table 4.1) according to the GBS data set. For each flanking SNP selected, we estimated pairwise estimates of LD within the deletion of the *LGS1* region (69.97–70.01 Mb) (Figure 4.4A, Table C.2). For instance, the distance between S5_69450954 (MAF = 0.42) and the *LGS1* deletion region is about 526 kb. Within this interval distance (69.97–70.01 Mb), which covers about 114 kb flanking region, R^2 was very low (0.008–0.07) for the 69 SNPs analyzed and minor allele frequencies were relatively low for the selected region (0.02–0.15) (Figure 4.4). Next, we considered *D′* values of 1 to account for only complete LD and analyzed 20 SNPs with the region.

KASP marker design: SNPs obtained from GBS data

We tested the KASP markers in mapping populations and diverse germplasm. Selected SNPs (Table C.1) were used to develop KASP-SNP assay and to identify genotypes differentiating for putative resistant and susceptible alleles controlling *Striga* responses. The physical distance between the deletion and the SNPs is quite large (Table 4.1), however, due to the number of missing SNPs and error rate in the GBS data, those SNPs were selected for further analysis. FAM and HEX alleles were then assigned at the SNP of interest for the reference genome (BTx623) and SRN39 respectively.

Test plate KASP assay

Based on the GBS data, predicted the alleles on each marker for the inbred lines tested were not consistent for hypothetical susceptible and resistant lines (Table 4.2, Figures C.1-2). Observed alleles for each genotype were based on the calls obtained from Intertek results according to the FAM and HEX fluorescences. Some markers differ between SRN39 and susceptible lines, while other markers differ within susceptible lines (Figure 4.5). Among the tested markers, the marker snpSB00246, upstream of the deletion, which only differentiated predicted resistant groups from susceptible groups (Figure 4.6, Table 4.2) was further investigated. For the predicted susceptible inbred (BTx623, El Mota, Mota Maradi, IRAT204, MR732, and Sepon82), genotype calls (G:G) were consistent with the expected allele for all biological replicates of each line.

In addition, we tested three Nigerien genotypes from GRIN (PI513676, PI513680, and PI513740) having the local name "Matche da Koumya", a locally preferred variety and parent of mapping population with SRN39 (Table 4.2). All genotypes were called G:G (homozygote for the G allele), as predicted except for one individual (called A:A) out of twelve tested. For the

alternative call A:A (homozygote for the A allele), genotypes PI277541 (short Framida), PI284980 (Framida), and PI465448 (Framida) have the A:A genotype in all of the individuals. Similarly in SRN39 obtained from two different seed sources, the majority of the individuals out of 16 tested were called A:A except for two individuals. For the artificial heterozygotes (a mix of predicted homozygotes at the G allele and SRN39), we observed 11 heterozygotes samples A:G out of 12 samples analyzed.

Genotyping results of snpSB00246 using inbred lines and mapping populations

We first checked the genotypes of inbred lines on marker snpSB00246 (G or A) (Figure 4.7). All of the 34 samples of BTx623 (including biological and technical replicates) are homozygotes for the G allele. Similar observations were made for Mota Maradi, MR732, and Sepon82, homozygotes for the G allele. In contrast, predicted resistant lines will have the alternative allele (A) on snpSB00246. For SRN39, 13 samples have the A allele while seven are heterozygotes at the marker (Table C.3). We found that most of the putative susceptible genotypes tested have a stable allelic state on snpSB00246 than the putative resistant genotypes on the alternative allele. Next, we tested F_2 and F_3 progenies of SRN39 by Mota Maradi and MR732. Homozygotes for the G allele in the F_2 have the same allele G in the next generation. Heterozygotes (A:G) progenies at F_2 were all homozygotes at the G allele in the next generation, while the homozygotes at the A allele were not consistent in the F_3 progenies with uncallable and unused biological replicates (Table 4.3). The deviation from the expected ratio (37.5% G:G, 25% A:G, 37.5% A:A) might be due to the leaf tissue collected from one or two plants within a plot in the field. To test that there is no statistical difference between the observed and the expected number of progenies in each class, most of the F_3 plants within the plot should be genotyped since the progenies derived from a single F_2 plant.

Genotyping results of snpSB00246 for mapping population SRN39 x Mota Maradi

To test whether the markers segregate in mapping populations, we genotyped F_3 progenies of SRN39 by Mota Maradi. A total of 71 samples, which include 56 F_3 progenies from each plot in the field and 15 parents were genotyped. We observed groups of F_3 progenies clustered with either Mota Maradi or SRN39 (Figure 4.8). Genotyping calls identified 26 progenies homozygotes for the A allele and 27 progenies homozygotes for the G allele. In four of the progenies, their HEX and FAM fluorescences were not sufficient to classify into the groups labeled as uncallable. Based on these results, we hypothesized that F_3 progenies clustered with Mota Maradi are susceptible and progenies clustered with SRN39 are resistant to *Striga*.

Germination assay

Under the hypothesis that KASP marker tag loci that control low germination stimulant in predicted resistant lines, we predicted a significantly higher number of germinated *Striga* seeds in the homozygotes for the G allele compared to the homozygotes A allele. On average 60 *Striga* seeds were germinated for the synthetic strigolactone GR24, the positive control, while no germination was observed in the DI water, the negative control (Figure 4.9). Next, we predicted that the susceptible parent Mota Maradi will have a similar germination rate as GR24. However, this was not the case since only three to five *Striga* seeds germinated. Root exudates of progenies from SRN39 x Mota Maradi did not induce a higher number of germinated seeds after 72 hours. In a few samples of predicted susceptible, we observed 1 to 3 germinated seeds. None of the predicted resistant varieties induced germination in the preconditioned *Striga* seeds.

Discussion

Development of KASP markers for diverse germplasm requires multiple testing and replicated genotypes

The use of multiple replicates for the same genotype is of significance to determine precisely the stability of the alleles in the germplasm for an optimized KASP assay. In our study, we tested multiple ($n \leq 8$) biological replicates (individual plants from the same accession) and technical replicates (two samples from the same individual) in the test plate for SNP verification (Figure 4.6). However, for initial screenings in the breeding program, a few replicates ($n = 1 – 4$) may be necessary to test the hypothesis of trait introgression into the background of the variety of interest to reduce possible experimental errors during tissue collection, contaminants during shipment, or low DNA concentration in some samples. For routine analysis, one to two replicates might be sufficient to test the hypothesis that the individuals have the similar genotypes calls. This procedure is reliable to test the hypotheses on the uniformity of the seeds and the stability of the marker in multiple germplasm (Figure 4.7, Table C.2). When using multiple replicates of the same individual, unexpected genotype calls can be negligible, in some cases, if more than 80% of the replicates carry the expected allele previously tested (Table 4.3).

The SNP quality assessment defined by Intertek has five classes (Very good, good, medium, bad, and inconclusive) with regards to the amplifications, the number of clusters, and the distribution of clusters. Of the 19 markers verified (Table C.1), eight of them were identified as medium to good markers for further analysis (Table 4.1). The consideration of BTx623 reference alignment might lead to the observation of "Inconclusive" or "Bad" (based on Intertek SNP verification analysis) due to the monomorphic status at the SNP which had not been detected in the GBS sequencing (Table C.1), thus a single cluster may be observed.

Presence of mutations around *LGS1* might not confer resistance

GBS often leads to incomplete data due to low coverage genotyping (Poland et al., 2012; Fu, 2014; Glaubitz et al., 2014), and referred to as missing SNPs. Genotype information in the accessions could be recovered using imputation methods based on linkage disequilibrium (Browning and Browning, 2009; Hickey et al., 2013). However, non-imputed data were used to identify polymorphisms at the region of interest hence a large number of missing SNPs in the sequences in the germplasm evaluated (Figure 4.4A-C, Table C.2), but they were sufficient to identify SNPs in LD (Davey et al., 2011). The high number of missing SNPs for the majority of the accessions within the extended region does not always infer the loss of function of *LGS1* due to the depth of coverage (Figure C.3). Given the high infestation of *Striga* in Africa over generations and its diverse mechanisms of resistance (Ejeta, 2007; Scholes and Press, 2008; Bellis et al., 2020), few varieties might be adapted to the environment and alleles controlling resistance could be present. Similar to our findings on the putative presence of *LGS1* deletion in West Africa (Figure 4.3) based on the sequencing data, the resistant accessions were also found in *Striga* infested regions of Africa (Bellis et al., 2020).

SNPs (n = 5) within the *LGS1* causative gene (Sobic.005G213600) have low to moderate allele frequencies (MAF between 0.02 to 0.14) suggesting that most of the genotypes may not carry the deletion of the gene and the limited number of resistant accessions in the collections carrying a deletion of *LGS1*. The correlations R2 between flanking SNPs and the *LGS1* region were low (Table C.2) possibly due to the differences in allele frequencies observed.

Deletion of *LGS1* segregating progenies could confer resistance in diverse *Striga* ecotypes

The material selected for the bi-parental mapping population and backcrosses originated from Niger (e.g. Mota Maradi) or released in Niger (e.g. Sepon82 and MR732) (Table 4.3) that do not have the deletion of *LGS1* (Figure 4.2) thus are putative susceptible genotypes. Those genotypes are present in the sorghum association panel (Casa et al., 2008) which makes population development with SRN39 as a donor parent possible in the United States. With two generations per year and marker selection, variety development will be much faster for participatory selection in targeted environments. Note, we conducted only generation advancements and backcrosses with the recurrent parent (SRN39) and screening for low germination inducing *Striga* assays are performed in the quarantine facility to identify putative resistant or susceptible progenies. We hypothesized that predicted resistant progenies derived from those families will demonstrate advantageous performance under *Striga* infested fields of West Africa and reduce long term generation advancement and reduced phenotyping cost.

The marker snpSB00246, upstream of the deletion, was only consistent among, known, putative susceptible and resistant genotypes (Figures 4.6 - 4.7, Table 4.1). In contrast, no downstream markers of our tested SNPs differ between putative susceptible and resistant genotypes (Table C.1, Figures C.1-2). Since SNPs were derived from the reduced representation library, some more tightly linked variants might be detected when analyzing whole-genome sequences from contrasting the genotypes and parental lines. Using GBS data, 0.05% (number of SNPs = 325) was captured in the "extended *LGS1* region" of 570 kb; in consideration with LD (~150 kb) about 0.02% is sufficient to identify markers. However, with variable MAF in this

region especially the low MAF within the "*LGS1* region", it might be possible to identify more variants with whole-genome sequences linked and in complete LD with the "*LGS1* region".

In the progenies $F_{2:3}$ of SRN39GMS × Mota Maradi (n = 56), the number of genotypes calls were 27 and 26 for G:G and A:A calls respectively. $F_{2:3}$ progenies with A:G calls were not observed. The deviation from the expected ratio (37.5% G:G, 25% A:G, 37.5% A:A) might be due to the leaf tissue collected from one or two plants within a plot in the field. To test that there is no statistical difference between the observed and the expected number of progenies in each class, most of the $F_{2:3}$ plants within the plot should be genotyped since the progenies derived from a single F_2 plant.

Findings in this study were based on SNPs in linkage with the deletion of *LGS1* previously cloned with the assumption that the deletion may confer *Striga* resistance in multiple genetic backgrounds and environments (Table C.2). PCR-based assays (Figure 4.2) could be used to test the hypothesis that *LGS1* deletions in the progenies selected based on the KASP genotyping results. This hypothesis might be consistent with multiple recurrent parent backgrounds of West African sorghum. In addition, genotype-by-environment interactions are more likely to occur due to the adaptation of the host and the parasite in multiple environments as observed in multi-location phenotyping of sorghum and millet under *Striga* infested fields (Haussmann et al., 2001; Kountche et al., 2013). In this study, we focused on one mechanism of *Striga* resistance, the low germination stimulant production with SRN39. Thus, targeting multiple sources of resistance might prevent yield losses due to *Striga* in West Africa. Moreover, field phenotyping and participatory selection in two contrasting environments (i.e. control and *Striga* infested fields) might assist further *Striga* marker and variety development. Once validated, breeder-friendly SNP *Striga* markers could facilitate breeding for the selection of

resistant varieties. Nevertheless, previous studies reported that environments might affect *Striga* resistance (Haussmann et al., 2004; Scholes and Press, 2008). In addition to the low germination stimulant resistant varieties, integrated controls of witchweed will be considered to efficiently reduce parasitic weed seed banks in smallholder farmers' fields. The advantage of understanding the life cycle of *Striga* provides control of resistance at each stage (Scholes and Press, 2008; Yoshida Satoko and Shirasu Ken, 2009).

Leveraging genomics to implement molecular breeding

Marker-assisted selection is widely used to identify causal variants for specific traits using both low and high-density markers (Singh and Singh, 2015). Markers provide advantages to accelerate plant breeding through the efficiency of the selection without phenotyping in multiple environments and has been a cost-effective genotyping method recently for monogenic and oligogenic traits of interest. This study describes the development of cost-effective KASP markers for *Striga* resistance in West Africa to facilitate breeding for *Striga* resistance. Using outsourcing genotyping, markers that tag the deletion of *LGS1* in the biparental population were developed (Table 4.2, Figure 4.8). The success of our genotyping relies on the ability to identify specific SNPs that tag the deletion of *LGS1* and the reduction of germinated *Striga* seeds in lab assays. As one to the resistance mechanism, low germination stimulant metabolites would reduce yield losses in *Striga* infested fields. Further experiments would be needed in fields to test the hypothesis that the putative resistant lines show less incidence of *Striga* infestation than the putative susceptible lines.

References

Atlin, G.N., J.E. Cairns, and B. Das. 2017. Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Global Food Security 12:31–37. doi:10.1016/j.gfs.2017.01.008

Bandaranayake, P.C.G., and J.I. Yoder. 2013. Haustorium Initiation and Early Development. Springer, Berlin, Heidelberg.

Bellis, E.S., E.A. Kelly, C.M. Lorts, H. Gao, V.L. DeLeo, G. Rouhan, A. Budden, G.B. Bhaskara, Z. Hu, R. Muscarella, M.P. Timko, B. Nebie, S.M. Runo, N.D. Chilcoat, T.E. Juenger, G.P. Morris, C.W. dePamphilis, and J.R. Lasky. 2020. Genomics of sorghum local adaptation to a parasitic plant. PNAS. doi:10.1073/pnas.1908707117

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Browning, B.L., and S.R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am J Hum Genet 84:210–223. doi:10.1016/j.ajhg.2009.01.005

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community Resources and Strategies for Association Mapping in Sorghum. Crop Science 48:30–40. doi:10.2135/cropsci2007.02.0080

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. doi:10.1098/rstb.2007.2170

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The Variant Call Format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. doi:10.1038/nrg3012

Deshpande, S.P., A. Mohamed, and C. Thomas Hash. 2013. Molecular Breeding for Striga Resistance in Sorghum. R.K. Varshney and R. Tuberosa, ed. John Wiley & Sons Ltd.

Ejeta, G. 2007. Breeding for Resistance in Sorghum: Exploitation of an Intricate Host–Parasite Biology. Crop Science 47:S-216-S-227. doi:10.2135/cropsci2007.04.0011IPBS

Faye, J.M., F. Maina, Z. Hu, D. Fonceka, N. Cisse, and G.P. Morris. 2019. Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. Ecology and Evolution 9:6038–6051. doi:10.1002/ece3.5187

Fu, Y.-B. 2014. Genetic Diversity Analysis of Highly Incomplete SNP Genotype Data with Imputations: An Empirical Assessment. G3 (Bethesda) 4:891–900. doi:10.1534/g3.114.010942

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One 9. doi:10.1371/journal.pone.0090346

Gobena, D., M. Shimels, P.J. Rich, C. Ruyter-Spira, H. Bouwmeester, S. Kanuganti, T. Mengiste, and G. Ejeta. 2017. Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. PNAS 201618965. doi:10.1073/pnas.1618965114

Haussmann, B.I.G., D.E. Hess, G.O. Omanya, R.T. Folkertsma, B.V.S. Reddy, M. Kayentao, H.G. Welz, and H.H. Geiger. 2004. Genomic regions influencing resistance to the parasitic weed <Emphasis Type="Italic">Striga hermonthica</Emphasis> in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016. doi:10.1007/s00122-004-1706-9

Haussmann, B.I.G., D.E. Hess, B.V.S. Reddy, S.Z. Mukuru, M. Kayentao, H.G. Welz, and H.H. Geiger. 2001. Pattern analysis of genotype × environment interaction for striga resistance and grain yield in African sorghum trials. Euphytica 122:297–308. doi:10.1023/A:1012909719137

Haussmann, B.I.G., D.E. Hess, H.-G. Welz, and H.H. Geiger. 2000. Improved methodologies for breeding striga-resistant sorghums. Field Crops Research 66:195–211. doi:10.1016/S0378-4290(00)00076-9

Hickey, JohnM., MatthewA. Cleveland, C. Maltecca, G. Gorjanc, B. Gredler, and A. Kranis. 2013. Genotype Imputation to Increase Sample Size in Pedigreed Populations. C. Gondro, J. van der Werf, B. Hayes, and J. van der Werf, ed. Methods in Molecular Biology. No. 1019. Humana Press.

Joel, D.M. 2013. The Haustorium and the Life Cycles of Parasitic Orobanchaceae. Springer, Berlin, Heidelberg.

Kapran, I., C. Grenier, and G. Ejeta. 2007. Introgression of genes for Striga resistance into African landraces of sorghum. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt 129–141

Kountche, B.A., S. Al-Babili, and B.I.G. Haussmann. 2017. Striga: A Persistent Problem on Millets. Elsevier.

Kountche, B.A., C.T. Hash, H. Dodo, O. Laoualy, M.D. Sanogo, A. Timbeli, Y. Vigouroux, D. This, R. Nijkamp, and B.I.G. Haussmann. 2013. Development of a pearl millet Striga-

resistant genepool: Response to five cycles of recurrent selection under Striga-infested field conditions in West Africa. Field Crops Research 154:82–90. doi:10.1016/j.fcr.2013.07.008

Lasky, J.R., H.D. Upadhyaya, P. Ramu, S. Deshpande, C.T. Hash, J. Bonnette, T.E. Juenger, K. Hyma, C. Acharya, S.E. Mitchell, E.S. Buckler, Z. Brenton, S. Kresovich, and G.P. Morris. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. Science Advances 1:e1400218. doi:10.1126/sciadv.1400218

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. doi:10.1093/bioinformatics/btp698

López-Ráez, J.A., K. Shirasu, and E. Foo. 2017. Strigolactones in Plant Interactions with Beneficial and Detrimental Organisms: The Yin and Yang. Trends in Plant Science 22:527–537. doi:10.1016/j.tplants.2017.03.011

Maina, F., S. Bouchet, S.R. Marla, Z. Hu, J. Wang, A. Mamadou, M. Abdou, A.-A. Saïdou, and G.P. Morris. 2018. Population genomics of sorghum (Sorghum bicolor) across diverse agroclimatic zones of Niger. Genome 1–10. doi:10.1139/gen-2017-0131

McCormick, R.F., S.K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, M. Amirebrahimi, B.D. Weers, B. McKinley, A. Mattison, D.T. Morishige, J. Grimwood, J. Schmutz, and J.E. Mullet. 2018. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal 93:338–354. doi:10.1111/tpj.13781

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. PNAS 110:453–458. doi:10.1073/pnas.1215985110

Olatoye, M.O., Z. Hu, F. Maina, and G.P. Morris. 2018. Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. G3: Genes, Genomes, Genetics 8:3269–3281. doi:10.1534/g3.118.200551

Olembo, K.N., F. M'mboyi, S. Kiplagat, J.K. Sitieney, and F.K. Oyugi. 2010. Sorghum Breeding in Sub-Saharan Africa: The Success Stories. African Biotechnology Stakeholders Forum

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D.G. Peterson, Mehboob-ur-Rahman, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, and D.S. Rokhsar. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556. doi:10.1038/nature07723

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLOS ONE 7:e32253. doi:10.1371/journal.pone.0032253

Scholes, J.D., and M.C. Press. 2008. Striga infestation of cereal crops – an unsolved problem in resource limited agriculture. Current Opinion in Plant Biology 11:180–186. doi:10.1016/j.pbi.2008.02.004

Singh, B.D., and A.K. Singh. 2015. Mapping Populations. Springer India.

Walker, T.S., and J. Alwang. 2015. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. CABI.

Yoder, J.I., and J.D. Scholes. 2010. Host plant resistance to parasitic weeds; recent progress and bottlenecks. Current Opinion in Plant Biology 13:478–484. doi:10.1016/j.pbi.2010.04.011

Yohannes, T., T. Abraha, D. Kiambi, R. Folkertsma, C. Tom Hash, K. Ngugi, E. Mutitu, N. Abraha, M. Weldetsion, C. Mugoya, C.W. Masiga, and S. de Villiers. 2015. Marker-assisted introgression improves Striga resistance in an Eritrean Farmer-Preferred Sorghum Variety. Field Crops Research 173:22–29. doi:10.1016/j.fcr.2014.12.008

Yoshida Satoko, and Shirasu Ken. 2009. Multiple layers of incompatibility to the parasitic witchweed, Striga hermonthica. New Phytologist 183:180–189. doi:10.1111/j.1469-8137.2009.02840.x

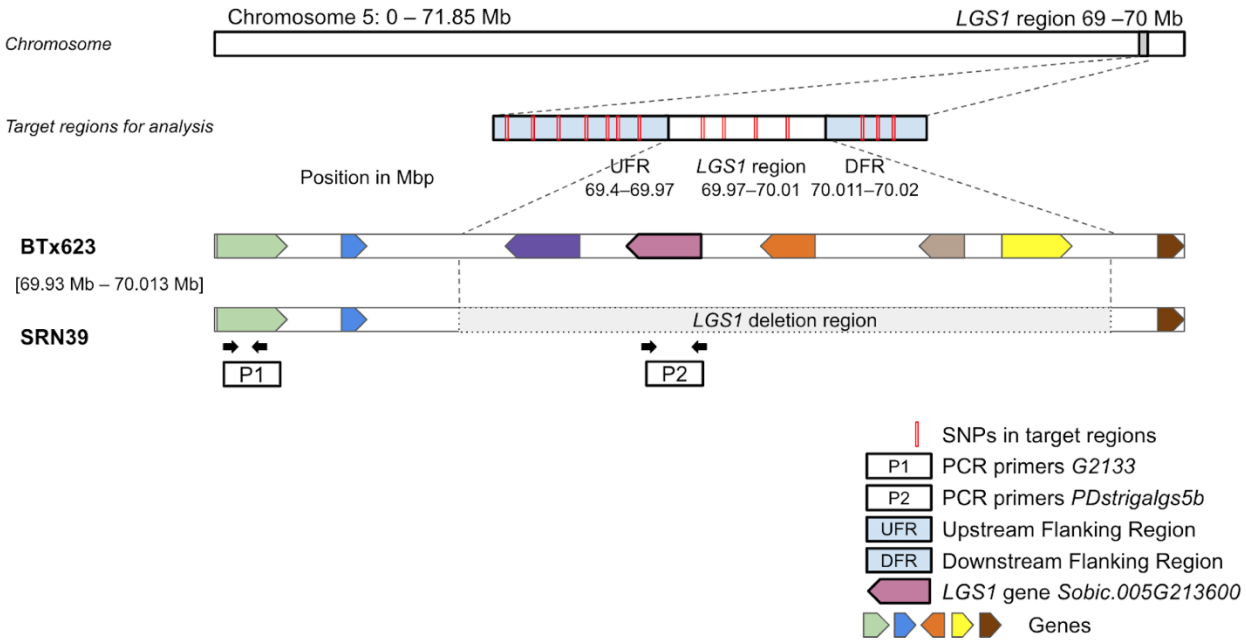


Figure 4.1. Schematic representation of the genomic region targeting *LGS1* deletion.

Genomic regions including the two sets of primers and genes were adapted from Gobena et al., (2017). "*LGS1* region" (69.97–70.011 Mb) on chromosome 5 and targeted flanking regions. Upstream Flanking Region (UFR, 69.4–69.97 Mb) and Downstream Flanking Region (DFR, 70.011–70.02 Mb), represented in blue, were selected for subsequent analysis.

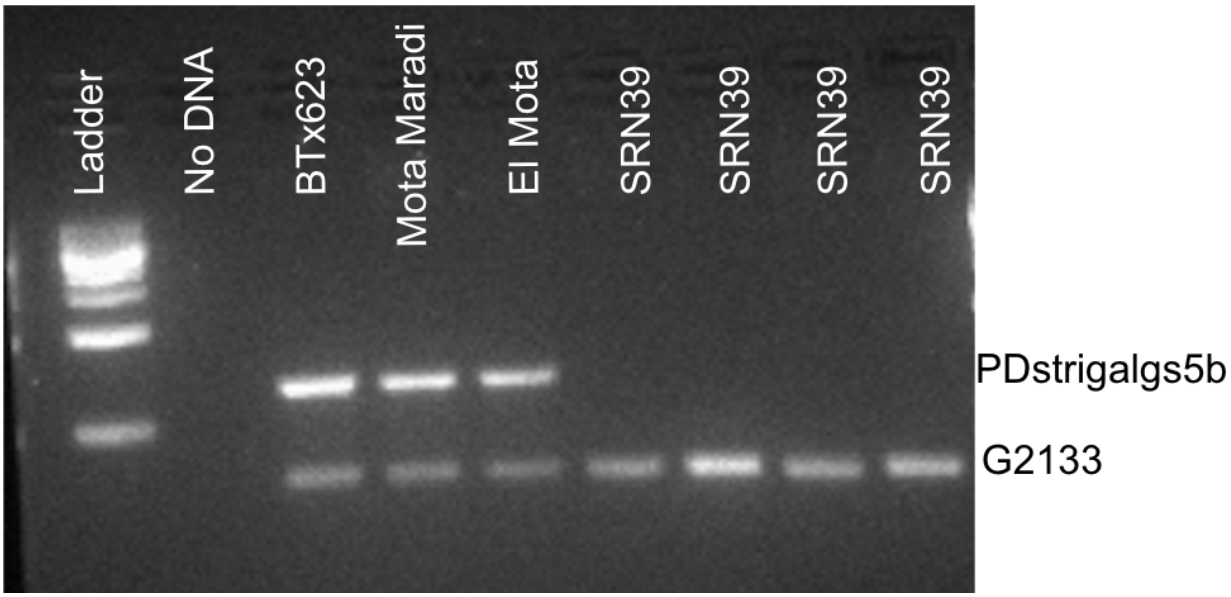


Figure 4.2. PCR analysis of *LGS1* deletion.

Primers designed from Gobena et al., (2017) were used in duplex. G2133 (non-deletion target) represents the common primer that amplifies all samples while PDstrigalgs5b (deletion target) represents the *LGS1* deletion primer that amplifies samples without deletion at *LGS1* gene Sobic.005G213600. Ladder: 500 bp bands ladder; No DNA: negative control. BTx623 is a known susceptible genotype; Mota Maradi and El Mota are hypothesized as *Striga* susceptible genotypes; SRN39 is a known *Striga* resistant genotype.

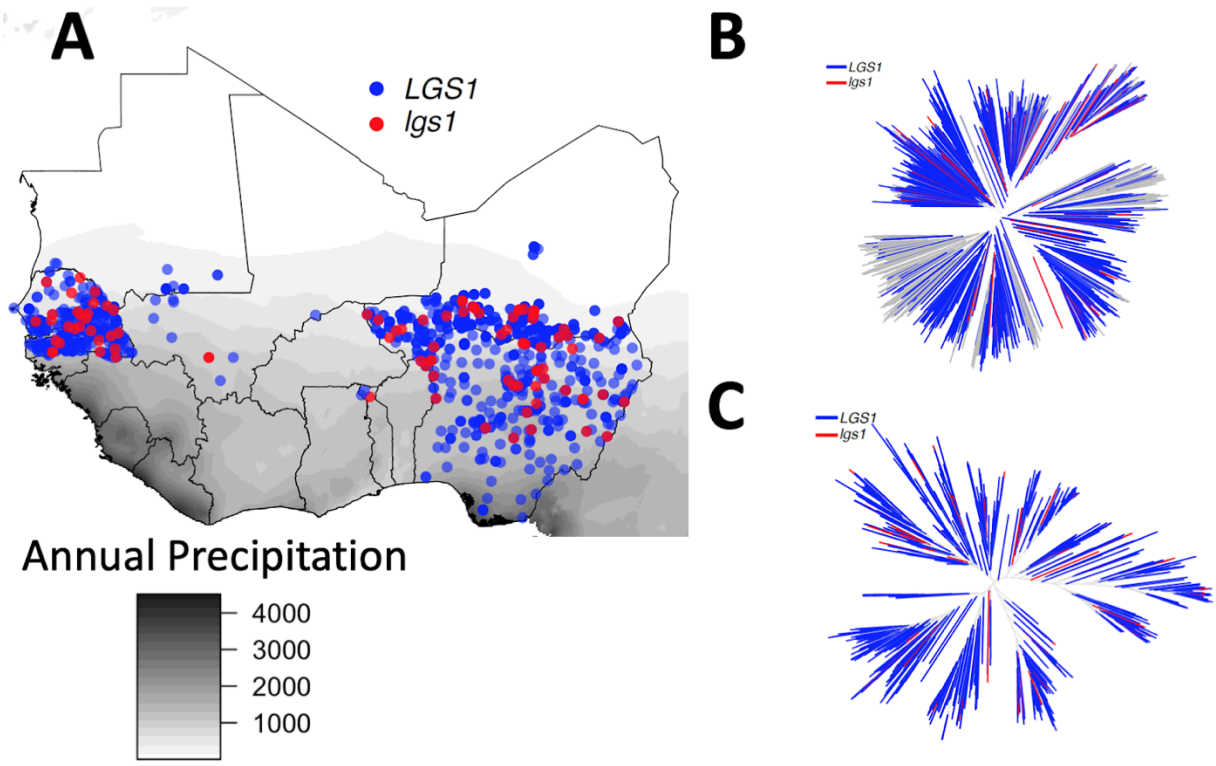


Figure 4.3. Inference of *LGS1* presence/absence variation in West African sorghum accessions.

Accessions are color-coded based on the hypothesized *LGS1* wild-type (in blue) and *lgs1* loss of function (in red) based on the number of missing SNPs within the "*LGS1* region" (69.97–70.011 Mb) from genotyping-by-sequencing. (A) Distribution of georeferenced accessions in Mali, Niger, Nigeria, Senegal, and Togo; (B) Neighbor-joining tree of chromosome 5; (C) Neighbor-joining tree of *LGS1* region (69.97–70.11 Mb) on chromosome 5.

Total number of SNPs within the "*LGS1* region" (69.97–70.011 Mb) is n = 21. *lgs1* cutoff: number of missing SNPs within the "*LGS1* region" is 21. *LGS1* cutoff: number of missing SNPs less than 21 SNPs.

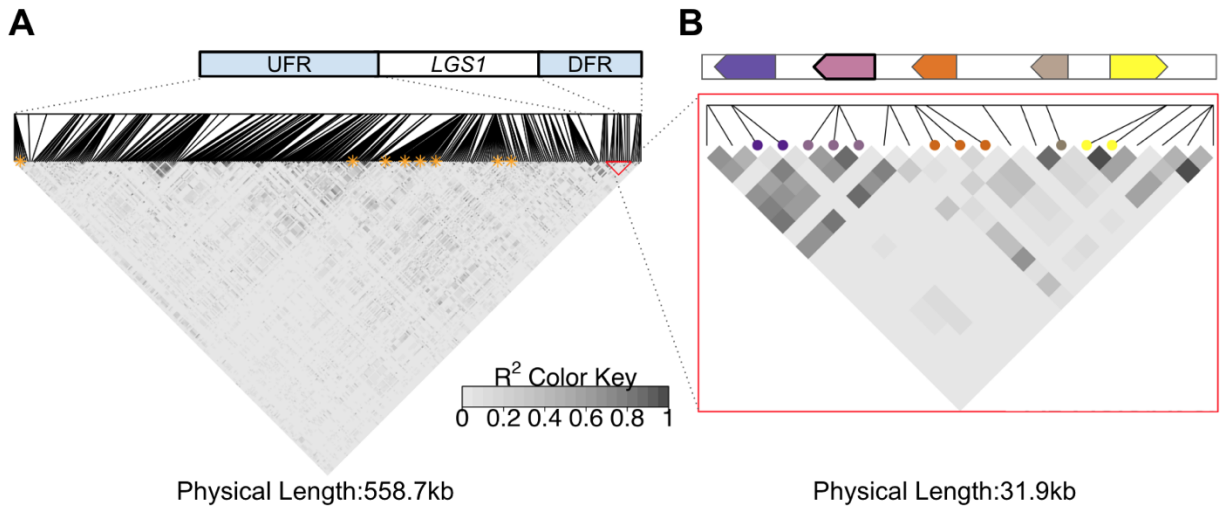


Figure 4.4. Linkage disequilibrium of the targeted region in this study.

(A) Pairwise R^2 based heatmap for the "extended *LGS1* region" (69.97–70.011 Mb), upstream flanking region (UFR), and downstream flanking region (DFR); orange stars indicate the position of single nucleotide polymorphisms (SNPs) used for marker development; red box indicates the "*LGS1* region" (69.97–70.011 Mb). (B) Pairwise R^2 based heatmap for the *LGS1* region (69.97–70.011 Mb) only. Colored boxed represent the genes within the region. Colored points represent SNPs found in their respective five genes within the deletion.

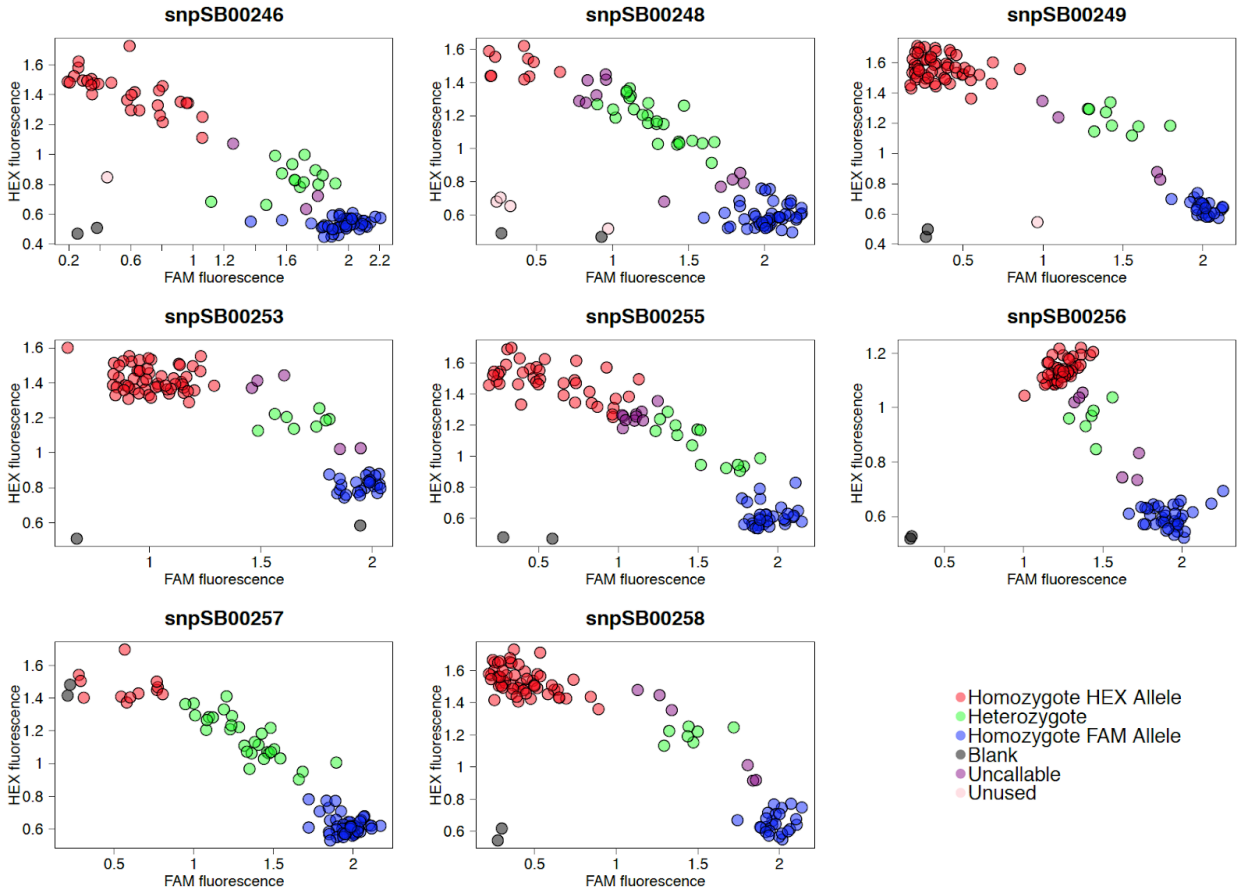


Figure 4.5. KASP genotyping assays of eight selected markers using the test plate of inbred accessions.

Red and blue colors are assigned based on the HEX fluorescence (reference allele of SRN39) and FAM fluorescence (reference allele of BTx623): red: homozygotes for the HEX allele; blue: homozygotes for the FAM allele. Green: heterozygotes; purple: uncallable for one of the three groups; Unused: no amplification; Blank: genotyping negative control. Genotypes used: BTx623, El Mota, IRAT 204, Mota Maradi, MR732, PI513676, PI513680, PI513749, Sepon 82, SRN39, PI656027, PI277541, PI284980, PI465448, Tx430.

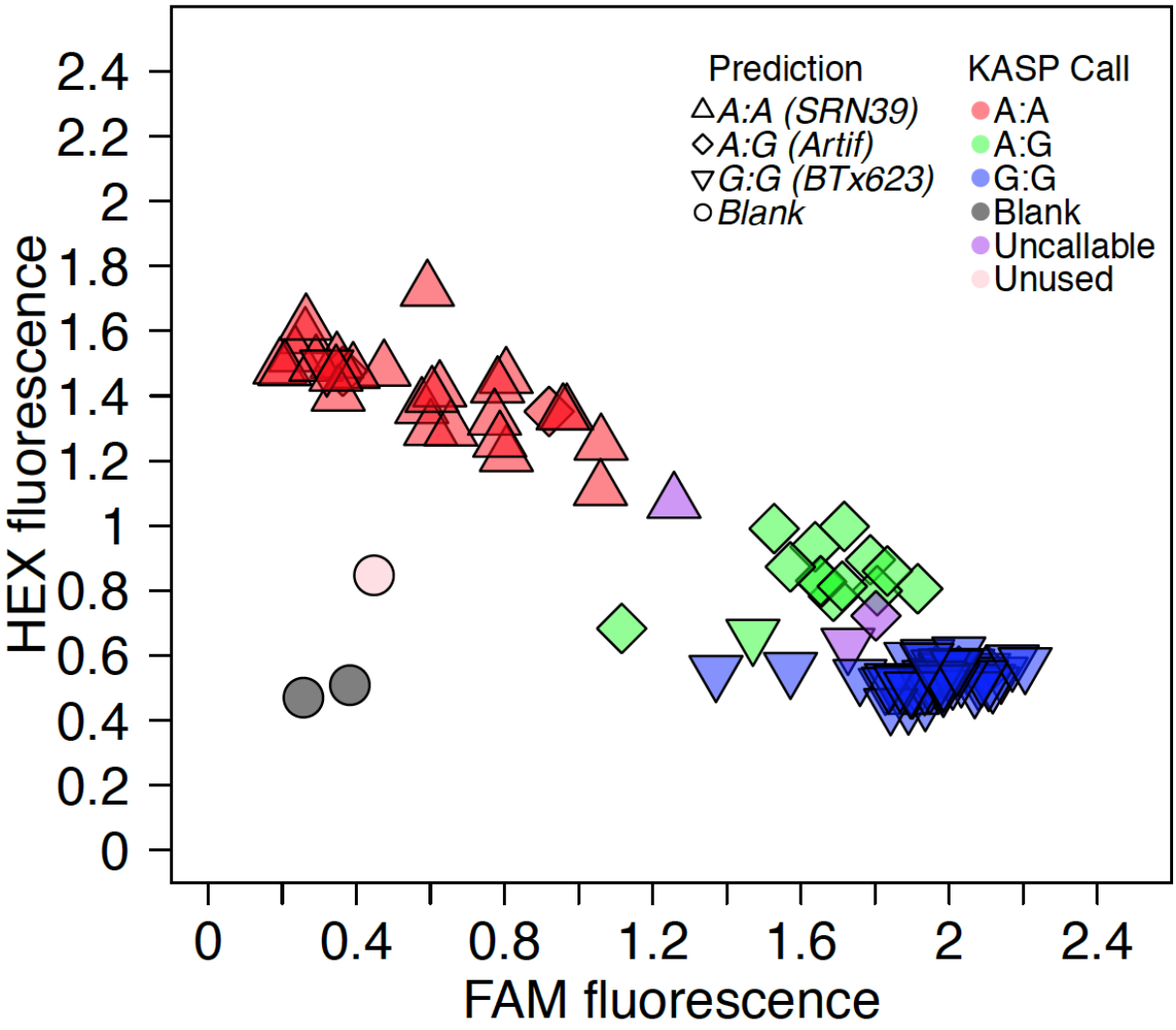


Figure 4.6. Genotyping for snpSB00246 for the test plate using KASP assay.

Expected KASP call for each sample are represented by point-up triangles (predicted A:A), point-down triangles (predicted G:G), diamonds (predicted A:G), and circles (blank) are a priori hypotheses based on the *Striga* resistance/susceptibility of the accessions. Observed KASP genotype calls are color-coded as follows: red, A:A; blue, G:G; green, A:G; purple, uncallable; pink, unused due to no amplification; black, blank, negative control. Artif: artificial heterozygotes, mix of leaf tissues between predicted A:A (SRN39GMS) and predicted G:G (BTx623, Mota Maradi, Sepon82, PI513676, PI513740, and PI513680).

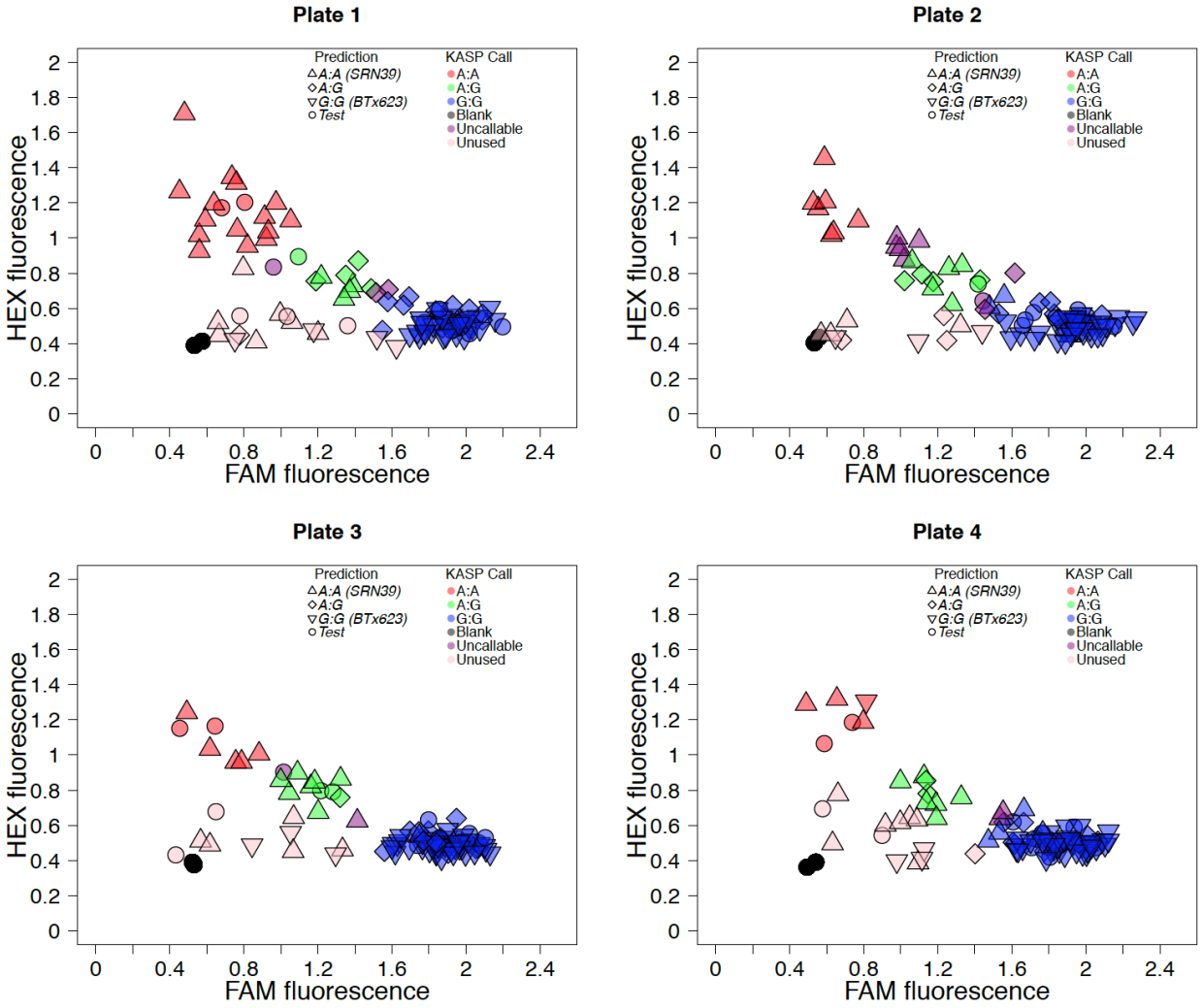


Figure 4.7. Genotyping assay for snpSB00246 using biological replicates of the genotypes in four plates.

Expected alleles for each genotype are represented by triangles, diamond shapes based on *a priori* information; circles represent genotypes that do not have *a priori* information. Observed genotype calls based on the KASP genotyping are color-coded in red: homozygotes for the HEX allele (A:A); blue: homozygotes for the FAM allele (G:G); green: Heterozygotes (A:G); purple: uncallable for one of the three groups; unused: no amplification; blank: genotyping negative control.

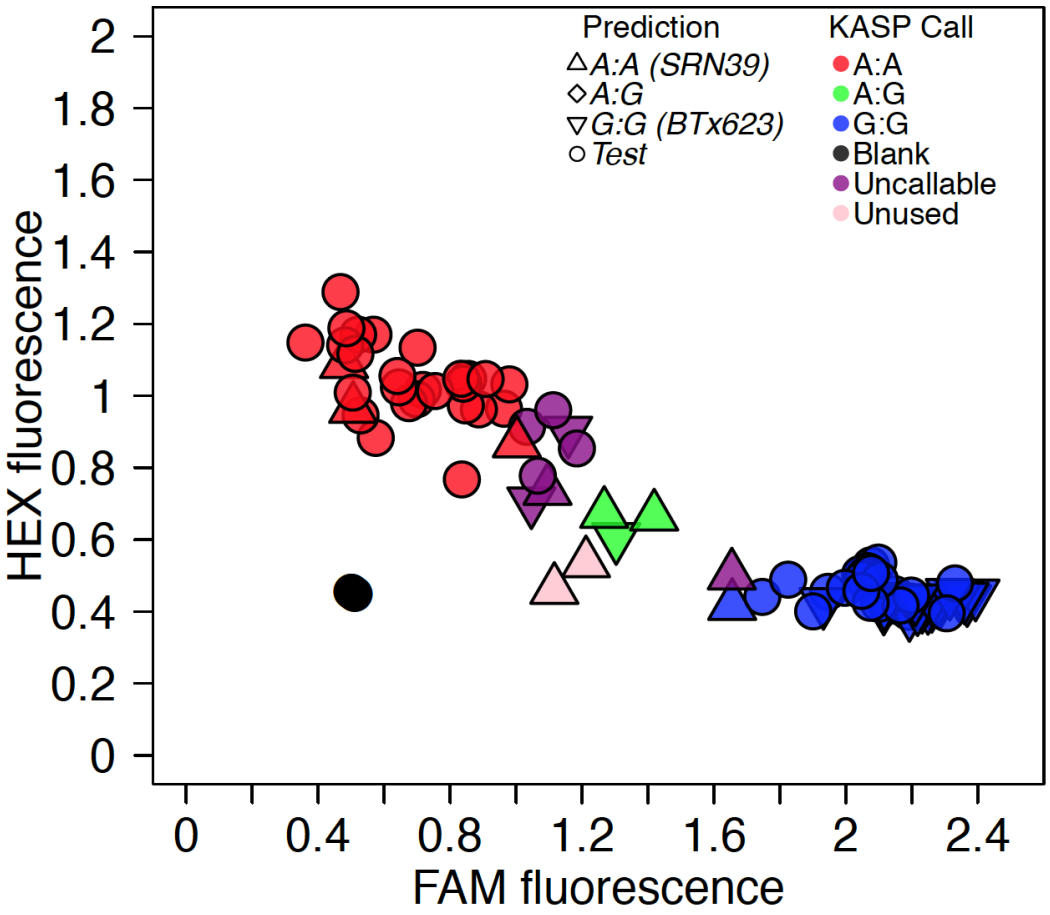


Figure 4.8. KASP assay for snpSB00246 of F2:3 progenies of SRN39 × Mota Maradi..

Expected KASP call for each sample are represented by point-up triangles (predicted A:A), point-down triangles (predicted G:G), diamonds (predicted A:G), and circles (Unknown or no *a priori* information) are *a priori* hypotheses based on the *Striga* resistance/susceptibility of the accessions. Observed KASP genotype calls are color-coded as follows: red, A:A; blue, G:G; green, A:G; purple, uncallable; pink, unused due to no amplification; black, blank, negative control.

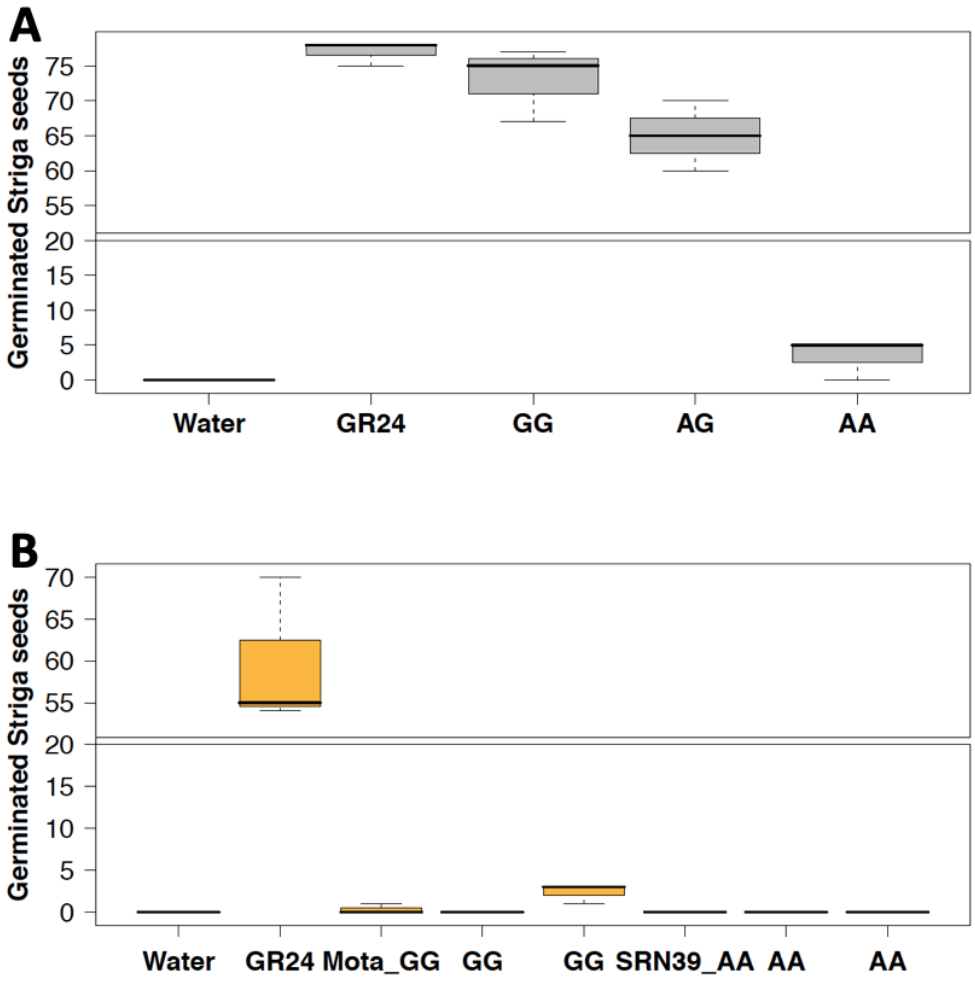


Figure 4.9. Germination of preconditioned *Striga* seeds 72 hours after root exudates application.

(A) Prediction for the number of germinated seeds for negative control (deionized water), negative control (GR24), homozygotes and heterozygotes genotypes. (B) Observation for the number of germinated seeds for negative control (deionized water), negative control (GR24), homozygotes and heterozygotes genotypes. Boxes indicate the lower and upper quartiles, and whiskers indicate the minimum and the maximum number of germinated seeds for each class.

Table 4.1. Candidate SNPs on chromosome 5 selected for KASP genotyping.

SNP	Type [a]	Distance to *LGS1* gene [b] (kb)	Average R^2 [c]	MAF [d]	Hypothesized SNP Effect [e]	Marker
S5_69450954	G/A	535	0.02	0.42	5 prime UTR	snpSB0246
S5_69794954	A/G	191	0.09	0.43	Non-synonymous coding	snpSB0253
S5_69847924	C/T	138	0.002	0.27	Synonymous coding	snpSB0257
S5_69851828	G/C	134	0.005	0.43	Non-synonymous coding	snpSB0255
S5_69852443	G/A	134	0.008	0.46	Synonymous coding	snpSB0258
S5_69854897	G/T	131	0.006	0.45	Upstream	snpSB0248
S5_69879981	T/G	106	0.007	0.28	Non-synonymous coding	snpSB0256
S5_69882454	T/C	104	0.011	0.35	Synonymous coding	snpSB0249

[a]Polymorphic Type: Major Allele / Minor Allele; [b]R^2: average value of R^2 between the SNP and the *LGS1* deletion (69.97–70.11 Mb); [c]*LGS1* gene: Sobic.005G213600; [d]MAF: Minor Allele frequency; [e]Hypothesized SNP Effect: obtained from https://phytozome.jgi.doe.gov

Table 4.2. Prediction and observation based on KASP genotyping of snpSB00246 for inbred genotypes in the test plate.

Genotype	N	Expected Allele Call	N with Expected Allele Call	Unexpected Allele Call
BTx623	8	G:G	8	-
El Mota	4	G:G	4	-
IRAT204	4	G:G	4	-
Mota Maradi	4	G:G	4	-
MR732	5	G:G	5	-
PI277541	2	A:A	2	-
PI284980	2	A:A	2	-
PI465448	2	A:A	2	-
PI513676	4	G:G	3	A:A
PI513680	4	G:G	4	-
PI513740	4	G:G	4	-
Sepon82	4	G:G	4	-
SRN39	8	A:A	6	G:G, Uncallable
PI656027 (SRN39)	8	A:A	8	-

N: Number of individuals; Uncallable: FAM and HEX fluorescence values were not sufficient to call the allele.

Table 4.3. Genotype calls of snpSB00246 for F_2 bi-parental populations.

Family	Line	F_2	Expected $F_{2:3}$	Observed $F_{2:3}$ (N^a)	Alternative allele [b]
SRN39 x MotaMaradi	P7	A:G	Segregating	G:G (3)	-
	P3	A:G	Segregating	G:G (3)	-
	P23	G:G	G:G	G:G (3)	-
	P15	A:A	A:A	A:A (1)	A:G, unused
	P38	A:A	A:A	A:A (1)	A:G, uncallable
	P9	A:A	A:A	A:A (1)	unused, uncallable
	P18	A:A	A:A	A:A (2)	A:G
SRN39 x MR732	P1	A:G	Segregating	G:G (3)	-
	P22	G:G	G:G	G:G (3)	-
	P38	G:G	G:G	G:G (3)	-
	P42	G:G	G:G	G:G (3)	-
	P63	A:G	Segregating	G:G (2)-A:A (1)	-

P24	A:A	A:A	A:A (2)	unused
P39	A:A	A:A	A:A (2)	
P40	A:G	Segregating	G:G (2)	uncallable
P65	A:G	Segregating	G:G (2)	unused
P37	A:A	A:A	A:A (1)	A:G, G:G

[a]N: number of lines with the allele out of three lines genotyped. [b]Alternative allele: unused, No PCR amplification; uncallable, FAM and HEX fluorescence values were not sufficient to call the allele.

Chapter 5 - Demand-led driven research and development to facilitate crop improvement

New sorghum varieties not enough to drive adoption in Niger

West African breeding programs developed and delivered stress-resilient and high yielding varieties to farmers, adapted to a wide range of growing environments. New sorghum varieties were developed in the early 1970s in collaboration with international partnerships. Nevertheless, improved varieties did not always satisfy the farmers' needs and end-users preferences where different agroclimatic zones and uneven precipitation patterns are frequent. In Niger, for instance, the *Striga* resistant variety SRN39 and the drought-tolerant variety IRAT204 have both a lower adoption rate, estimated to be 0.07% and 2.7% of the total cultivated sorghum area respectively (Walker and Alwang, 2015). Both varieties had lower adoption rates than purified landrace varieties (Ministère de l'agriculture du Niger, 2012; Walker and Alwang, 2015). SRN39 (IAR, 1991), IRAT204 (French Institute,1966), and SEPON82 (ICRISAT, 1982) were internationally released varieties that were agronomically evaluated in Niger for specific tolerance and resistance to stressors rather than demand-led varieties (e.g. Mota Maradi and MDK) which fit the end-users requirements for an adapted and appreciated variety (Ministère de l'agriculture du Niger, 2012). In addition, known drought-tolerant lines such as BTx642 and Tx7000 were not used for breeding in Niger even though those lines carry stay-green alleles. Towards breeders' goals of developing locally-adapted varieties with enhanced tolerance to biotic and abiotic stressors, the introgression of favorable alleles into local varieties might provide insights into delivering farmer preferred varieties. The local variety Mota Maradi is adapted to Sahelo-Soudanian agroclimatic zones and a reference variety for adaptation and grain yield in most of the seed companies in Niger (e.g. La Sahélienne Des Semences, LSDS-Halal).

Genomic studies provide resources for crop improvement in West Africa

The research presented in this thesis assembled three proposed approaches for implementing genomics-enabled breeding. This research focuses on understanding the unique and complex genetic diversity of West African sorghum germplasm to help breeders deliver better adapted and adopted varieties (Figure 1.1) using GBS for Nigerien, Nigerian, and Senegalese collections in GRIN (Olatoye et al., 2018; Faye et al., 2019). Those results provide a basic starting point for establishing the GEB in West Africa. Based on our conceptual roadmap for the establishment of genomics-enabled breeding, the studies in the chapters are complementing each other for delivering markers and traits to breeders in West Africa. The first approach focused on population genomics analysis to better understand the genetic diversity of the West African sorghum germplasm focusing on the Nigerien sorghum. The second approach highlighted quantitative genomics to identify genomic regions associated with drought tolerance in a managed environment. The third approach took advantage of the cloned gene associated with *Striga* resistance to develop variants linked and in linkage disequilibrium with *LGS1*.

Our proposed framework may work independently as presented in the chapters; however, it is useful to link discoveries to the crop improvement programs for major crops in West Africa. With the recent advancement in sequencing technologies, reference genomes of pearl millet and cowpea are available (Varshney et al., 2017; Lonardi et al., 2019). Those resources might increase our knowledge of the existing genetic diversity, their interaction with other crops (i.e intercropping of sorghum and cowpea) in West Africa, and the ability to develop new varieties through genomic-enabled breeding.

Genomics to characterize the diversity of sorghum

To get a deeper insight into the genetic basis of local adaptation, farmers' and end-users' preferences, genomic analyses were conducted for the West African sorghum germplasm. In the second chapter, the Nigerien germplasm available at USDA-GRIN was used for population genomic analysis. Given that sorghum is cultivated by many ethnic groups and across precipitation gradients of Niger, it was hypothesized that the adaptation to agroclimatic zones has acted on loci controlling maturity and grain pigmentation. The genetic diversity existing in the Nigerien sorghum highlights six genetic groups with respect to the agroclimatic zones. Those groups were mainly identified based on the botanical types, grain colors, and agroclimatic zones. We found evidence of local adaptation to the sub-humid zone of Niger, such as the SNPs colocalizing with the flowering time adaptation gene of Vgt1 and abph1 that are critical for specific botanical types. In the southern Niger, where higher precipitations and longer growing seasons are observed (Figure 2.1), specific ideotypes (i.e. open panicle, late maturity) might be considered to limit grain mold damages for instance. Guinea botanical type is more prevalent in this region, so taking into account local adaptation of guinea in the humid regions of West Africa should be considered in the breeding product profiles.

With respect to farmer preferences, the local names Mota and Jenjari were analyzed to identify signatures of selection that acted on those sympatric varieties. We found evidence of balancing selection on Mota and Jenjari varieties grown in the same geographical area. Even though one of the hypotheses was on grain pigmentation genes, other hypotheses (derived from the observation higher pairwise FST across the genome in Figure 2.5) in which maturity groups and grain quality control farmers' and end-users' preferences on local varieties can be tested.

Dissecting the genetic basis of water-deficit in a managed environment

To investigate the genetic variation of West African sorghum under drought-stress, genome-wide association studies were conducted in the WASAP to account for historical recombination in the panel. We hypothesized that the WASAP harbors unique variations conferring drought tolerance in addition to stay-green loci. The study analyzed the ecophysiological genomics of pre-flowering drought tolerance in West African sorghum. This study was conducted to better understand the WASAP in the managed environment with two water treatments. The specific hypothesis relies on the oligogenic variation of West African sorghum under drought stress. The use of diverse germplasm for association studies led us to identify genomic regions associated with grain weight and vegetative biomass under water deficit. This experiment was conducted in the lysimeter which is a controlled environment in PVC tubes. The well-watered treatment was closely related to on-station conditions at ICRISAT-Sadoré in the rainy season. Future studies for the drought experiment are to test that the drought-induced trial reflects the station experiment and farmers' fields. Furthermore, in the lysimeter experiment, testing relevant hypotheses underlying product developments will be of significance to deliver traits and markers to breeders especially for complex traits difficult to phenotype in field conditions. Given the existing diversity in the WASAP and the adaptation derived from a wide range of agroclimatic zones, a testing hypothesis might underlie the phenotypic plasticity of some genotypes in well-watered and water-stressed environments. for phenotypic plasticity of some genotypes can be hypothesized when grown in a common garden for well-watered and water-stressed treatments.

Even though the pre-flowering drought imposition reflects our predictions for the check lines, more studies are needed to confirm whether the performance of some genotypes is similar

to the farmers' field conditions. The water-stressed treatment was imposed at the flag-leaf appearance, and that was considered in this study as a pre-flowering water deficit. We tested the hypothesis whether those SNPs are in LD with genes nearby to reduce the effect of false-positive associations. Moreover, the marker-trait associations (MTA) observed might be linked to variants that balance yield components reduction under drought stress.

In order to develop broadly adapted varieties, combining multiple desirable traits would be possible. Major breeding targeted traits common in West Africa represent a starting point to deliver broadly adapted varieties. The international partnership and the germplasm exchange allow for field evaluations in West African countries of breeding lines. For drought-tolerant lines, the known post-flowering drought tolerance BTx642 is one of the donor parents in breeding programs. Stay-green loci (Stg1–4) in BTx642 converted to KASP markers were made available to the scientific community. Breeders are now taking this advantage of outsourcing KASP genotyping to test the hypothesis that the stay-green loci control pre-flowering drought tolerance in progenies of BTx642 and locally preferred recurrent germplasm. Marker-assisted backcross selection could then be the method in such traits for breeders to recover the genome of the parent in the progenies with stay-green favorable loci.

Future climate variabilities will lead to increased temperatures, and with reduced annual precipitations, crop productions will decrease. Water-deficit conditions and high-temperature pressures are among abiotic stressors that affect sorghum production (Prasad et al., 2008, 2015). Hence, enhancing tolerance to drought and heat stress might provide insights for product development in targeted agroclimatic zones.

Towards broad and durable *Striga* resistance

Striga hermonthica resistant variety SRN39 has a deletion on chromosome 5 responsible for the low germination stimulant resistance (Gobena et al., 2017). In chapter 4, high-density markers were generated for the West African sorghum and the sorghum association panel to identify genotypes carrying the deletion on *LGS1*. Based on the empirical distribution of missing SNPs, sets of random 21 SNPs and 21 consecutive SNPs on chromosome 5 were analyzed suggesting that the missing SNPs observed are not due to the low coverage. With a false discovery rate of 35% (Figure C.3), only 1– 4 accessions in each of the West African sorghum assayed will have a deletion, thus a putative resistant line. The SNPs used for KASP analysis are either common for all the susceptible lines or specific in the case of the differentiation between the resistant line and one susceptible line. To predict the presence or absence of *LGS1* deletion in multiple genetic backgrounds, we used diagnostic markers that tag the deletion. Along with the KASP markers, in lab germination assays provide insights into the production of strigolactones responsible for the *Strig*a seeds germination. In this study, *Striga* seeds originating from Mali were used for the assay. Subsequent studies might focus on testing the hypothesis that the mechanism of sorghum resistance due to *LGS1* interacts with Striga ecotypes and local adaptation.

Given multiple mechanisms of resistance to *Striga hermonthica*, pyramiding those resistant traits into a preferred variety could provide a broader resistance in West Africa. While this study focuses on one of the resistance mechanisms of sorghum, for instance, the low germination stimulant, other resistant lines should provide more insights into future product developments. The *Striga* resistant sorghum variety N13 has a mechanical barrier through a hypersensitivity mechanism that prevents haustorial attachment with the parasite (Haussmann et

al., 2000b; Yohannes et al., 2015). Genomic regions have been identified in bi-parental populations with N13 as the resistant parent (Haussmann et al., 2000a, 2001). Yet, one mechanism might be sufficient in some agroclimatic zones and breeding targets. In specific breeding product profiles, the resistance mechanisms could be both integrated via gene pyramiding to reduce yield losses caused by the parasite. For instance, the resistance mechanisms due to *LGS1* (SRN39) and mechanical barrier (N13) could be combined into a new locally-adapted variety instead of developing two product developments for *Striga* resistance. SRN39 and N13 are potential candidate donors for developing *Striga* resistant varieties in sub-Saharan Africa.

Hypothesis-driven research for genomics-enabled breeding

In West Africa, staple crops are cultivated under limiting factors reducing yield in smallholder farming systems. Breeders seek to improve staple crops in this region where multiple limiting factors are occurring with temporal and spatial variability. The major limitation in small breeding programs remains the lack of resources available to achieve the goals. For instance, some of the traits to improve are often challenging in phenotyping or controlled by multiple genes. As crop improvement involves a combined effort of interdisciplinary approaches to meet the needs of farmers and end-users with respect to crop management systems, agroclimatic zones, land pressure, and soil type, our research focuses on genomic studies.

Previous breeding programs in West Africa considered farmers and end-users' needs for their product developments for broad agroclimatic adaptation (Ndjeunga et al., 2015). However, the development of the new varieties often relies on agronomic performances in research stations with limited participatory selection and multi-environment trials, representing a small fraction of the targeted population environments (Ceccarelli and Grando, 2007). Though a few product

developments remain challenging due to the lack of selection in specific agroclimatic zones, farmer's preferences coupled with other disciplinary teams. By using the integration of multidisciplinary approaches, the variability of some traits that need to be included in the pipeline as well as accounted for under the social and production systems in the targeted end-users' agroclimatic zones, can be better understood. Soil infertility and soil erosion are prevalent in farmers' fields (Abdoulaye and Sanders, 2005; Leiser et al., 2014; Gemenet et al., 2016). The decentralized participatory selection might capture significant insights in breeding programs in highly changing environments and variable needs (e.g. maturity, grain yield increase, tolerance, grain quality, fodder yield, and fodder quality).

Future integrated methods to account for such variability in West Africa will require a deep understanding of farmers' and end-users' needs and lessons learned from advanced crop improvement programs. Product profiles developed by breeding programs will certainly provide insights into the varieties to improve or to develop for specific regions (Cobb et al., 2019). Briefly, a breeding product profile includes information regarding the product development for a specific agroclimatic zone and among certain end-users while considering gender. In addition, photoperiod, maturity, and parental lines are included in the profile. The product profile guides breeders to set priorities and review management and resources for new varieties that meet the need for end-users. Those product profiles developed by breeding programs in West Africa are targets for future product developments for new varieties locally and/or broadly adapted.

Genomics-enabled breeding to support current crop improvements in the Sahel

Regarding the impact of genomics-enabled breeding in West Africa, this approach focused mainly on a few aspects of breeding. Yet, further disciplinary teams, once associated in

the pipeline of the crop improvement would drive better adoption in smallholder farming systems and end-user's preference. Data management programs such as "Breeding Management System, BMS" (https://excellenceinbreeding.org/) could support breeding strategies. In this study, we focused, on the ecophysiological genomics, on the hypothesis for a few genes of major effects controlling a trait of interest, for instance, drought stress. When few markers are associated with the traits of interest, marker-assisted backcrossing could be beneficial for recovering the genome of the recurrent parent. In addition, the polygenic variation could be associated with the desirable trait, for instance where small effect QTLs are associated with the trait. In this case, genomic selection might contribute to future breeding programs in developing countries (Muleta et al., 2019).

In this dissertation, SNPs were identified for local adaptation, grain pigmentation, pre-flowering drought tolerance, and *Striga* resistance that are candidates for marker development deliverable to breeders (Table 5.1). My hypothesis is that those SNPs are good candidates for developing new versions of local varieties carrying resistant/tolerant alleles for specific stress. In such cases, a few cycles of backcrossing to the recurrent parent using foreground and background makers. While background markers are specific to the recurrent parent used by the breeding parent, SNPs could be converted into KASP markers for breeding programs in West Africa. For instance, marker-trait associations (MTAs) associated with vegetative biomass (S2_57620549, S2_57664163, and S2_57663973) and grain weight (S5_16199662) are additional hypothetical candidates for KASP marker development under the hypothesis that those MTAs could confer tolerance to water deficit in West African sorghum. Similarly, for the SNPs linked and in linkage disequilibrium with the loss of function of *LGS1*, the marker snpSB00246 (S5_69450954) was among the best predictor for putative resistant and susceptible

to *Striga hermonthica*. In-lab germination and field assays will test the hypothesis that the variant (A or G) at the marker differentiates *Striga hermonthica* response in West African sorghum.

There are several recipient lines based on the local adaptation (e.g agroclimatic zones) and farmer preferences. In West Africa, introgression of a few traits for resistance and/or tolerance could provide better adapted and more appreciated varieties of new versions of Mota Maradi (Niger), Kapelga (Burkina Faso), Nganda (Senegal), and Sorvato1 (Togo). Drought tolerant lines, BTx642 and Tx7000, are some putative lines for allele introgression into a local background. BTx642, the post-flowering drought tolerant that we used as one of check lines in the third chapter showed a reduction of grain weight in the pre-flowering water stress in comparison with Tx7000, the pre-flowering drought tolerant. With the stay-green (*Stg1–4*) KASP markers available for the scientific community, breeding for drought tolerance with BTx642 as a donor is now possible in small breeding programs of West Africa. Future product developments will provide both resistance to *Striga* and tolerance to drought stress into a locally preferred variety for a sustainable yield advantage under multiple stressors. Hypotheses to evaluate might thus be that the introgression of favorable alleles into a local genetic background confers drought and *Striga* resistance in Sahelian smallholder farming systems. In addition, product developments for developing grain mold and midge resistant varieties are of significant importance in sub-Saharan Africa. Potential donors might include resistant varieties (e.g. Sureno and ICSV88032 for grain mold and midge resistance respectively).

Strategies of genomic-assisted breeding may incorporate diverse studies and multi-environment trials for a better knowledge of the genetic basis of crop adaptation and improvement. Taking into consideration farming and socio-agroecological systems, increased

land pressure, and multiple stressors can bring all together to enhance tolerance in farmers' fields. In the case of *Striga hermonthica*, integrated management will, for instance, reduce the seed banks in farmers' fields. The management should be cost-effective and sustainable for smallholder farmers such as suicidal germination through trap crops, crop rotation, and chemical compounds that induce *Striga* germination without the host plant for the attachment and parasitism (Botanga et al., 2003; Yoder and Scholes, 2010; Kountche et al., 2019). Through the international partnerships, *Striga* phenotyping facility in the south-central of Niger, Konni was implemented providing parasite evaluation and screening during rainy seasons in this highly infested area. Thus, putative *Striga*-resistant genotypes carrying the deletion of *LGS1* could be tested in on-station and neighboring farmer's fields for participatory evaluation selection. In this regard, competing hypotheses on the performance of putative resistant genotypes on-station and farmer's fields might be tested in the phenotyping facility that represents one of the targeted populations of environments.

References

Abdoulaye, T., and J.H. Sanders. 2005. Stages and determinants of fertilizer use in semiarid African agriculture: the Niger experience. Agricultural Economics 32:167–179. doi:10.1111/j.0169-5150.2005.00011.x

Botanga, C.J., S.O. Alabi, C.A. Echekwu, and S.T.O. Lagoke. 2003. Genetics of Suicidal Germination of Striga hermonthica (Del.) Benth by Cotton. Crop Science 43:483–488. doi:10.2135/cropsci2003.4830

Ceccarelli, S., and S. Grando. 2007. Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155:349–360. doi:10.1007/s10681-006-9336-8

Cobb, J.N., R.U. Juma, P.S. Biswas, J.D. Arbelaez, J. Rutkoski, G. Atlin, T. Hagen, M. Quinn, and E.H. Ng. 2019. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theor Appl Genet 132:627–645. doi:10.1007/s00122-019-03317-0

Faye, J.M., F. Maina, Z. Hu, D. Fonceka, N. Cisse, and G.P. Morris. 2019. Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. Ecology and Evolution 9:6038–6051. doi:10.1002/ece3.5187

Gemenet, D.C., W.L. Leiser, F. Beggi, L.H. Herrmann, V. Vadez, H.F.W. Rattunde, E. Weltzien, C.T. Hash, A. Buerkert, and B.I.G. Haussmann. 2016. Overcoming Phosphorus Deficiency in West African Pearl Millet and Sorghum Production Systems: Promising Options for Crop Improvement. Front Plant Sci 7. doi:10.3389/fpls.2016.01389

Gobena, D., M. Shimels, P.J. Rich, C. Ruyter-Spira, H. Bouwmeester, S. Kanuganti, T. Mengiste, and G. Ejeta. 2017. Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. PNAS 201618965. doi:10.1073/pnas.1618965114

Haussmann, B.I.G., D.E. Hess, B.V.S. Reddy, S.Z. Mukuru, M. Kayentao, H.G. Welz, and H.H. Geiger. 2001. Pattern analysis of genotype × environment interaction for striga resistance and grain yield in African sorghum trials. Euphytica 122:297–308. doi:10.1023/A:1012909719137

Haussmann, B.I.G., D.E. Hess, B.V.S. Reddy, H.G. Welz, and H.H. Geiger. 2000a. Analysis of resistance to <Emphasis Type="BoldItalic">Striga hermonthica</Emphasis> in diallel crosses of sorghum. Euphytica 116:33–40. doi:10.1023/A:1004046001009

Haussmann, B.I.G., D.E. Hess, H.-G. Welz, and H.H. Geiger. 2000b. Improved methodologies for breeding striga-resistant sorghums. Field Crops Research 66:195–211. doi:10.1016/S0378-4290(00)00076-9

Kountche, B.A., M. Jamil, D. Yonli, M.P. Nikiema, D. Blanco-Ania, T. Asami, B. Zwanenburg, and S. Al-Babili. 2019. Suicidal germination as a control strategy for Striga hermonthica (Benth.) in smallholder farms of sub-Saharan Africa. PLANTS, PEOPLE, PLANET 1:107–118. doi:10.1002/ppp3.32

Leiser, W.L., H.F.W. Rattunde, E. Weltzien, N. Cisse, M. Abdou, A. Diallo, A.O. Tourè, J.V. Magalhaes, and B.I. Haussmann. 2014. Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African Sorghum. BMC Plant Biology 14:206. doi:10.1186/s12870-014-0206-6

Lonardi, S., M. Muñoz-Amatriaín, Q. Liang, S. Shu, S.I. Wanamaker, S. Lo, J. Tanskanen, A.H. Schulman, T. Zhu, M.-C. Luo, H. Alhakami, R. Ounit, A.M. Hasan, J. Verdier, P.A. Roberts, J.R.P. Santos, A. Ndeve, J. Doležel, J. Vrána, S.A. Hokin, A.D. Farmer, S.B. Cannon, and T.J. Close. 2019. The genome of cowpea (Vigna unguiculata [L.] Walp.). The Plant Journal 98:767–782. doi:10.1111/tpj.14349

Ministère de l'agriculture du Niger. 2012. Catalogue national des espèces et variétés végétales du Niger

Muleta, K.T., G. Pressoir, and G.P. Morris. 2019. Optimizing Genomic Selection for a Sorghum Breeding Program in Haiti: A Simulation Study. G3: Genes, Genomes, Genetics 9:391–401. doi:10.1534/g3.118.200932

Ndjeunga, J., K. Mausch, and F. Simtowe. 2015. Assessing the effectiveness of agricultural R&D for groundnut, pearl millet, pigeonpea and sorghum in West and Central Africa and East and Southern Africa. T.S. Walker and J. Alwang, ed. CABI, Wallingford.

Olatoye, M.O., Z. Hu, F. Maina, and G.P. Morris. 2018. Genomic Signatures of Adaptation to a Precipitation Gradient in Nigerian Sorghum. G3: Genes, Genomes, Genetics 8:3269–3281. doi:10.1534/g3.118.200551

Prasad, P.V.V., M. Djanaguiraman, R. Perumal, and I.A. Ciampitti. 2015. Impact of high temperature stress on floret fertility and individual grain weight of grain sorghum: sensitive stages and thresholds for temperature and duration. Front. Plant Sci. 6. doi:10.3389/fpls.2015.00820

Prasad, P.V.V., S.A. Staggenborg, Z. Ristic, L.R. Ahuja, V.R. Reddy, S.A. Saseendran, and Q. Yu. 2008. Impacts of Drought and/or Heat Stress on Physiological, Developmental, Growth, and Yield Processes of Crop Plants. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, H. Zhang, Y. Zhao, X. Wang, A. Rathore, R.K. Srivastava, A. Chitikineni, G. Fan, P. Bajaj, S. Punnuri, S.K. Gupta, H. Wang, Y. Jiang, M. Couderc, M.A.V.S.K. Katta, D.R. Paudel, K.D. Mungra, W. Chen, K.R. Harris-Shultz, V. Garg, N. Desai, D. Doddamani, N.A. Kane, J.A. Conner, A. Ghatak, P. Chaturvedi, S. Subramaniam, O.P. Yadav, C. Berthouly-Salazar, F. Hamidou, J. Wang, X. Liang, J. Clotault, H.D. Upadhyaya, P. Cubry, B. Rhoné, M.C. Gueye, R. Sunkar, C. Dupuy, F. Sparvoli, S. Cheng, R.S. Mahala, B. Singh, R.S. Yadav, E. Lyons, S.K. Datta, C.T. Hash, K.M. Devos, E. Buckler, J.L. Bennetzen, A.H. Paterson, P. Ozias-Akins, S. Grando, J. Wang, T. Mohapatra, W. Weckwerth, J.C. Reif, X. Liu, Y. Vigouroux, and X. Xu. 2017. Pearl millet genome sequence provides a resource to

improve agronomic traits in arid environments. Nat Biotech advance online publication. doi:10.1038/nbt.3943

Walker, T.S., and J. Alwang. 2015. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa. CABI.

Yoder, J.I., and J.D. Scholes. 2010. Host plant resistance to parasitic weeds; recent progress and bottlenecks. Current Opinion in Plant Biology 13:478–484. doi:10.1016/j.pbi.2010.04.011

Yohannes, T., T. Abraha, D. Kiambi, R. Folkertsma, C. Tom Hash, K. Ngugi, E. Mutitu, N. Abraha, M. Weldetsion, C. Mugoya, C.W. Masiga, and S. de Villiers. 2015. Marker-assisted introgression improves Striga resistance in an Eritrean Farmer-Preferred Sorghum Variety. Field Crops Research 173:22–29. doi:10.1016/j.fcr.2014.12.008

Table 5.1. Putative candidate SNPs for breeder-friendly marker development

Trait	Phenotype	SNP	MAF	Method	Population
Flowering	Genome scan	S6_51192692	0.38	Pairwise F_{ST}	Nigerien GRIN [a]
	Genome scan	S9_2371665	0.4	Pairwise F_{ST}	Nigerien GRIN
	Genome scan	S9_57517028	0.13	Pairwise F_{ST}	Nigerien GRIN
	WW	S4_64398335	0.07	GWAS	WASAP [b]
	Diff WW-WS	S5_52090779	0.07	GWAS	WASAP
	WS	S3_19589652	0.13	GWAS	WASAP
	WS	S4_14172212	0.08	GWAS	WASAP
Grain pigmentation	Genome scan	S4_62359202	0.34	Pairwise F_{ST}	Nigerien GRIN
	Genome scan	S4_62389180	0.13	Pairwise F_{ST}	Nigerien GRIN
Grain weight	WS	S7_50055849	0.13	GWAS	WASAP
	WS	S6_46923493	0.05	GWAS	WASAP
	WS	S6_18075344	0.18	GWAS	WASAP
Plant height	WS	S7_59459123	0.06	GWAS	WASAP
	WW	S7_59412395	0.05	GWAS	WASAP
	Diff WW-WS	S4_149131	0.04	GWAS	WASAP

	WS	S7_50055849	0.13	GWAS	WASAP
	WS	S9_56534065	0.07	GWAS	WASAP
	WW	S6_2900751	0.09	GWAS	WASAP
Panicle length	WS	S5_70062719	0.32	GWAS	WASAP
Striga resistance	*LGS1*	S5_69450954	0.42	LD analysis	West African and
	LGS1	S5_69847924	0.27	LD analysis	SAP [c]
	LGS1	S5_69852443	0.46	LD analysis	
	LGS1	S5_69879981	0.28	LD analysis	
	LGS1	S5_69882454	0.35	LD analysis	
Temperature	Annual [d]	S1_7769638	0.08	GWAS	Nigerien GRIN
	Annual	S1_7769764	0.08	GWAS	Nigerien GRIN
Vegetative biomass	WS	S7_50055849	0.14	GWAS	WASAP

Appendix A - Supplementary Material Chapter 2

This section contains supplementary materials for Chapter 2.

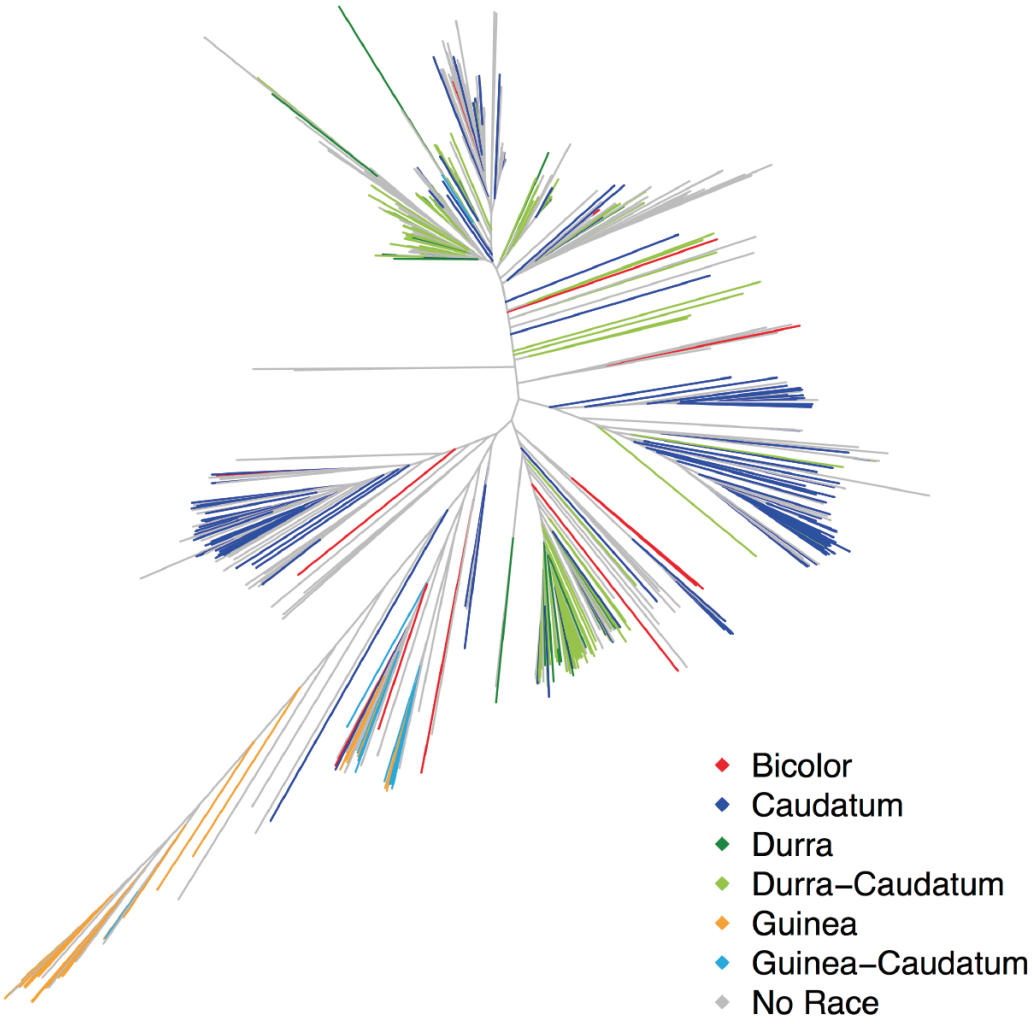


Figure A.1. Neighbor-joining tree of the Nigerien germplasm color-coded by botanical types obtained from the passport data.

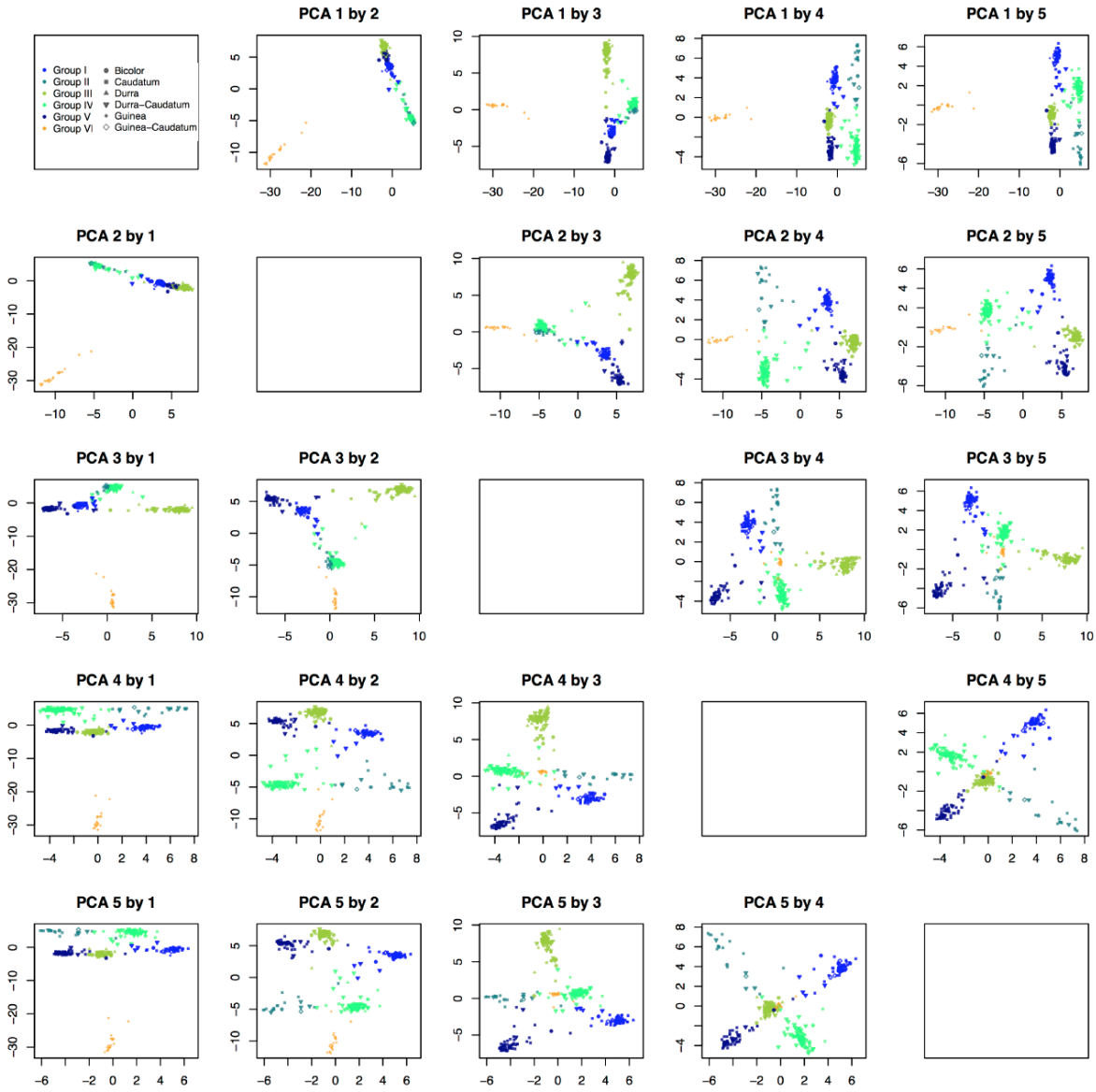


Figure A.2. Pairwise scatter plot of the five principal components based on the DAPC analysis.

PC1, PC2, PC3, PC4 and PC5 explain 44%, 18%, 15%, 13%, and 9% of the total variation respectively.

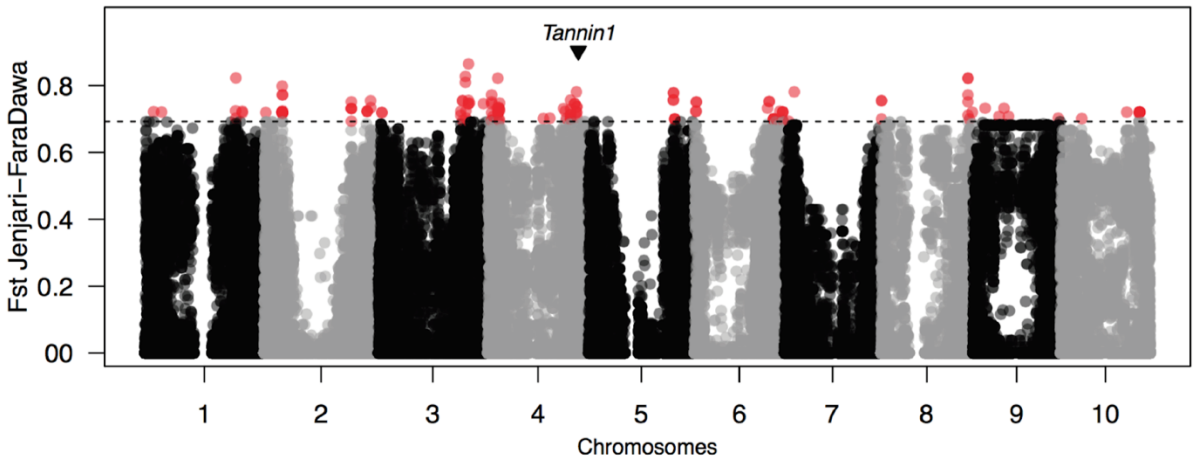


Figure A.3. F_{ST} genome scan between Jenjari and Fara Dawa local varieties.

The x-axis represents the chromosomes and the y-axis the F_{ST} values. Each point represents the SNP marker. The top 0.1% was used to set the threshold. The colors represent the significant SNPs above the threshold that differentiated the local names. F_{ST} was calculated using Weir and Cockerham's method. The highest F_{ST} value (0.86, p-value < 2.2e-16) is located on chromosome 3 (S3_61439066).

Table A.1. *A priori* candidate genes list.

Chromosome	Trait	Gene Name	Start	Stop	Gene ID
1	Flowering	SbD8	9,381,697	9,384,098	Sb01g010660
1	Flowering	LUH	155,315	162,004	Sobic.001G001900
1	Maturity	ids1/ts6	2,736,885	2,740,680	Sobic.001G036800
1	Maturity	zmm4/mads4	6,699,580	6,708,049	Sobic.001G086500
1	Maturity	PhyC/Ma5	6,748,036	6,753,421	Sobic.001G087100
1	Maturity	KN1	8,180,004	8,187,693	Sobic.001G106200
1	Flowering	PhyA	8,708,287	8,721,152	Sobic.001G111500
1	Maturity	d8	9,381,697	9,384,098	Sobic.001G120900
1	Maturity	SbEHD1	21,860,030	21,867,056	Sobic.001G227900
1	Maturity	ID1	25,113,243	25,117,555	Sobic.001G242900
1	Flowering	THE1/CRCK3	54,724,247	54,735,144	Sobic.001G279900
1	Flowering	SP1	55,301,073	55,304,219	Sobic.001G282000
1	Flowering	HAM3/LOM3	58,833,880	58,837,256	Sobic.001G304500
1	Height	GH3.2/DFL2	61,899,347	61,901,912	Sobic.001G331200
1	Height	Dwarf8	62,864,836	62,867,833	Sobic.001G341400
1	Maturity	PhyB/Ma3	68,034,103	68,043,358	Sobic.001G394400
2	Flowering	HD6	982,093	988,818	Sb02g001110
2	Maturity	sid1	8,923,187	8,929,483	Sobic.002G083600
2	Flowering	ID1	46,356,015	46,359,637	Sobic.002G153900
2	Maturity	ra1	58,696,835	58,697,696	Sobic.002G197700
2	Maturity	tsh4	63,576,564	63,580,059	Sobic.002G247800
2	Maturity	DLF1	66,203,768	66,205,133	Sobic.002G280800
2	Flowering	IDD1	71,688,818	71,692,504	Sobic.002G353700
2	Maturity	ra3	73,798,565	73,802,581	Sobic.002G381600
2	Maturity	bd1	75,943,774	75,945,236	Sobic.002G411000
2	Stay Green	Stg3a	57,146,102	61,786,369	
2	Stay Green	Stg3b	62,714,219	64,672,324	
3	Maturity	SbCN12	62,753,997	62,755,638	Sb03g034580
3	Flowering	TCP24	1,672,644	1,677,633	Sobic.003G018700

3	Maturity	GI	3,821,973	3,830,666	Sobic.003G040900
3	Maturity	vt2	4,756,775	4,761,141	Sobic.003G052700
3	Maturity	ra2	4,788,870	4,791,162	Sobic.003G052900
3	Maturity	CLV1	5,734,677	5,739,236	Sobic.003G067600
3	Stay Green	SbPIN2/Stg2	57,437,629	57,442,175	Sobic.003G235800
3	Maturity	SbCN12	62,747,946	62,749,919	Sobic.003G295300
3	Stay Green	SbPIN4/Stg1	65,308,098	65,311,355	Sobic.003G327500
3	Maturity	ba1	66,583,460	66,584,226	Sobic.003G344501
4	Maturity	SbCN2	51,451,057	51,452,704	Sobic.004G165100
4	Maturity	Hd1/TOC1	56,625,893	56,628,817	Sobic.004G216700
4	Flowering	IDD4	62,792,662	62,802,656	Sobic.004G285400
5	Flowering	Sb-TFL1	3,523,470	3,528,654	Sobic.005G038400
5	Height	gdd1/Dwarf1	60,955,631	60,961,062	Sobic.005G144600
5	Stay Green	Stg4			
6	Maturity	GHD7/Ma6	697,459	700,101	Sobic.006G004400
6	Maturity	SbPRR37/Ma1	40,304,883	40,316,799	Sobic.006G057866
6	Maturity	SbCN4	42,914,834	42,916,517	Sobic.006G068300
6	Maturity	abph1	51,253,987	51,255,100	Sobic.006G151800
6	Maturity	zfl1	55,289,735	55,293,320	Sobic.006G201600
6	Maturity	zfl2	56,217,985	56,221,715	Sobic.006G201600
6	Height	haf1	58,107,163	58,114,954	Sobic.006G240400
6	Height	Dw2	56,085,564	56,090,604	Sobic.006G067700
7	Flowering	LHY	4,745,530	4,757,815	Sobic.007G047400
7	Maturity	rel2	5,282,384	5,291,841	Sobic.007G051700
7	Height	qPHT7.1			
7	Height	Dw3	59,821,905	59,829,910	Sobic.007G163800
8	Maturity	bif2	60,472,114	60,473,745	Sobic.008G170500
9	Maturity	SbCN8	55,147,567	55,148,912	Sb09g025750
9	Maturity	Vgt1/Rap2.7	2,179,301	2,183,620	Sobic.009G024600
9	Maturity	SbCN8	54,961,462	54,963,786	Sobic.009G199900
9	Maturity	SbFL9.1	58,743,654	58,746,644	Sobic.009G249900

9	Height	*Dw1*(SbHT9.1)	57,038,653	57,041,166	Sobic.009G229800
10	Maturity	baf1	2,204,328	2,206,240	Sobic.010G027300
10	Maturity	SbCN15/Hd3a	3,499,965	3,502,278	Sobic.010G045100
10	Maturity	SbCO	12,353,901	12,355,900	Sobic.010G115800
10	Flowering	CRY2	22,156,857	22,163,945	Sobic.010G138000
10	Maturity	APO1	56,254,112	56,255,950	Sobic.010G220400
10	Maturity	td1	60,242,646	60,247,066	Sobic.010G267700

Appendix B - Supplementary Material Chapter 3

This section contains supplementary materials for Chapter 3.

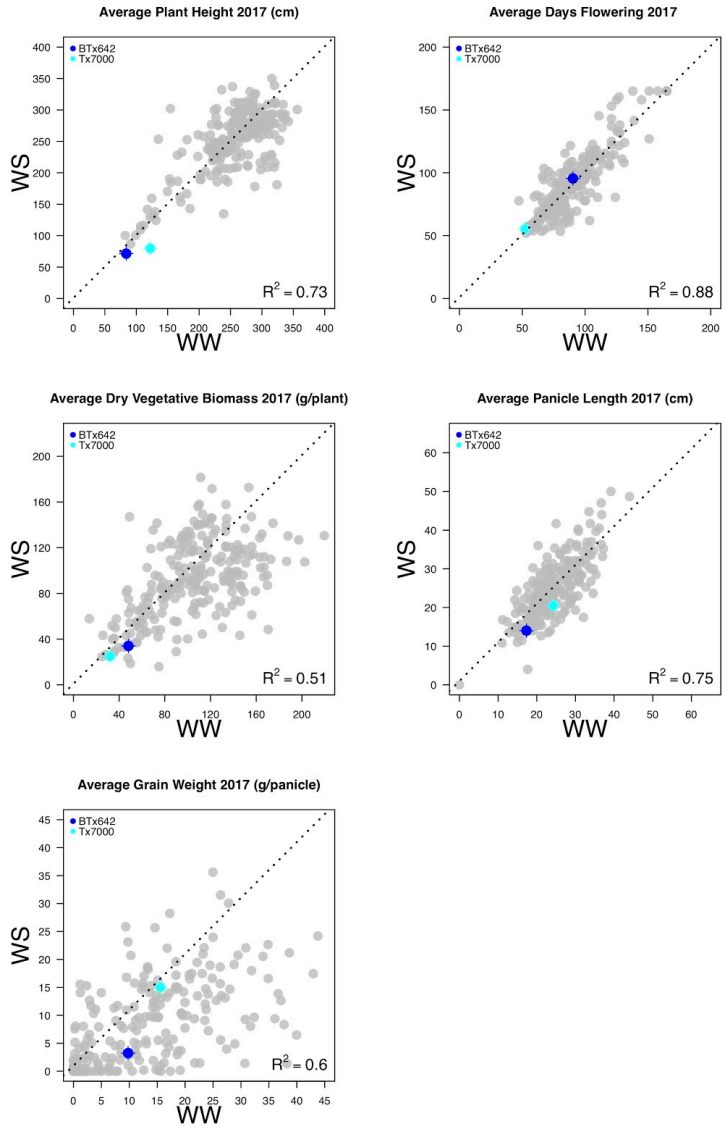


Figure B.1. (1) Correlations for plant height, days to flowering, vegetative biomass, and grain weight between WW and WS for 2017 experiments.

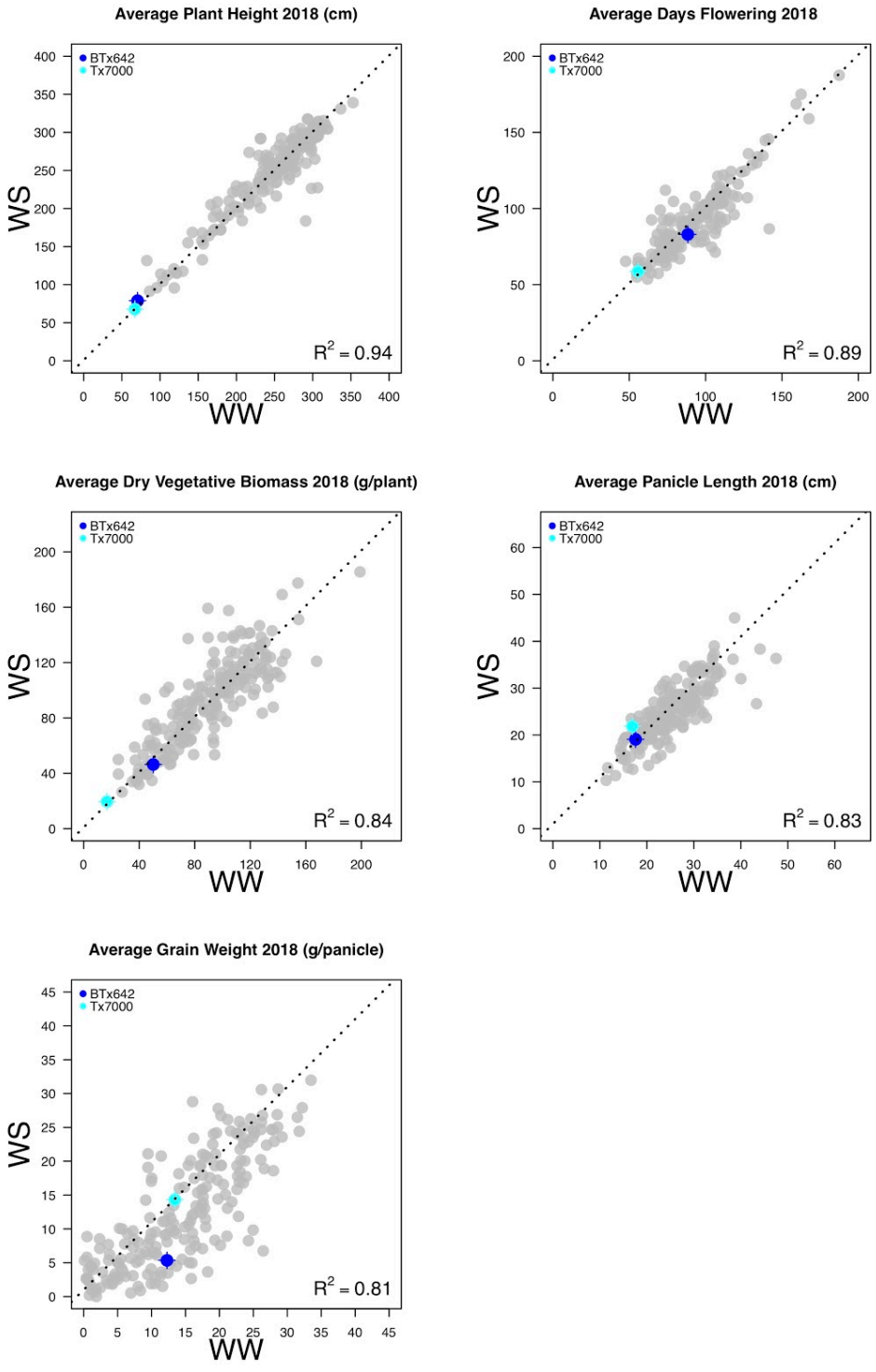


Figure B.1. (2) Correlations for plant height, days to flowering, vegetative biomass, and grain weight between WW and WS for 2018 experiments.

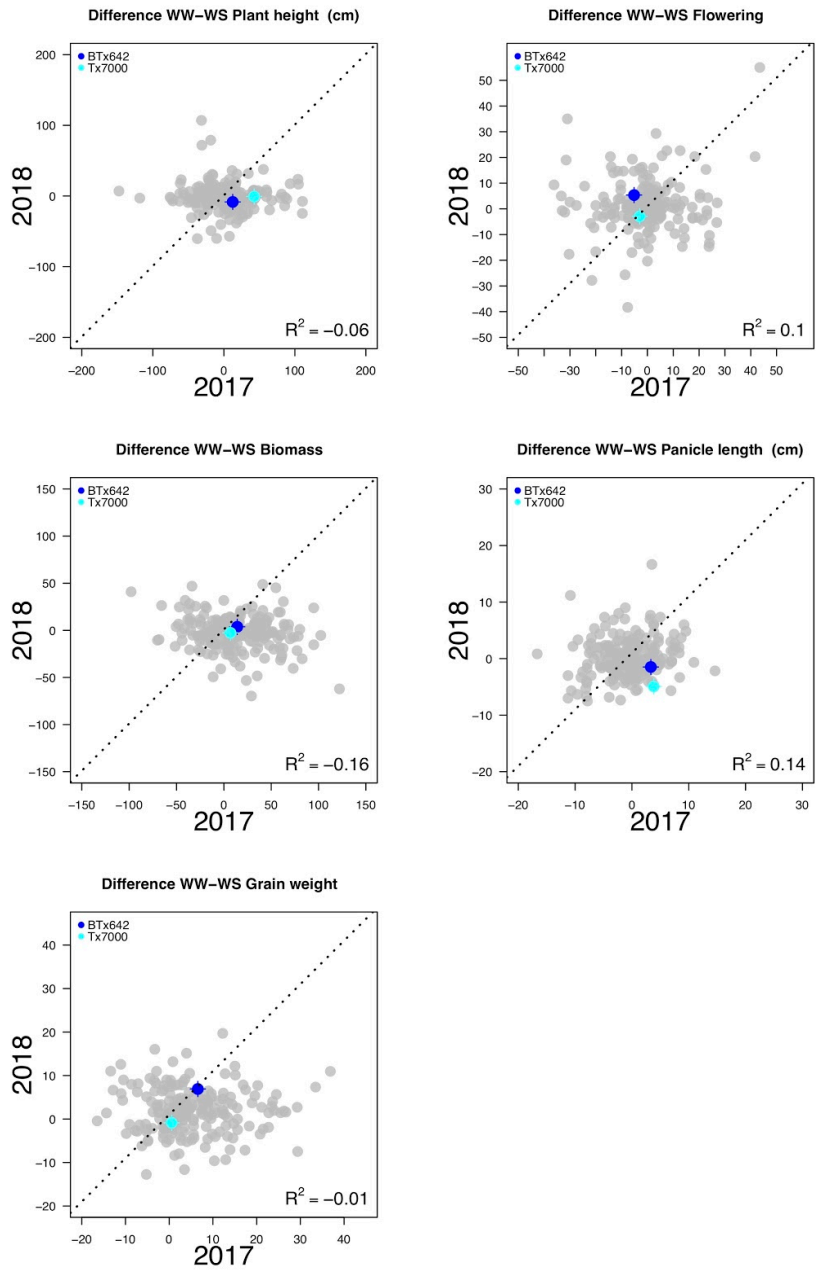


Figure B.2. Correlations of 2017 and 2018 differences between well-watered and water-stressed treatments for plant height, days to flowering, vegetative biomass, panicle length, and grain weight.

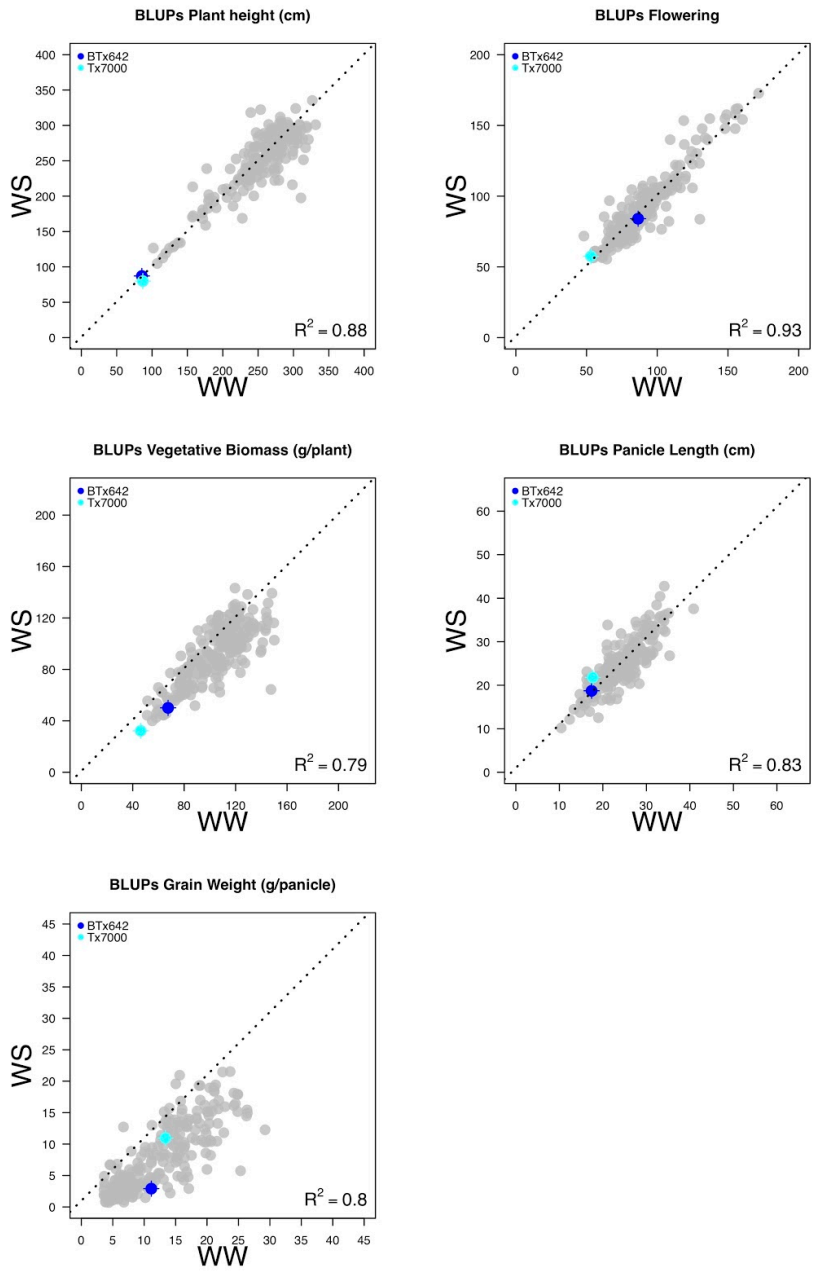


Figure B.3. Correlations of BLUP values within treatments and across years for plant height, days to flowering, vegetative biomass, panicle length, and grain weight.

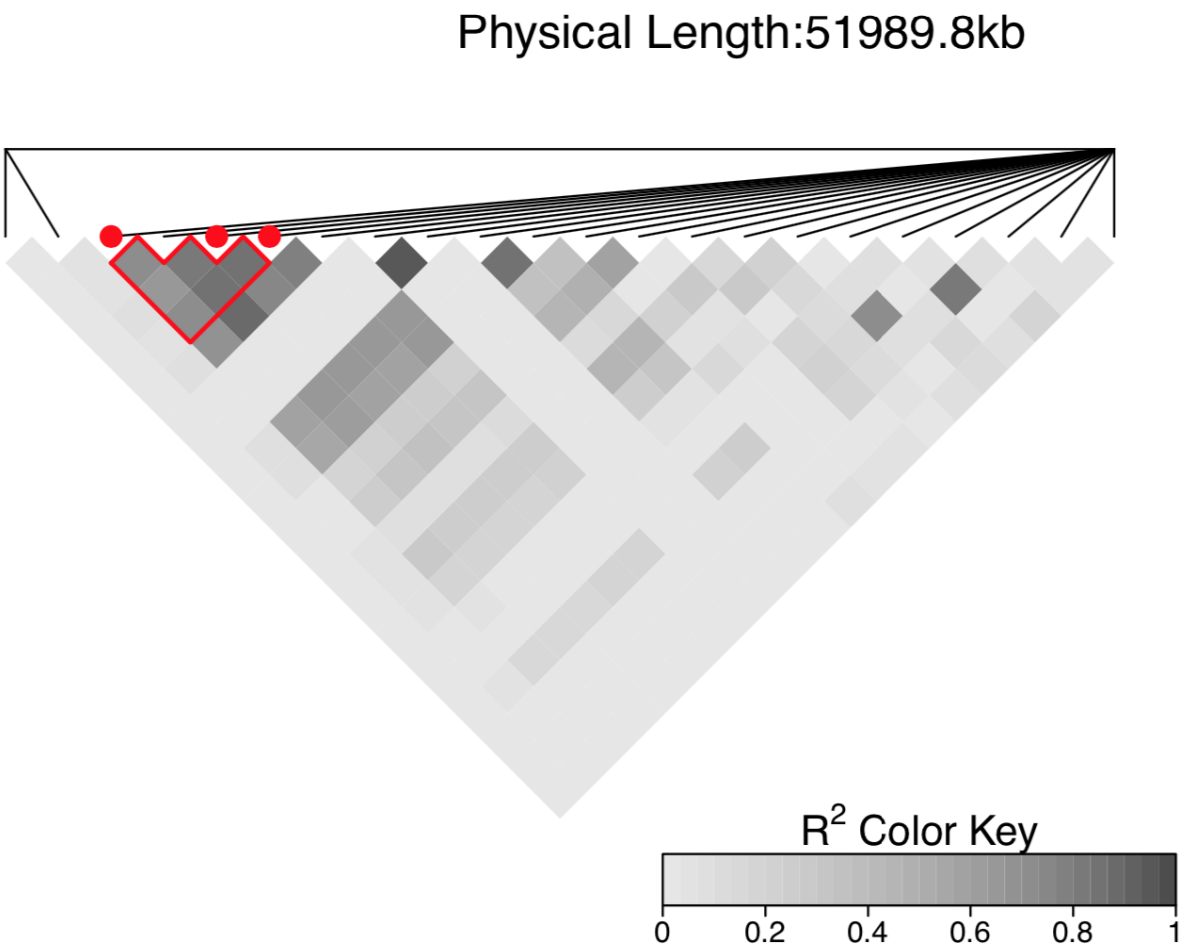


Figure B.4. Linkage disequilibrium for MTAs for vegetative biomass under water-stressed treatment.

Pairwise R^2 based heatmap on chromosome 2 for the MTAs S2_57620549, S2_57664163, and S2_57663973 represented in red points. The region selected, within the stay-green gene *Stg3a* (57,14– 61,78 Mb), is composed of 22 SNPs including the MTAs for vegetative biomass under water-stressed treatment.

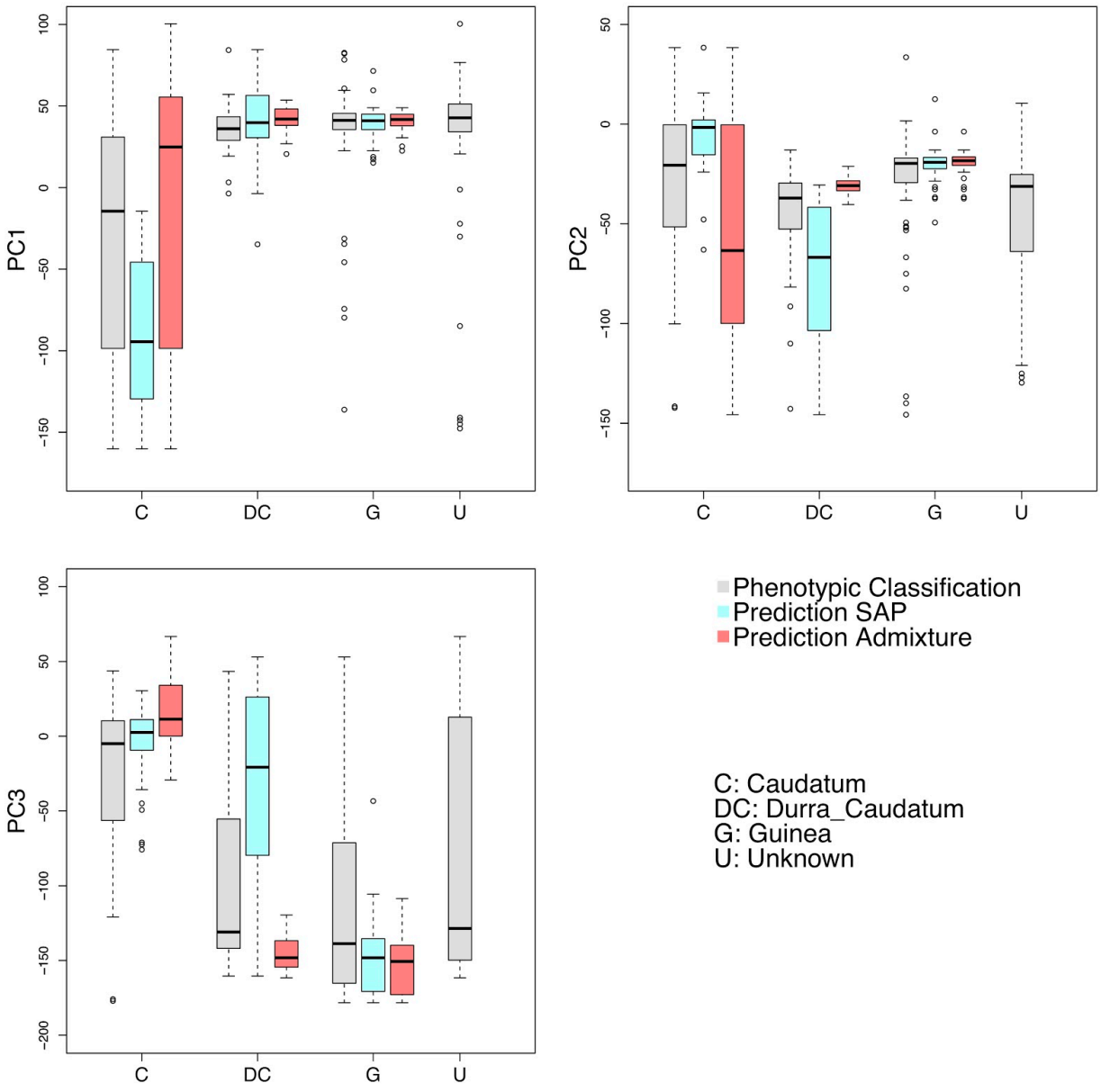


Figure B.5. Major botanical types variation in the WASAP_Lysi based on principal components of the Sorghum association Panel.

The X-axis represents the types while y-axis, principal components of Sorghum Association Panel (SAP), PC1 (A), PC2 (B), and PC3 (C). Boxplots in grey are original botanical types based

on phenotypic classification, in cyan are predicted groups based on SAP, and s in red are predicted groups based on admixture.

Table B.1. List of genotypes assayed in the study.

The list of the genotypes and sequencing data are available in Faye, Maina et al, *in prep*

Table B.2. Average performance of the genotypes based on the country of origin. 95% confidence interval was estimated based on the mean value.

	Treatment[a]	Mali	Niger	Senegal	Togo
Plant Height	WW (95% CI)	255 (13.9)	249 (7.3)	245 (18.42)	257 (24.3)
	WS (95% CI)	254 (11.47)	245 (6.7)	247 (16.7)	253 (23.5)
	WW-WS (95% CI)	1.96 (6.8)	-0.07 (8.24)	13.97 (7.4)	-2.12 (18.42)
	P-value	ns	ns	ns	ns
Days Flowering	WW (95% CI)	84 (5.1)	87 (2.5)	99 (5.9)	114 (12.3)
	WS (95% CI)	83 (5.2)	87 (2.5)	96 (5.9)	104 (13.17)
	WW-WS (95% CI)	0.87 (3)	-1 (1.5)	-2.01 (3.29)	-2.57 (6.85)
	P-value	ns	ns	ns	ns
Vegetative Biomass	WW (95% CI)	83 (8.6)	90 (4.7)	91 (9.1)	104 (14.6)
	WS (95% CI)	88.8 (7.8)	94 (4.8)	89 (9)	95 (12.58)
	WW-WS (95% CI)	3.9 (3.57)	10.01 (2.72)	20.43 (4.5)	17.98 (7.51)
	P-value	ns	1.01e-4 (***)	1.7e-3 (**)	4.7e-4 (***)
Grain Weight	WW (95% CI)	16.9 (2.05)	14.51 (1.12)	12.42 (2.65)	16.4 (4.07)
	WS (95% CI)	12.3 (2.17)	12.52 (1.18)	12.49 (2.54)	14 (3.5)
	WW-WS (95% CI)	4.6 (1.32)	4.7 (0.77)	4.26 (1.07)	4.05 (1.35)
	P-value	3.7e-4 (***)	1.3e-10 (***)	7.86e-3 (**)	1.06e-2 (*)

[a]Treatment: WW, well-watered; WS, water-stressed, WW-WS, difference between well-watered and water-stressed treatments.

Table B.3. Average performance of the genotypes based on botanical types. 95% confidence interval was estimated based on the mean value.

	Treatment[a]	Caudatum	Durra-Caudatum	Guinea
Plant Height	WW (95% CI)	233 (20.5)	259 (9.1)	281 (8.4)
	WS (95% CI)	228 (18.6)	256 (8.6)	278 (7.7)
	WW-WS (95% CI)	4.95 (6)	0.27 (3.3)	3.40 (7.78)
	P-value	ns	ns	ns
Days Flowering	WW (95% CI)	89 (5.7)	81 (6.9)	105 (8.04)
	WS (95% CI)	88 (5.5)	82 (7.1)	108 (8.6)
	WW-WS (95% CI)	-1.46 (3)	-0.46 (1.4)	-2.5 (2.85)
	P-value	ns	ns	ns
Vegetative Biomass	WW (95% CI)	94 (7.8)	102 (5.6)	115 (5.4)
	WS (95% CI)	80 (7.2)	89 (5.5)	101 (5)
	WW-WS (95% CI)	14.65 (3.1)	9.8 (2.75)	13.72 (4.1)
	P-value	1e-4 (**)	1.35e-3 (**)	3e-4 (***)
Grain Weight	WW (95% CI)	16.6 (2.2)	12.5 (1.4)	9.6 (1.2)
	WS (95% CI)	10.4 (1.9)	8.1 (1.2)	5.7 (1.09)
	WW-WS (95% CI)	5.56 (1.32)	4.65 (0.72)	3.92 (0.67)
	P-value	5.62e-05 (***)	1.41e-05 (***)	6.18e-06 (***)

[a]Treatment: WW, well-watered; WS, water-stressed, WW-WS, difference between well-watered and water-stressed treatments.

Table B.4. Marker-trait associations in the multilocus mixed linear model.

Traits	Treatment	SNP ID	MAF	-Logpvalue
Flowering	WW	S4_64398335	0.07	8.59
	WW	S8_54132955	0.06	4.85
	WW	S10_59513462	0.24	4.43
	WW	S2_26746971	0.06	4.39
	WW	S6_50716126	0.21	4.23
	WW	S4_54423729	0.05	4.21
	WW	S9_21035124	0.09	4.07
	WW	S7_57835652	0.09	4.05
	WS	S3_19589652	0.13	7.43
	WS	S4_14172212	0.08	6.87
	WS	S7_50055849	0.13	5.24
	WS	S1_75289395	0.07	5.11
	WS	S6_40060292	0.07	4.86
	WS	S6_15746352	0.23	4.83
	WS	S9_5668457	0.05	4.80

	WS	S9_49015597	0.35	4.76
Flowering	Difference (WW-WS)	S5_52090779	0.07	7.53
	Difference (WW-WS)	S6_29203686	0.07	6.06
	Difference (WW-WS)	S6_49881551	0.05	5.09
	Difference (WW-WS)	S1_49299750	0.46	4.90
	Difference (WW-WS)	S5_64932515	0.05	4.29
	Difference (WW-WS)	S7_59982074	0.35	4.12
	Difference (WW-WS)	S4_60577046	0.05	4.04
	Difference (WW-WS)	S7_32532696	0.07	3.99
Grain Weight	WW	S3_51606304	0.05	5.51
	WW	S3_16557388	0.17	4.58
	WW	S7_50055849	0.13	4.48
	WW	S1_19574771	0.09	4.42
	WW	S10_1339316	0.28	4.20
	WW	S3_22918748	0.1	4.08
	WW	S1_60422880	0.05	3.96
	WW	S4_21963875	0.13	3.93

	WS	S7_50055849	0.13	9.83
	WS	S6_46923493	0.05	6.65
Grain Weight	WS	S6_18075344	0.18	6.40
	WS	S1_58554242	0.06	4.89
	WS	S1_58557225	0.06	4.89
	WS	S9_52955476	0.38	4.54
	WS	S1_49924274	0.16	4.53
	WS	S2_32292081	0.30	4.38
	Difference (WW-WS)	S6_14260723	0.17	4.65
	Difference (WW-WS)	S1_58561343	0.14	4.26
	Difference (WW-WS)	S2_487459	0.13	4.23
	Difference (WW-WS)	S1_60898587	0.12	3.98
	Difference (WW-WS)	S6_465640	0.06	3.98
	Difference (WW-WS)	S5_16199662	0.09	3.95
	Difference (WW-WS)	S9_5525152	0.07	3.95
	Difference (WW-WS)	S9_53870340	0.11	3.93
Panicle Length	WW	S5_70062719	0.32	4.29

	WW	S2_73868064	0.07	4.22
	WW	S8_616145	0.32	4.11
	WW	S6_47319583	0.14	4.09
Panicle Length	WW	S5_2454070	0.48	4.02
	WW	S3_52408369	0.08	3.99
	WW	S6_2687200	0.22	3.84
	WW	S1_11081212	0.07	3.83
	WS	S10_57284231	0.05	6.14
	WS	S3_6603686	0.05	4.72
	WS	S5_70062719	0.32	4.58
	WS	S4_34268663	0.13	4.54
	WS	S8_882696	0.2	4.39
	WS	S3_50715050	0.13	4.35
	WS	S1_80001381	0.23	3.93
	WS	S1_80001417	0.23	3.93
	Difference (WW-WS)	S9_44566031	0.02	5.71
	Difference (WW-WS)	S9_44566032	0.02	5.71

	Difference (WW-WS)	S5_7259022	0.015	4.89
	Difference (WW-WS)	S8_58280907	0.45	4.55
	Difference (WW-WS)	S8_58112022	0.32	4.54
	Difference (WW-WS)	S8_58843141	0.05	4.53
	Difference (WW-WS)	S6_47843380	0.12	4.51
	Difference (WW-WS)	S2_73315348	0.4	4.41
	Difference (WW-WS)	S9_6790652	0.34	4.41
	Difference (WW-WS)	S6_47843433	0.12	4.35
Plant Height	WW	S7_59412395	0.05	19.31
	WW	S6_2900751	0.01	6.51
	WW	S5_65127595	0.059	5.00
	WW	S1_25832594	0.17	4.88
	WW	S10_5159486	0.11	4.70
	WW	S5_48158357	0.05	4.38
	WW	S9_43204765	0.1	4.29
	WW	S7_61574727	0.12	4.25
	WS	S7_59459123	0.06	23.27

	WS	S7_50055849	0.13	7.44
	WS	S9_56534065	0.07	6.64
	WS	S5_65128216	0.04	4.69
	WS	S4_12207762	0.22	4.27
	WS	S6_52379005	0.31	4.20
Plant Height	WS	S4_548416	0.06	4.14
	WS	S10_4711064	0.31	4.10
	Difference (WW-WS)	S4_149131	0.04	7.70
	Difference (WW-WS)	S8_12113370	0.03	5.79
	Difference (WW-WS)	S9_49741060	0.05	5.26
	Difference (WW-WS)	S4_548416	0.06	5.21
	Difference (WW-WS)	S7_39910362	0.05	4.86
	Difference (WW-WS)	S5_7637497	0.15	4.75
	Difference (WW-WS)	S7_9057897	0.04	4.73
	Difference (WW-WS)	S10_59035223	0.09	4.69
Vegetative Biomass	WW	S7_59459123	0.06	5.45
	WW	S5_67433835	0.37	5.31

	WW	S7_59393451	0.07	5.22
	WW	S7_50055849	0.13	5.19
	WW	S7_32268022	0.06	5.06
	WW	S7_62162231	0.10	4.87
	WW	S7_59412395	0.05	4.76
	WW	S2_72712959	0.19	4.71
Vegetative Biomass	WS	S7_50055849	0.13	6.14
	WS	S4_55772490	0.29	4.56
	WS	S4_55780219	0.3	4.47
	WS	S7_59459123	0.06	4.44
	WS	S5_11830740	0.09	4.42
	WS	S7_49831209	0.29	4.37
	WS	S6_40796183	0.11	4.29
	WS	S5_68814273	0.10	4.21
	Difference (WW-WS)	S7_64933026	0.09	5.21
	Difference (WW-WS)	S1_32818953	0.07	5.03
	Difference (WW-WS)	S7_10503230	0.29	4.63

Difference (WW-WS)	S2_57620549	0.4	4.37
Difference (WW-WS)	S2_57664163	0.36	4.36
Difference (WW-WS)	S8_61651261	0.07	4.35
Difference (WW-WS)	S4_60492994	0.48	4.24
Difference (WW-WS)	S2_57663973	0.37	4.20

Appendix C - Supplementary Material Chapter 4

This section contains supplementary materials for Chapter 4.

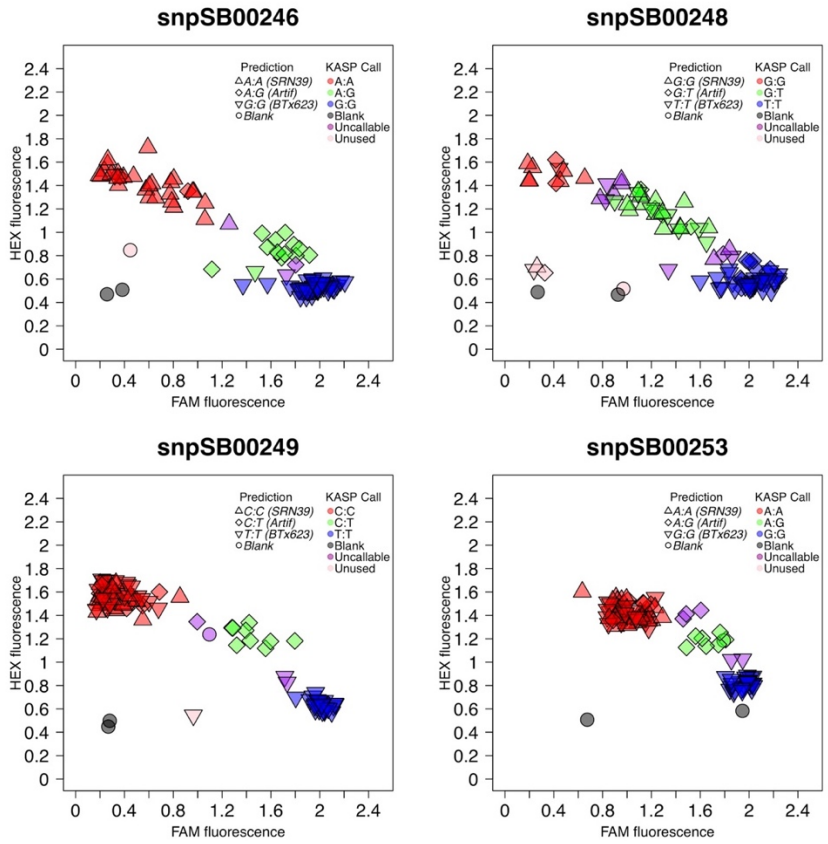


Figure C.1. KASP genotyping assays of four markers using the test plate of inbred genotypes.

Markers: snpSB00246, snpSB00248, snpSB00249, snpSB00253. Colors are assigned based on the HEX fluorescence (reference allele of SRN39) and FAM fluorescence (reference allele of BTx623). Red: homozygotes for the HEX allele; blue: homozygotes for the FAM allele; green: heterozygotes; purple: uncallable for one of the three groups; Unused: no amplification; Blank: genotyping negative control. Genotypes used: BTx623, El Mota, IRAT 204, Mota Maradi,

MR732, PI513676, PI513680, PI513749, Sepon 82, SRN39, PI656027, PI277541, PI284980, PI465448, Tx430.

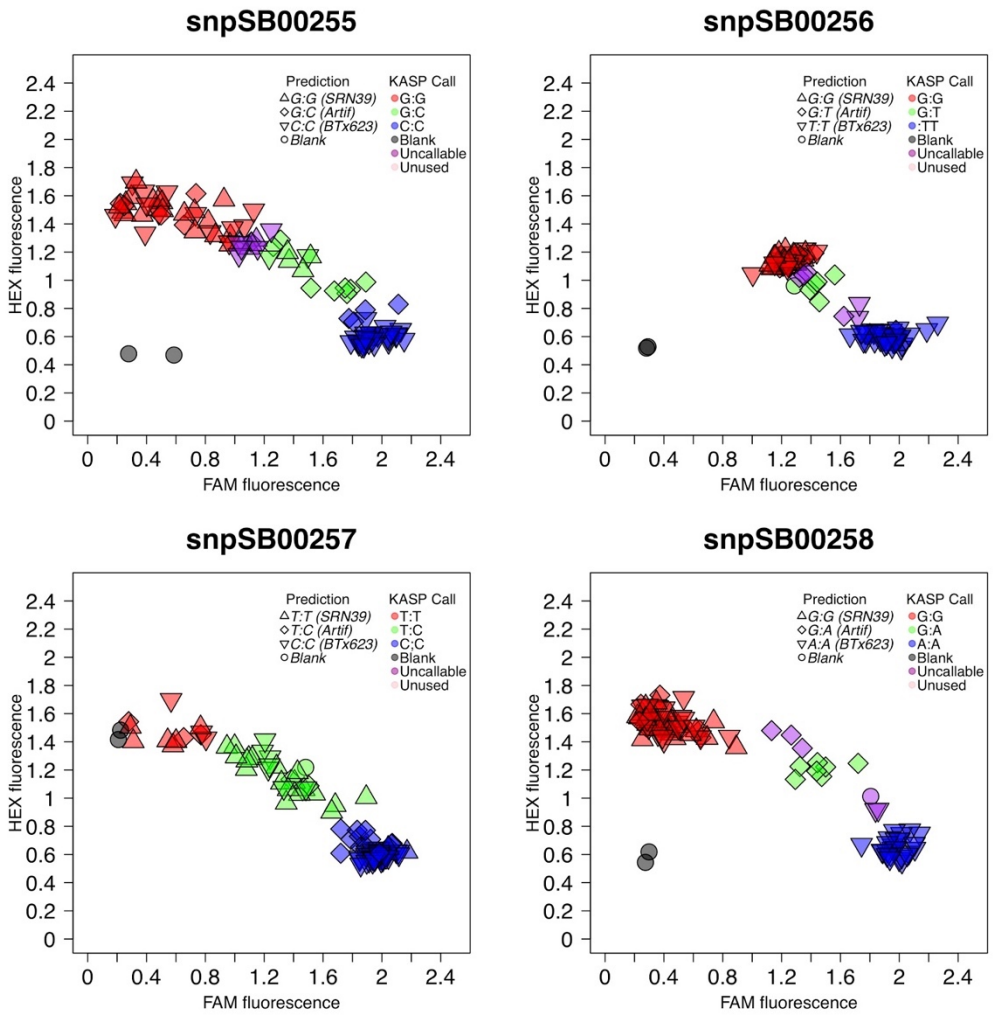


Figure C.2. KASP genotyping assays of four markers using the test plate of inbred genotypes.

Markers: snpSB00255, snpSB00256, snpSB00257, snpSB00258. Colors are assigned based on the HEX fluorescence (reference allele of SRN39) and FAM fluorescence (reference allele of BTx623). Red: homozygotes for the HEX allele; blue: homozygotes for the FAM allele; green: heterozygotes; purple: uncallable for one of the three groups; Unused: no amplification; Blank: genotyping negative control. Genotypes used: BTx623, El Mota, IRAT 204, Mota Maradi, MR732, PI513676, PI513680, PI513749, Sepon 82, SRN39, PI656027, PI277541, PI284980, PI465448, Tx430.

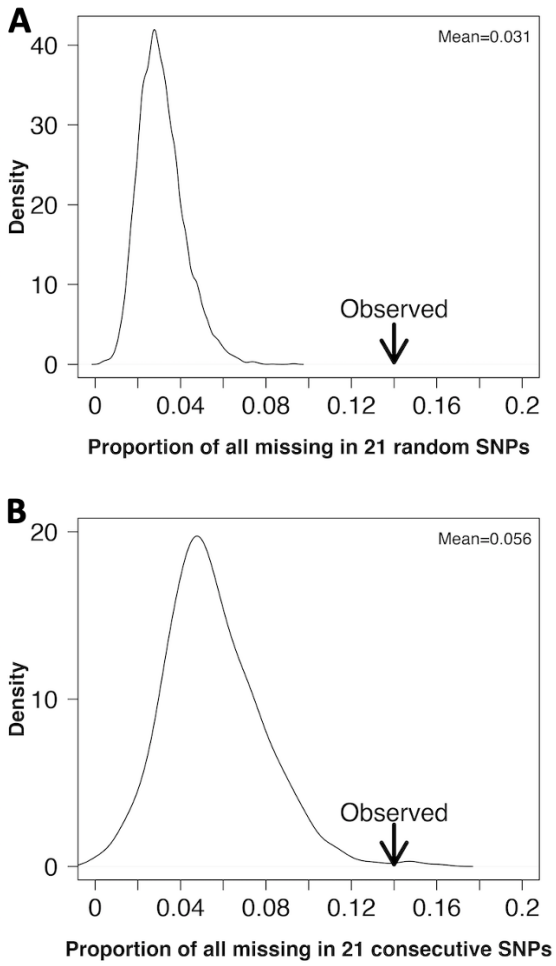


Figure C.3. Empirical distribution of missing SNPs on chromosome 5.

(A) A random 21 SNPs on chromosome 5 were selected for each of the 10,000 permutations. The average percentage of all missings in 21 random SNPs is 3%. (B) Consecutive 21 SNPs on chromosome 5 (total number of SNP = 14,307) were selected for 682 iterations (number of iterations = 14,307/21). The average percentage of all missings in 21 consecutive SNPs is 5%. In a set of West African germplasm (Niger, Nigeria, and Senegal), 14% (black arrow) of the accessions have all missing SNPs in the *LGS1* deletion region (69.97–70.011 Mb).

Table C.1: Allele-specific markers used for SNP verification in KASP genotyping.

Marker	SNP	BTx623	SRN39	Mota M	MR732	[a]SNP quality assessment
snpSB0243	S5_69191749	C	A	[b]N	A	Inconclusive
snpSB0244	S5_69944130	A	G	N	N	Bad
snpSB0245	S5_69847941	C	T	N	C	Bad
snpSB0246	S5_69450954	G	A	N	G	Medium
snpSB0247	S5_69163716	T	T	N	T	Inconclusive
snpSB0248	S5_69854897	T	G	N	G	Medium
snpSB0249	S5_69882454	T	C	N	C	Very good
snpSB0250	S5_69163798	G	T	N	G	Inconclusive
snpSB0251	S5_69725830	T	K	N	T	Inconclusive
snpSB0252	S5_70124251	C	T	T	T	Inconclusive
snpSB0253	S5_69794954	G	A	N	N	Good
snpSB0254	S5_70124285	C	G	G	G	Inconclusive
snpSB0255	S5_69851828	C	G	N	G	Good
snpSB0256	S5_69879981	T	G	G	T	Good
snpSB0257	S5_69847924	C	T	N	C	Medium
snpSB0258	S5_69852443	A	G	G	G	Very good
snpSB0259	S5_70124271	A	G	G	G	Inconclusive
snpSB0260	S5_70124265	G	A	A	A	Inconclusive
snpSB0261	S5_69742573	G	R	N	G	Inconclusive
snpSB0262	S5_69994954	A	M	N	N	Inconclusive
snpSB0263	S5_70018205	C	T	N	N	Inconclusive

[a]SNP quality assessment key: Bad: no amplification; Medium: scattered amplification, homozygotes and heterozygotes are close; Good: distinct cluster of homozygotes classes;Very

Good: three distinct clusters; Inconclusive: one compact cluster. [b]N: missing SNP not called in the GBS data.

Table C.2. Linkage disequilibrium using physical distance around the deletion of *LGS1* for selected SNP-KASP markers.

Within *LGS1* gene[a]	Position 1 (MAF)	Position 2 (MAF)	R^2	KASP Marker
Yes	69985710 (0.14)	69450954 (0.42)	0.019	snpSB0246
Yes	69985884 (0.1)	69450954 (0.42)	0.020	snpSB0246
Yes	69985912 (0.03)	69450954 (0.42)	0.008	snpSB0246
Yes	69985914 (0.12)	69450954 (0.42)	0.018	snpSB0246
	69977595 (0.12)	69450954 (0.42)	0.017	snpSB0246
	69990672 (0.02)	69450954 (0.42)	0.039	snpSB0246
	69989017 (0.08)	69882454 (0.35)	0.025	snpSB0249
	69990672 (0.02)	69882454 (0.35)	0.004	snpSB0249
	70003572 (0.01)	69882454 (0.35)	0.005	snpSB0249
	69989050 (0.05)	69794954 (0.43)	0.013	snpSB0253
	69994985 (0.02)	69794954 (0.43)	0.006	snpSB0253
Yes	69985887 (0.02)	69851828 (0.46)	0.004	snpSB0255
	70002782 (0.01)	69851828 (0.46)	0.007	snpSB0255
	69981082 (0.02)	69879981 (0.35)	0.009	snpSB0256
	69989050 (0.05)	69879981 (0.35)	0.013	snpSB0256
	69994985 (0.02)	69879981 (0.35)	0.001	snpSB0256
	69996995 (0.04)	69879981 (0.35)	0.006	snpSB0256
Yes	69985887 (0.02)	69847924 (0.46)	0.003	snpSB0257
	70002782 (0.01)	69847924 (0.46)	0.002	snpSB0257
	70003572 (0.01)	69847924 (0.46)	0.003	snpSB0257
	70002782 (0.01)	69852443 (0.46)	0.008	snpSB0258

[a]*LGS1* gene: Sobic.005G213600

Table C.3. Genotype calls of snpSB00246 for inbred genotypes with biological and technical replicates.

Genotype ID	Biological Replicate	Technical Replicate	Genotype ID	Biological Replicate	Technical Replicate
BTx623	G:G	G:G	PI656027	A:A	A:A
BTx623	G:G	G:G	PI656027	A:G	A:A
BTx623	G:G	G:G	PI656027	A:A	G:G
BTx623	G:G	G:G	PI656027	A:A	Uncallable
BTx623	G:G	G:G	PI656027	Unused	
BTx623	G:G	G:G	PI656027	G:G	Unused
BTx623	G:G	G:G	PI656027	A:G	A:A
BTx623	G:G		PI656027	A:G	
BTx623	G:G		PI656027	Unused	
BTx623	G:G		PI656027	A:G	G:G
BTx623	G:G	G:G	PI656027	G:G	
BTx623	G:G	G:G	PI656027	A:G	Unused
BTx623	G:G		PI656027	Uncallable	A:A
BTx623	G:G		PI656027	G:G	
BTx623	G:G		PI656027	Unused	
BTx623	G:G		PI656027	Unused	
BTx623	G:G		PI656027	A:A	
BTx623	G:G		PI656027	A:G	
BTx623	G:G		PI656027	Unused	
BTx623	G:G		PI656027	G:G	
BTx623	G:G		PI656027	Unused	
BTx623	G:G		PI656027	Uncallable	

Table C.3. Genotype calls of snpSB00246 for inbred genotypes with biological and technical replicates. (Cont.)

Genotype ID	Biological Replicate	Technical Replicate	Genotype ID	Biological Replicate	Technical Replicate
BTx623	G:G		PI656027	A:G	
BTx623	G:G		PI656027	A:G	
EL_MOTA	G:G	G:G	MOTA_MARADI	G:G	G:G
EL_MOTA	G:G	G:G	MOTA_MARADI	G:G	G:G
EL_MOTA	G:G	G:G	MOTA_MARADI	G:G	G:G
EL_MOTA	G:G	G:G	MOTA_MARADI	G:G	G:G
EL_MOTA	G:G	G:G	MOTA_MARADI	G:G	G:G
EL_MOTA	G:G		MOTA_MARADI	G:G	
EL_MOTA	Unused		MOTA_MARADI	G:G	
EL_MOTA	Unused		MOTA_MARADI	G:G	
EL_MOTA	G:G		MOTA_MARADI	G:G	
IRAT204	G:G	G:G	PI277541	Uncallable	A:A
IRAT204	G:G	G:G	PI277541	A:G	A:G
IRAT204	G:G	G:G	PI277541	A:A	
IRAT204	G:G	G:G	PI284980	A:G	A:A
IRAT204	G:G	G:G	PI284980	A:G	Uncallable
IRAT204	G:G		PI284980	G:G	
IRAT204	G:G		PI465448	A:A	A:A
IRAT204	G:G		PI465448	A:A	
IRAT204	G:G		PI465448	A:A	A:A
SRN39	Unused	A:A	PI563137	A:A	Uncallable

SRN39	A:G	Uncallable	PI563137	A:A	Unused
SRN39	A:A	A:G	PI563137	A:A	
SRN39	A:A	A:A			
SRN39	A:A	Unused			
SRN39	A:A				
SRN39	Unused				
SRN39	A:G				
SRN39	Unused				