IDENTIFICATION AND CHARACTERIZATION OF GENES CONTROLLING THE ALKALI SPREADING PHENOTYPE IN SORGHUM AND THEIR IMPACT ON STARCH QUALITY

by

Stefanie Griebel

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agronomy

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mitchell R. Tuinstra, Chair
College of Agriculture, Department of Agronomy
Dr. Clifford F. Weil
College of Agriculture, Department of Agronomy
Dr. Bruce A. Craig
College of Science, Department of Statistics
Dr. Bruce R. Hamaker
College of Agriculture, Department of Food Science

Approved by:

Dr. Ron Turco
Head of the Graduate Program

*

For my beloved Dad Dieter and beloved Mom Brigitte

Für meinen geliebten Papa Dieter und geliebte Mama Brigitte

*

For my beloved brother Tobias and my grandparents

Für meinen geliebten Bruder Tobias, Oma und Opa

*

ACKNOWLEDGMENTS

I deeply appreciate the guidance, support and direction from my major advisor Prof. Dr. Mitchell Tuinstra during my time at Purdue University. He was always there when needed and shaped my future career goals. He gave me the desired and great opportunity to work on sorghum breeding for human nutrition and health related to global food security and African development.

I thank Prof. Dr. Osvaldo Campanella and my Ph.D. advisory team: Prof. Dr. Clifford Weil, Prof. Dr. Bruce Craig, and Prof. Dr. Bruce Hamaker, for their invaluable guidance and time for many individual meetings to make a great project successful.

I am thankful to Dr. Gregor Koliantz, Prof. Dr. Lee Schweitzer, Dr. Michael Mashtare, my major advisor Prof. Dr. Mitchell Tuinstra and Karen Neubauer, for their great support on building my teaching expertise at Purdue and obtaining the Advanced Graduate Teaching Certificate.

My research was made possible by conducting research in several labs and I am thankful for the members of the Prof. Tuinstra Lab with Andy Linvill and Eugene Glover, the Prof. Ejeta lab with Prof. Dr. Gebisa Ejeta, Dr. Patrick Rich and Dr. Adedayo Adeyanju, the Prof. Volenec Lab with Suzanne Cunningham, the Prof. Campanella Lab, Prof. Hamaker Lab and Prof. Weil Lab. Also thank you to Charles Addo-Quaye for great long-distance support and Richard P. Westerman. And a long list of people who I could not put in here due to space constrains.

Thank you to Prof. Dr. Richard Grant, Prof. Dr. Joseph Anderson, Prof. Dr. Linda Lee, Connie Foster, Jenny Kelly, Cheryl Long, Karen Clymer and Patrica Oliver for valuable administrative support. Thank you to Prof. Dr. George and Jane Van Scoyoc and Prof. Dr. John and Sheila Graveel for their hospitality.

Many thanks to the Department of Agronomy, which funded me over five semesters working as a teaching assistant. I further acknowledge USAID and the Sorghum and Millet Innovation Lab, who funded my PhD research and several conferences. This study is made possible by the generous support of the American People through the United States Agency for International Development

(USAID). The contents are the responsibility of Stefanie Griebel and do not necessarily reflect the views of USAID or the United States Government.

My family and friends from Germany, Uganda, Haiti, USA and Zambia for your invaluable encouragement, uncountable hours of phone calls, coffees at Lavazza and prayers during my time in the USA.

My deepest thanks to all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Griebel, Stefanie. PhD
Institution: Purdue University
Degree Received: May 2019
Title: Identification and Characterization of Genes Controlling the Alkali Spreading Phenotype in Sorghum and Their Impact on Starch Quality
Committee Chair: Mitchell Tuinstra

Sorghum [*Sorghum bicolor* (L.) Moench] is a staple food for millions of people in Africa and South Asia. It is mainly consumed for its starch. The starch composition and structure in the seed endosperm determines cooking properties, processing quality, and starch digestibility.

An assay to measure the alkali spreading value (ASV) of sorghum is described. The assay was used to identify sorghum EMS mutants with variation in starch composition. The ASV mutants (ASV+) exhibited a range of starch thermal properties with starch gelatinization temperatures (GT) being lower or higher than samples from Tx623 or Sepon82. The ASV+ phenotypes were found to be correlated with starch related traits such as enthalpy ($r = -0.53$) and range of starch GT (T_c-T_o) ($r = 0.60$).

Genes controling the ASV phenotype of sorghum and their impact on starch quality traits are described. Whole genome re-sequencing of sorghum EMS mutants exhibiting an ASV+ phenotype was used to identify single nucleotide polymorphisms (SNPs) in candidate genes *Sobic.004G163700* and *Sobic.010G093400*. The two genes were identified as a *SbeIIb*, a putative sorghum homolog of *amylose extender*, and as a *SSIIa*, respectively. Linkage analysis showed that the mutations in *Sobic.010G093400* and *Sobic.004G163700* co-segregated with the ASV phenotype. The *ssIIa*-mutants exhibited normal amylose values, lower starch GT and lower final viscosity than the wild type. The *sbeIIb*-mutants exhibited higher amylose content, higher starch GT and lower peak and final viscosity with poor gel consistency compared to the wild type and

ssIIa-mutants. An allele dosage test indicated that the *sbeIIb*-mutants had an allele dosage dependent effect on amylose content. Double mutants of *sbeIIb* and *ssIIa* showed that amylose content, starch thermal properties and paste viscosity profiles resemble the *sbeIIb* parent.

A study of ASV phenotypes in a panel of more than 750 sorghum conversion lines revealed genetic variation for the ASV phenotype. A few SC-lines exhibiting stable expression of the ASV+ phenotype over two growing seasons. Most of these lines were described as belonging to the working group Nandyal, durra types from India described as producing 'glutinous grains'. Whole genome resequencing discovered common SNPs in genes associated with starch biosynthesis. A genome wide association study (GWAS) identified a significant SNP that could be associated with the starch biosynthesis gene *Sobic.010G273800,* and with candidate genes *Sobic.010G274800* and *Sobic.010G275001* both annotated as glucosyltransferases. Grain samples from SC489, SC491, SC587 and SC589 exhibited a consistent ASV+ phenotype with lower or similar starch GT, similar amylose content, and similar high viscosity and gel consistency compared to controls.

CHAPTER 1. LITERATURE REVIEW

1.1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a staple food for millions of people in Africa and South Asia (House et al., 2000). Sorghum is mainly consumed for its starch. The starch composition and structure in the seed endosperm determines cooking properties, processing quality, and digestibility (Tetlow and Emes, 2014). This chapter reviews mainly starch biosynthesis and metabolism and the genes that are known to influence key starch quality traits. .

1.2 Importance of Sorghum bicolor for global food security

Sorghum [*Sorghum bicolor (L.) Moench*] is a C4 cereal (Kimber et al., 2013; Paterson, 2013) and is now widely grown around the world as an annual crop. Sorghum is commercially important in semiarid tropical and dry temperate environments (Kimber, 2000) in Africa, India, China, Europe, Australia and Northern America. However, sorghum is mostly grown on the African continent and across the world mainly in low-income, food-deficient countries (FAOSTAT, 2016, 2019). Sorghum is an important carbohydrate source. The utilization of sorghum is regionally dependent and concentrated on animal feed in the western world (House et al., 2000) and for human consumption in Africa and Asia (House et al., 2000). Traditional African foods made from sorghum are thick and thin porridges, fermented breads, breads and biscuits from flour blends (max 50% sorghum flour), cooked products, boiled whole or pearled sorghum, sorghum malt and alcoholic beverages such as beer (Rooney and Waniska, 2000; House et al., 2000). Sorghum is an indigenous crop and has great economic value for smallholder farmers in Africa, because it is less expensive than other imported cereals (Taylor and Dewar, 2000; House et al., 2000). The malting and brewing industries are key users and contribute substantially to local

rural economies (House et al., 2000). In many countries, sorghum is valued as a gluten free grain (NuLifeMarket, 2019) for humans with gluten allergies.

1.3 Sorghum's Genetic Diversity and Genomics

1.3.1 Sorghum Phenotypic Diversity – Race Classification

The African continent provides the greatest diversity of sorghum germplasm (Kimber et al., 2013). The genus Sorghum consists of three species, *S. halepense* (wild sorghum), *S. propinquum* (wild sorghum) and *S. bicolor* (wild and cultivated sorghums), *which* diversified during domestication (Dahlberg, 2000). There are three subspecies of *S. bicolor*, namely *S. bicolor subsp. bicolor*, *S. bicolor subsp. drummondii* and *S. bicolor subsp. verticilliflorum* (Dahlberg, 2000). The cultivated sorghums belong to the subspecies *S. bicolor subsp. bicolor* and can be classified into five races based on plant - panicle morphologies (Dahlberg, 2000; Harlan and de Wet, 1972). The five basic cultivated races are bicolor, guinea, kafir, caudatum and durra, which can combine into intermediate races (Dahlberg, 2000; Harlan and de Wet, 1972; Kimber et al. 2013). Sorghums were first classified by Snowden in 1936, who's classification was modified over decades by the scientists Murty and Govil, Harlan, de Wet and Dahlberg (Dahlberg, 2000; Harlan and de Wet, 1972). Race designations are used to further sub divide each race and intermediate races into working groups (Dahlberg, 2000; Harlan and de Wet, 1972).

The race bicolor is most closely related to the wild species and exhibits open panicles, small seeds, low yields (Harlan and de Wet, 1972; Dahlberg, 2000), high pigmentation, and of a medium height (Dahlberg, 2000). There is no distinct geographical region for bicolor types and they occur across Africa, India and other parts of Asia (Kimber et al., 2013).

The other races of sorghum developed either out of the bicolor race or are results of crosses between bicolor types and wild sorghums (Dahlberg, 2000; Kimber et al., 2013). The possible

introgression of bicolor and wild races formed the race durra (Dahlberg, 2000), a very important and predominant race for farmers in Ethiopia, India, Northern Africa and the Arabic world (Harlan and de Wet, 1972; Kimber et al., 2013). Durras are well adapted to very dry regions (Harlan and de Wet, 1972; Dahlberg, 2000; Kimber et al., 2013), such as Sudan and India (Dahlberg, 2000) so that the densely packed panicles are not exposed to wet conditions causing mold (Kimber et al., 2013; Harlan and de Wet, 1972). Durras are long or short-season sorghums, where the short-season ones are well known as Milo types from Africa and Nandyal types from India (Harlan and de Wet 1972). The race durra is characterized by white seeds (Kimber et al., 2013), smaller, ovate, stiff and dense panicles in comparison to the bicolor race (Dahlberg, 2000).

The guinea race dominates in West Africa and the Savannas of less dry zones as it also can be grown in high rainfall environments with good yield potential (Harlan and de Wet, 1972; Dahlberg 2000). The race is also important in South Asia, southern Africa (Dahlberg, 2000) and the high rainfall mountain ranges in Eastern Africa (Kimber et al., 2013). It is assumed that guineas further developed from selections of the bicolors in West Africa (Dahlberg, 2000) and likely by introgression of a wild sorghum (Kimber et al., 2013). Guineas are characterized with open panicles to dry quickly in high rainfall regions and ensures good storage ability (Harlan and de Wet, 1972).

The caudatum race is very important for plant breeders and smallholder farmer. Open panicle types can be grown in higher rainfall areas, and densely packed panicles in drier areas (Harlan and de Wet, 1972). The race is associated with high yields and good seed quality (Dahlberg, 2000). Caudatums probably developed out of a cross from bicolor and a wild sorghum on the African continent (Dahlberg, 2000; Kimber et al., 2013). The caudatums are of importance in middle and Eastern-Africa such as in Nigeria, Chad, Sudan, Ethiopia and Uganda (Dahlberg, 2000).

The kafir race developed out of bicolor sorghum types (Kimber et al., 2013) and seems to be more important as part of an intermediate race than as a race alone (Harlan and de Wet, 1972). The kafir types are most important in southern Africa (Dahlberg, 2000) and other temperate environments.

Intermediate races such as the kafir-caudatum are important in the USA sorghum breeding programs, guinea-kafir types are a dominant race in India, while guinea-caudatum are important in Nigeria, Chad and Sudan (Harlan and de Wet, 1972).

1.3.2 Sorghum Genomic Diversity

The sorghum race classification system is based on botanical descriptors (Dahlberg, 2000; Harlan and de Wet, 1972). Genomic studies confirmed the classification system of five sorghum races and intermediate races originating from Africa and Asia (Morris et al., 2013; Deu et al., 2006; Brenton et al., 2016; Billot et al., 2013; Sukumaran et al., 2012); with northeast Africa as the center of diversity (Paterson et al., 2009). The studies confirmed that the race durra dominates in dry areas of Arabia, India and Africa, the race guinea occurs mainly in West Africa and some in Southern Africa, the race kafir dominates in Southern Africa, and the race caudatum occurs mainly in Central and Eastern Africa (Morris et al., 2013; Billot et al., 2013; Sukumaran et al., 2012). The race bicolor was the most diverse and was distributed across geographic regions (Morris et al., 2013; Deu et al. 2006; Brenton et al. 2016; Billot et a. 2013). The bicolors were confirmed to be the oldest and most primitive of the races (Deu et al., 2006).

Sorghum bicolor subsp. bicolor is a diploid with 2n=20 (Rooney, 2000; Paterson et al, 2009). The sorghum genome is relatively small (Paterson et al. 2013) with a genome size of approximately 730Mb, based on a genome assembly of Tx623 (Paterson et al., 2009). Tx623 is a breeding line

and the standard sorghum reference genome used in genomic studies (Paterson, 2013; Morris et al., 2013).

Sorghum and rice are closely related. Both share a whole genome duplication event approximately 70 Mya (Paterson et al., 2004) and diverged approximately 42Mya (Paterson, 2013). Rice and sorghum have had no other duplication and thus have fewer duplicated genes than maize (Paterson, 2013). Maize diverged from sorghum approximately 11.9 Mya (Swigonova et al., 2004) therefore, maize and sorghum are more closely related than sorghum and rice (Paterson 2013). The three species developed and specialized separately but due to common ancestors have conserved chromosome regions, which makes comparative genomics a valuable tool to compare gene sequences and putative functions across grass species (Paterson et al., 2004, Paterson et al., 2009).

1.4 Germplasm Sources for Gene Identification in Plants

1.4.1 The Sorghum Conversion Program and Sorghum Diversity Panels

The second largest collection of sorghum accessions (approximately 36,000 accessions; Kimber et al., 2013) is stored at ICRISAT, one of the centers of the Consultative Group for International Agricultural Research (CGIAR) (Billot et al., 2013; Kimber et al., 2013). The US Department of Agriculture's Germplasm Repository Information Network (GRIN) stores more than 20.000 accessions (USDA ARS, 2019) and the USA National Center for Genetic Resources Preservation (USDA-ARS-PGRCU) more than approximately 40,000 accessions (Kimber et al., 2013). These collections bear great genetic potential to be elaborated.

Sorghum improvement often begins with efforts to improve adaptation of exotic and imported sorghums to a particular geographic region (Stephens et al., 1967; Kimber et al., 2013). The largest number of accessions in the sorghum collections are from the tropics and are photoperiod sensitive. In order to be used easily in temperate environments, these accessions

require adaptation to long day environments (Stephens et al., 1967; Kimber et al., 2013). The Sorghum Conversion Program was first described by Stephens et al. (1967) and followed up by Rosenow et al. (1997a, 1997b). In the conversion program, exotic sorghum accessions were crossed as the pollinator to a 4-dwarf, early-flowering line from the USA under short-day conditions in Puerto Rico (Stephens et al., 1967). The resulting hybrids were also propagated in Puerto Rio and F_2 progeny were sent to Chillicothe, Texas for field evaluation under long-day conditions. Dwarf, early-flowering F_2 plants were selected and the F_3 progeny were sent to Puerto Rico for backcrossing with the alien sorghum as the recurrent parent. This process was repeated for four generations of back crossing (Stephens et al., 1967; Rosenow et al., 1997a; Rosenow et al., 1997b) except for the final backcross when the alien sorghum was used as the female parent to bring the converted line back into its original cytoplasm (Stephens et al., 1967; Rosenow et al., 1997a; Rosenow et al., 1997b). These sorghum conversion lines (SC lines) (Stephens et al., 1967) capture the genotype of the original alien sorghum in a dwarf and photoperiod insensitive idiotype. At present, more than 700 converted sorghum lines are available (Hayes et al., 2015; Kimber et al., 2013). The collection of SC-lines are the basis for sorghum diversity panels used in several research studies in the USA and Europe, such as the Sorghum Association Panel (SAP), including approximately 228 SC lines (Casa et al., 2008, Morris et al., 2013; Cuevas et al., 2017; Boyles et al., 2017; Adeyanju et al., 2015; Mace et al., 2013; Sukumaran et al., 2012; Shenstone et al., 2018) and the Sorghum Conversion Panel with approximately 788 SC lines (Hayes et al. 2015). Other sorghum diversity panels and core collections are also available (Deu et al., 2006; Morris et al., 2013; Billot et al., 2013; Brenton et al., 2016; Cuevas et al., 2017; Kimber et al., 2013; Mace et al., 2013).

Sorghum Diversity Panels are often used in Genome Wide Association Studies (GWAS) to study complex traits (Morris et al., 2013, Boyles et al., 2017; Adeyanju et al., 2015) and identify novel QTLs. GWAS uses genomic markers such as those created from genotyping-by-sequencing (GBS) that are spread across the entire genome (Morris et al., 2013; Boyles et al., 2017), not providing a whole genome sequence but a reduced representation of the genome (Torkamaneh et al., 2016; Elshire et al., 2011). This provides the opportunity to pursue high throughput genotyping of many individuals at low coverage and low cost (Glaubitz et al., 2014). Association studies depend on linkage disequilibrium (LD) in the population, which is influenced by population structure, recombination and selection among others (Yu and Buckler, 2006). The success of GWAS depends on the LD between the marker and causative polymorphism and based on rapidity of LD decay, the marker density needs to be adjusted so that, ideally, one GBS marker is to be found in LD with a causal polymorphism (Yu and Buckler, 2006; Glaubitz et al., 2014).

The sorghum reference genome from Tx623 (Paterson et al., 2009; Morris et al., 2013; Paterson, 2013) is commonly used to develop GBS data for GWAS in sorghum (Morris et al., 2013). The LD decay rate is reported at the point where the maximum LD (r^2 = average pairwise correlation coefficient, Huang et al., 2010) reaches half of its value (Mace et al., 2013; Huang et al., 2010) and provides the average distance of LD decay (Mace et al., 2013). In sorghum, LD decay is reported within approximately 15kb (Kimber et al., 2013) with slower LD decay in inbreds (19.7kb) than in landraces (10.3kb) (Mace et al., 2013). In contrast, a large set of more than 900 sorghum landraces and conversion lines showed LD decay of approximately 1kb (Morris et al., 2013), however these studies report background noises up to approximately 150kb (Mace et al., 2013; Morris et al., 2013). For comparison, in maize, the LD decay is faster with approximately 2kb (Kimber et al., 2013), while in rice LD decay occurs within approximately 123-167kb, varying

by chromosomes (Huang et al., 2010). Slow LD decay like in rice is not really suitable for GWAS as the resolution is too low to identify single gene candidates (Huang et al., 2012; Huang et al., 2010). A faster LD decay provides evidence of smaller haplotype blocks across the genome, so that the window of significant GWAS signals and candidate genes is smaller.Sorghum thus has a suitable LD decay for GWAS (Kimber et al., 2013).

The software GAPIT (Genome Association and Prediction Integrated Tool) is often used for GWAS (Lipka et al., 2012, Boyles et al., 2017, Morris et al., 2013, Adeyanju et al., 2015) as it provides high prediction accuracy and is computationally fast (Lipka et al., 2012). The statistical mixed models from GAPIT associate the phenotypes with genotype-by-sequencing data to identify novel QTLs while accounting for kinship and population structure (Lipka et al., 2012). In association studies, the map resolution is higher when historical and evolutionary recombination events are taken into account, which is different than in structured populations used for linkage analysis based on family pedigrees (Yu and Buckler, 2006). Different types of mixed linear models are used for GWAS (Huang et al., 2010; Zhao et al., 2011; Morris et al., 2013; Adeyanju et al., 2015).

The Logistic Mixed Model (LMM) was recently proposed for the analysis of binary traits in diversity panels (Shenstone et al., 2018). Population size of the diversity panel is critical in association analyses and, ideally, large panels are required to have enough power and genetic diversity to detect effects of genes in the genome reliably (Huang et al., 2012) and to ensure enough statistical power to detect marker associations with causal variants (Sham et al. 2014). GWAS generally enable predictions of candidate genes that require further evaluation.

1.4.2 Mutagenized EMS Population

Another strategy to identify genes in plants is through the creation of novel mutants using methods such as ethyl methane sulfonate (EMS) treatments of seeds to produce populations of chemically induced variants that can be used in forward or reverse genetic analyses (Greene et al., 2003; Addo-Quaye et al., 2018; Addo-Quaye et al., 2017; Jiao et al., 2016). EMS populations generally represent thousands of novel genetic variants not represented in the standing variation of a species (Jiao et al., 2016). In sorghum, the reference genome, Tx623, has been mutagenized by chemical mutagenesis (Jiao et al., 2016; Addo-Quaye et al., 2017; Addo-Quaye et al., 2018). EMS creates single base mutations across the genome (Westergaard, 1957; Loveless 1958; Greene et al. 2003). Typical SNP mutations are changes of G/C to A/T (Greene et al., 2003). In Arabidopsis, approximately 5% of EMS mutations result in truncation mutations, usually either nonsense mutations or splice site changes resulting in mRNA/protein truncation or loss (Greene et al., 2003). Next generation sequencing analysis can be used to identify SNPs in mutant genotypes of an EMS population (Addo-Quaye et al. 2018, Addo-Quaye et al. 2017). Functional annotations of identified SNPs are created using comparative genomics and SNPeff (Addo-Quaye et al. 2018, Addo-Quaye et al. 2017).

1.4.3 Follow up from GWAS and EMS Mutagenesis

Several methods are available to characterize candidate genes, such as linkage analysis, mutant populations, comparative genomics and association studies, and these methods often provide complementary information about gene(s) of interest (Yu and Buckler, 2006). The identification of a hypothetical gene from GWAS and from sequenced EMS mutants requires further genetic studies and validation. Structured bi-parental populations such as F_2 or back cross (BC) populations can be used to identify linkage between a phenotype and a gene (Yu and Buckler,

2006). F_2 and BC populations are of a lower mapping resolution than association mapping studies and are suitable only for a small number of alleles but can be rapidly developed (Yu and Buckler, 2006). Recombinant inbred line (RIL) populations can also be used for linkage mapping with quantitative traits (Yu and Buckler, 2006; Boyles et al., 2017) but require many years of development (Yu and Buckler, 2006).

A common strategy to relate SNP mutations in candidate genes to a phenotype is through co-segregation analysis for genotype and phenotype (Jiao et al. 2016, Addo-Quaye et al. 2017, Zhang et al. 2004) in bi-parental F_2, or RIL populations (Boyles et al., 2017). A disadvantage of F_2 populations is the low level of recombination and thus result in large linkage blocks across chromosomes making it difficult to narrow down the region carrying the causal polymorphism.

1.5 Starch Biosynthesis and Genetic Regulation

1.5.1 Importance of Starch and Starch Structure

Carbohydrates, especially starch, are a major human energy source (BeMiller and Huber, 2008) and transfer unique textural and nutritional properties to foods (BeMiller, 2014; BeMiller and Huber, 2008; Biliaderis, 2009). Unmodified starches are valuable for a variety of food products including snack foods and baby foods (BeMiller and Huber, 2008).

The sorghum endosperm contains the starch granules, protein matrix and protein bodies (Waniska and Rooney, 2000). A starch molecule is composed of two glucose polymers, amylose and amylopectin (Tetlow et al., 2004; BeMiller and Huber, 2008; Jane, 2009) that interact to produce a starch granule together with intermediate molecules of amylose and amylopectin, phosphate monoester (Jane, 2009), ash, lipids and proteins (Jane, 2009, BeMiller and Huber, 2008). Amylose and amylopectin are the key influencers of starch functionality (Jane, 2009), where amorphous and predominant the crystalline regions play a key role (Tetlow and Emes, 2014) in

starch GT (Biliaderis, 2009). Amylose and amylopectin are linked via (1→ 4)-α-D-glucosyl units and have branching points of (1→ 6)-linkages (BeMiller and Huber, 2008; Jane, 2009; Shannon et al., 2009; Preiss, 2009) at 5% in amylopectin (Preiss, 2009; BeMiller and Huber, 2008; Jane, 2009) and only 0.3-0.5% in the linear amylose (BeMiller and Huber, 2008; Preiss 2009; Jane, 2009). Amylopectin is one of the largest polymers (BeMiller and Huber, 2008). The average chain length and degree of polymerization (DP) of amylose is between 100 and 10.000 glucosyl units, while the average chain length of amylopectin is only 20-30 glucosyl units but a DP of 10,000 to 100,000 (Shannon et al., 2009).

The amylose and amylopectin contents of grain starch vary by species but are approximately 70-80% amylopectin and 15-30% amylose in maize, rice, wheat (Jane, 2009; Biliaderis, 2009; Shannon et al., 2009; BeMiller and Huber 2008), and sorghum (Waniska and Rooney, 2000, Sang et al., 2008) and occasionally as high as 35% (Shannon et al., 2009). Non-mutant starches usually have more amylopectin than amylose (Shannon et al., 2009). Mutant starches such as waxy can be composed of 100% amylopectin and high amylose starches have been reported with more than 50% apparent amylose (Jane, 2009; BeMiller and Huber 2008).

1.5.2 Sorghum Starch Biosynthesis Pathway and Candidate Genes

The starch synthesis pathway (SSP) is a conserved pathway across species with a very conserved final starch structure (Tetlow et al., 2004). Across species, a key group of enzymes located in the endosperm is involved in starch biosynthesis including Granule Bound Starch Synthases (GBSSs), Starch Synthases (SSs), Starch Branching Enzymes (SBEs), Isoamylases (debranching enzyme), and pullulanases (debranching enzyme) (Tetlow et al. 2004). The different enzymes vary in their catalytic activities (Tetlow, 2004; Tetlow and Emes, 2014; Guan and Preiss 1993; Nakamura 2015). The SSs and SBEs act predominantly on amylopectin (Tetlow and Emes,

2014; Nakamura, 2015), while the GBSSI is responsible for amylose biosynthesis (Nakamura 2015). Starch synthases are ADP-glucose-dependent (alpha 1,4-glucan-4-glucosyl transferases; Boyer, 1985) transferases, while SBEs are 1,4-alpha-glucan 6-glucosyl transferases (Tetlow and Emes, 2014; Boyer, 1985).

Starch synthases can be grouped as GBSS and starch synthases I-IV (Preiss 2009; Tetlow et al. 2004). The GBSS enzymes are majorly involved in the elongation of amylose chains, while the starch synthases I-V are responsible for the elongation of the amylopectin glucan chain at the alpha-(1—> 4)-linkages (Tetlow et al. 2004, Nakamura 2015). Each enzyme class has several isoforms, where the SSIIa is involved in the elongation of short chains in monocots (<10DP chains are created by SSI) into intermediate size glucan chains (DP 12-24) (Tetlow et al., 2004). Two SSII classes exist, which are SSIIa, found in endosperm, and SSIIb found in chloroplasts (Tetlow et al., 2004).

The classes of SBEs in maize, rice, barley, wheat are BEI, BEII (Nakamura, 2015) and BE3, BE4 in rice (Preiss, 2009). Tetlow and Emes (2014) used amino acid sequences to find similarity and classified the SBEs of cereals into three groups, (1) SBEIIb from sorghum, maize, barley, wheat and SBE-3 from rice, (2) SBEIIa from sorghum, maize, rice, barley and wheat and (3) with SBEI from maize, sorghum, wheat and barley (Tetlow and Emes, 2014).

Boyer (1985) purified soluble SS and different SBE's in developing sorghum seeds and reported high similarity with maize soluble SS and SBE's. The SBEIIb (NCBI Genbank number AY304539 and AY304540) of sorghum was first identified from BAC clones and showed high sequence similarity with maize, rice, barley and wheat (Mutisya et al., 2003). SBEIIb in sorghum has been reported to be a single copy gene (Mutisya et al., 2003) and is expressed endosperm specific (Mutisya et al., 2006). The SBEs are responsible for the creation of branch points as alpha-

(1→6) linkages and by doing that transfer glucan chains from one position to another by cleaving alpha-(1→4)-linkages (Tetlow et al., 2004; Preiss, 2009; Tetlow and Emes, 2014; Nakamura, 2015). These chain transfers can be either an intrachain or interchain transfer (Tetlow and Emes, 2014; Nakamura 2015). The SBEs cleave the original bonds and thus create nonreducing-ends, which are than elongated by starch synthases (Tetlow and Emes, 2014). SBEI transfers longer glucan chains and SBEII (SBEIIa, SBEIIb) transfer shorter glucan chains (Tetlow et al., 2004; Preiss, 2009). SBEII act mostly on amylopectin (Tetlow et al. 2004; Tetlow and Emes 2014) and transfer shorter chains of degree of polymerization (DP) 6-14 in corn, rice, wheat, potato, and especially SBEIIb is an endosperm specific enzyme transferring chains of DP 6-7 (Tetlow and Emes, 2014).

Different classes of enzymes work together as a protein complex to create the starch molecule (Tetlow et al. 2004; Preiss 2009; Ahmed et al., 2015; Liu et al. 2012b) and also affect each other if one is mutated (Tetlow et al. 2004; Ahmed et al., 2015).

Campbell et al. (2016) evaluated the starch synthesis pathway (SSP) of sorghum. The sorghum SSP is multidirectional and consists of three SSPs (SSP-1, SSP-2, SSP-3), all resulting in ADP-Glc (Campbell et al., 2016). ADP-Glc is the starting product for starch synthesis of amylose and amylopectin (Campbell et al., 2016; Preiss, 2009) to synthesize the (1 → 4) α-glucosidic linkages in both polymers (Preiss, 2009). Campbell et al. (2016) reported genes involved in the primary sorghum SSP in the amyloplast and their maize homologs like among others starch synthases such as "*Sobic.010G022600.1*, similar to GBSS1 precursor", a homolog of *Waxy (Wx*) in maize; "*Sobic.010G093400.1*, similar to starch synthase isoform zSTSII-1", homolog to SSIIa in maize; "*Sobic.002G116000.1*, similar to GBSS IIa", homolog to SSII in maize; and other genes. A second group of genes represented mostly branching and debranching enzymes

were reported for example "*Sobic.004G163700.1*, similar to 1,4-alpha-glucan-branching enzyme 2, chloroplast precursor", homolog to Ae1 in maize; "*Sobic.010G273800.1*, similar to starch branching enzyme I precursor", no maize homolog; "*Sobic.003G213800.2*, similar to putative 1,4-alpha-glucan branching enzyme", no maize homolog; and other genes (Campbell et al., 2016).

Starch mutants have been used in sorghum, maize, potato, barley, wheat and rice to identify the role of different enzymes in starch biosynthesis (Preiss, 2009; Nakamura, 2015).

1.5.3 Starch Branching Enzyme Genes and *Amylose Extender* Starches

The group of SBEII act predominantly on amylopectin (Tetlow et al., 2004; Tetlow and Emes, 2014; Nakamura, 2015). Amylose content is an important starch quality parameter for product quality (BeMiller and Huber, 2008). Amylose content depends on the starch structure, the ratio of amylose to amylopectin, and is influenced by environmental effects such as temperature (Jane, 2009). The apparent amylose levels from iodine analysis vary across sorghum species and range between 15.5 and 47.3 (Taylor et al., 1997; Beta et al., 2000).

The *amylose extender (ae)* is a known mutant starch type resulting in high amylose starches in maize (Shannon et al., 2009; Li et al., 2008; Liu et al., 2012a; Tetlow and Emes, 2014; Nakamura, 2015b) and rice (Tetlow and Emes, 2014, Nishi et al., 2001, Nakamura, 2015b) and is regulated by the *SBEIIb* (Jane, 2009; Tetlow and Emes, 2014; Nishi et al., 2001; Nakamura, 2015b). A sorghum study on natural variation reported that both genes *SSIIa* and *SBEIIb* (GenBank accession AY304539, AY304540) result in increased amylose values (Hill et al., 2012). The *ae* in maize is reported as a recessive gene with allele dosage dependent influence on amylose content in the triploid endosperm; therefore, the wild type allele is not completely dominant over the mutant allele *ae* (Shannon et al., 2009). Such high amylose maize starches result in higher starch GT conclusion temperatures (Tc) (Li et al., 2008; Shannon et al., 2009) even higher than 100°C (Li et

al., 2008). The environment, modifier genes and the genetic background also influence the *ae* phenotype (Shannon et al. 2009). Genotypes with an *ae* mutation and high levels of amylose are highly correlated with resistant starch content and are less enzyme-digestible (Li et al., 2008, Tetlow and Emes, 2014; Jane, 2009). *Amylose extender* starches from maize and high amylose rice mutants have longer internal branch chains of amylopectin, are less branched in outer chains, have fewer short chains, which is a result of the lack of *SBEIIb* transferring short chains in the starch molecule (Jane, 2009; Tetlow and Emes 2014). Those high amylose mutants therefore consist mostly of amylose and intermediate products (Jane, 2009). It is reported that mutations in the *SBEIIb* have very specific catalytic functions that cannot be complemented by other starch branching enzymes (Nakamura, 2015). The SBEs determine branching frequency as they are able to set branch points at varying locations and transfer longer or shorter branch chains resulting in starches that differ in digestibility thereby influencing food processing and starch digestibility (Tetlow and Emes, 2014).

1.5.4 Starch Synthases Genes and Starch GT

An important food product quality parameter is starch GT. Starch GT varies between cultivars and between species (Shannon et al., 2009), with sorghum being one of the species with a higher starch GT (Akingbala et al., 1988; Beta et al., 2000; Taylor, 1992; Taylor et al., 1997; Waniska and Rooney, 2000). Sorghum starch gelatinization has been reported between 71°C to 80°C (Waniska and Rooney, 2000), with variation by geography such as sorghum produced in Zimbabwe with a mean of 67.4°C for peak starch GT (T_p) (Beta et al. 2000), in South Africa 63.0°C (T_o) to 70.5°C (T_c) (Taylor et al., 1997) and in India ranging from 69°C (T_o) to 83.4°C (T_c) (Akingbala et al., 1988). For sorghum starch improvement, it may be advantageous to reduce starch GT to improve the mashing and malting steps in sorghum beer production (House et al.,

2000). The high starch GT of sorghum requires higher mashing temperatures, which result in an inactivation of necessary alpha- and beta amylases for saccharification and with that low levels of fermentable sugars (Taylor and Dewar, 2000; Taylor, 1992).

The *Su2* gene codes for SSIIa in maize (Zhang et al., 2004; Liu et al., 2012b). The *ssIIa* mutations in maize result in lower starch GT than the wild type (Zhang et al., 2004; Shannon et al., 2009; Preiss, 2009; Liu et al., 2012b), up to 15-20°C lower onset starch GT (To) (Zhang et al., 2004; Liu et al., 2012b). The *su2* mutants in maize exhibit a strong drop in enthalpy of 4.2 J/g in comparison to the wild type with 15.8 J/g (Liu et al., 2012b). In rice, the *ALK* gene encodes for the SSSII-3 and mutations can result in varying starch GT (Gao et al. 2011). A sorghum study of *SBEIIb* and *SSIIa,* reported that varying haplotypes of both genes result into both, low and high starch GT (Hill et al., 2012) and did not show a clear distinction in functionality of these genes. A positive correlation between amylose content and starch GT was reported in sorghum (Hill et al., 2012). It has been reported in other species, that the low starch GT was related to the amylopectin structure of *ssII* mutants, that showed an increase in short chains and a lack of intermediate chains (Preiss, 2009). The *ssII* mutants result in different phenotypes across species with higher amylose values in maize (Zhang et al., 2004; Liu et al., 2012b), barley and wheat (Luo et al., 2015), and a reduction of amylose content in other studies (Preiss, 2009) and nearly no change in amylose content in rice (Luo et al., 2015). The *su2* mutants exhibited a normal total starch content (Shannon et al., 2009) but *ssII* mutants were reduced in starch content in developing corn endosperm (Liu et al., 2012b). In contrast to *ae* mutants, the *Su2* wild type allele is dominant over the *su2* allele (Shannon et al., 2009). The *su2* mutant phenotype is also affected by the environment and genetic background (Shannon et al., 2009). The creation of double mutants of *amylose extender* and *sugary*

2 (su2) resulted in a variation of amylose values most similar to the *ae* mutant, with the starch GT also being more similar to the *ae* parent (Shannon et al., 2009).

1.6 The Alkali Spreading Phenotype

1.6.1 The Alkali Spreading Test

Little et al. (1958) developed a test to measure the alkali spreading value (ASV) on milled white rice varieties. Six kernels per variety were placed in a transparent plastic box with 10ml of 1.7% potassium hydroxide (KOH) solution for 23 hours and scored for the alkali spreading value (ASV) (Little et al., 1958). The ASV scores were based on extent of dispersion and alkali clearing value on a scale of 1 to 7 (Little et al., 1958). The grading scheme is "1 = grain not affected, 2 = kernel swollen, 3 = kernel swollen, collar incomplete and narrow, 4 = kernel swollen, collar complete and wide, 5 = kernel split or segmented, collar complete and wide, 6 = Grain dispersed merging with collar, 7= Grain completely dispersed and intermingled" (Juliano et al., 1982; Little et al., 1958). The score of 1 is well distinguishable from a score of 6 or 7, while the intermediate classes are more difficult to evaluate. Little et al. (1958) observed different patterns of degradation and dispersion between genotypes. The alkali test was further modified in a follow up study using 1.4% KOH for rice varieties that exhibit a wide range of ASV scores (Bhattacharya and Sowbhagya, 1972).

Not much is known about the ASV in sorghum. An earlier research study by Waniska (1976) tried to adapt the alkali test to sorghum and treated whole, pearled sorghum kernels with 5N NaOH. The sorghum seeds were scored for alkali disintegration and dispersion on a scale from 1 to 6 to distinguish between waxy and non-waxy genotypes (Waniska 1976). A study by McKneight (2015) used 1.5% KOH on cut sorghum seeds in an Eppendorf tube and showed that the alkali spreading value occurred after 18h on the low starch GT mutant SbEMS 4308.

1.6.2 The Alkali Spreading Phenotype and Starch Gelatinization

Starch GT influences rheological properties and functionality of starches and determines end-product quality (Biliaderis, 2009; BeMiller and Huber, 2008). Three different treatments are known to induce starch gelatinization: hot starch gelatinization (BeMiller and Huber, 2008), a cold starch gelatinization (Bhattacharya, 1979; Ragheb et al., 1995) or food extrusion using a high shear process (Bhattacharya and Hanna, 1987). Hot starch gelatinization occurs by adding water in the presence of heat (Bhattacharya, 1979; BeMiller and Huber 2008), while a cold gelatinization occurs by adding an alkali solution (Bhattacharya, 1979). Cold gelatinization can be forced by adding potassium hydroxide (KOH) or sodium hydroxide (NaOH) solution (Ragheb et al., 1995) such as in the case of the ASV assay (Little et al., 1958; Griebel et al., 2019; Waniska, 1976).

The ASV assay is used in rice research to identify varieties with variations in starch GT (Mogga et al. 2018; Huang et al., 2010, Huang et al., 2012; Song et al., 2019). Many studies have shown that a strong expressed ASV (ASV 6, 7) is correlated with low starch GT in rice (Bhattacharya, 1979; Bhattacharya and Sowbhagya, 1972; Juliano et al., 1964; Mariotti et al., 2010; Tan and Corke, 2002; Gao et al. 2011). Thus, ASV has been used in rice as an indirect estimate for starch GT without direct validation using DSC (Huang et al., 2010, Huang et al., 2012; Song et al., 2019). Recent studies comparing rice RILs and landraces for ASV and starch GT measured by DSC demonstrated only a weak correlation between ASV and DSC values (Cuevas et al., 2010). A strong expressed ASV occurred with lower and higher DSC values (Cuevas et al., 2010). The ASV intensity of disintegration in rice relates to whether an intermediate or low starch GT exists, since an ASV of 5 to 7 is considered as low starch GT <70°C, while an ASV below 5 is considered as intermediate starch GT >70°C (Mogga et al., 2018, Juliano, 1982).

1.6.3 Genetic Variants associated with the Alkali Spreading Phenotype

The key gene for ASV and starch GT in rice is reported to be the *ALK* gene encoding for SSII-3 (Gao et al., 2011; Tian et al., 2009). Depending on the SSII-3 allele, higher or lower starch GT (starch GT measured in ASV values) were reported (Tian et al., 2009). The GT, indirectly measured using ASV, with a score >5 was related to higher but also lower amylose content (Tian et al., 2009). Minor genes that are associated with starch GT are *Wx, SBE3, ISA* and *SSIV-2* in rice (Tian et al., 2009).

Only a few GWAS for ASV have been reported in rice (Zhao et al., 2011; Mogga et al., 2018; Song et al., 2019; Huang et al., 2010; Huang et al., 2012). A GWAS study on 517 rice landraces estimated starch GT by ASV and found a hit on chromosome 7, at position 6,404,473 close by candidate genes including amylase inhibitors (*Os07g0213700*b; *Os07g0214300*b; *Os07g0215500*b) (Huang et al., 2012). In a related study, a compressed MLM was used to identify a significant SNP for ASV on Chromosome 6 at position 6,726,252 around approximately 24kb away from the *ALK* gene (Huang et al., 2010). Another GWAS in rice used a mixed model approach and found significant peaks on Chromosomes 2, 3, 6, 7 and 8 with a hit on Chromosome 6 within 200kb of the known candidate gene *SSII-3* and on chromosome 8 within 200kb of the known *SSII-2* gene (Zhao et al., 2011). A naïve model approach resulted in significant signals on Chromosomes 1, 3, 4, 5, 6, 7 and 8, with regions on Chromosome 4 being within 200kb of the *PUL* gene and on Chromosome 6 within 200kb of the *SSII-3* gene (Zhao et al., 2011). The ASV GWAS hits varied between years (Year 1: Chr 1, 5, 6, 7, 8, 9, 11; Year 2: Chr. 3, 6 and 12) and could not be confirmed across years and or in a combined analysis (Zhao et al., 2011). Another rice study on only 59 varieties used GLM and MLM in TASSEL to perform GWAS for ASV and identified ten significant QTLs on Chromosome 1, 3, 4,6, 7, 8, 9, 10 (Mogga et al., 2018). Song et

al. (2019) evaluated the USDA rice mini core collection of 217 accessions and found one significant GWAS hit for ASV on Chromosome 4, at position 23527208 (SNP gi04_23527208).

The alkali spreading phenotype is not well known in sorghum. One study hypothesized that the *SSIIa* is the causal gene for ASV, as a mutant SbEMS 4308 exhibited the ASV phenotype and had a SNP mutation in *Sobic.010G093400* (McKneight, 2015). So far, no candidate genes for ASV have been proved to be causal in either mutant populations or in the standing variation in sorghum. Also, no GWAS has been conducted yet on the ASV phenotype in Sorghum.

1.6.4 Genetic Interactions of ASV and related Starch Quality Characteristics

The interactions of low amylose content, high gel consistency and a low starch GT results into soft and sticky rice (Wang et al., 2007). The paste viscosity, amylose content and gel consistency are very important traits in cooking quality, while ASV, pasting temperature, and pasting time, are important in the rice cooking process (Wang et al., 2007). Cooking quality is primarily controlled by the Wx gene and minor effected by the *ALK* gene. The cooking process is controlled by the *ALK* gene with minor effects by *Wx* (Wang et al., 2007). The major gene controlling amylose content in rice is *Wx* with minor gene effects of *SSII-3, SSIII-2, SSI, PUL* and *AGPPlar* (Tian et al., 2009).

Gel consistency (GC) is primarily controlled by *Wx* (Tian et al. 2009). The *ALK (SSII-3*) gene also has a minor impact on GC (Tian et al., 2009; Gao et al., 2011) and amylose content (Tian et al., 2009). Other genes controlling GC are *ISA, SBE3* and *AGPiso* (Tian et al., 2009). GC is negatively correlated with amylose content (r=-0.91) in rice (Tian et al., 2009). A transgenic rice study reported positive correlation of (1) amylose content and (2) GC with ASV (Gao et al., 2011). The same study reported a negative correlations of (1) starch GT, (2) amylose content and (3) gel consistency with the variant of *ALK* transgene is used (Gao et al. 2011). Depending on the *ALK*

transgene the ASV scores varied from low to high (Gao et al., 2011). Peak viscosity was positive correlated with gel consistency and negatively correlated with starch GT (Gao et al., 2011).

Additive gene action between *Wx* and *SSII-3* were reported to influence amylose content and starch GT (measure as ASV) in rice (Tian et al., 2009). Higher amylose contents are observed together with an increase in starch GT (starch GT measure as ASV) but also high ASV values with low amylose contents (Tian et al., 2009).

In sorghum, no correlation was observed between amylose content and starch GT (T_o, T_p) (Taylor et al. 1997), while another study observed a positive correlation between amylose content and T_p (0.73) (Beta et al., 2000). A negative correlation between amylose and peak viscosity was reported in sorghum (Beta et al., 2000) but was not found by Taylor et al. (1997). A sorghum study on natural variation reported a positive correlation between amylose levels and final viscosity, peak time and setback, while the peak viscosity and breakdown were negative correlated with amylose content (Hill et al., 2012). The rheological behaviour of starches depends on granule swelling, solubilization of amylose and amylopectin during gelatinization (Biliaderis, 2009). Amylopectin is the driving force on granule swelling (Biliaderis, 2009). Amylose retrogrades in aqueous solutions whereas amylopectin is stable (Shannon et al., 2009; Jane, 2009). Starches with more amylose content are structural unstable and often do not form gels (Biliaderis, 2009). The strength or softness of a gel is also influenced by other factors such as fats, proteins and water content (BeMiller and Huber, 2008).

1.7 Research Objectives of this Dissertation

The goal of this dissertation is to dissect the phenotypic and genetic architecture of a relatively new and unique phenotype, the alkali spreading phenotype, in sorghum seed endosperm to determine genotypes with improved food processing, starch quality and dietary characteristics.

This should enhance sorghum grain quality for production of improved food products for low-income groups in Africa and India. The identified genetic variants can be introgressed into commercial African varieties.

Therefore, the objectives of my dissertation are:

Experimental objective I: Assess phenotypic variation of the alkali spreading phenotype and the relationship to starch thermal properties in mature sorghum seeds using an EMS population.

Experimental objective II: Identify and assess EMS genetic variants causing the alkali spreading phenotype in sorghum seeds and characterize their functional impact on starch quality traits.

Experimental objective III: Identify and characterize genetic variants of the alkali spreading phenotype in sorghum seed in a sorghum diversity panel.

1.8 Broader Impact: Building the foundation for new sorghum varieties to fight global food insecurity

The goal of this dissertation is to improve grain quality for sorghum food products of low-income groups consuming sorghum in India and Africa who depend on the daily carbohydrate intake of sorghum as a staple food. This research will help in the long run to develop sorghum varieties with improved sorghum starch quality traits such as lower starch GT, varying pasting profiles, and different starch digestibilities (Griebel et al. 2019; Griebel et al. 2019b). Sorghums with lower starch GT could be favourable for the local beer brewers and thus create new markets and networks for smallholder farmers. The development of sorghum starches with new and unique processing characteristic also opens the door for unique food products and develops new markets. The improvement of sorghum starches that are gluten free and vary in starch digestibility may contribute dietary solutions for humans with diabetes (Englyst et al. 2007) and gluten allergies. This is of great economic value for smallholder farmers in Africa and India, who grow sorghum

for their own consumption and selling surplus to the urban population or food processing companies. This will likely improve the economic situation of smallholder farmers which is related to better health, better livelihood conditions and better chances for those farmers children for higher school education.

Therefore, this research will bring a long-term great benefit to the global sorghum consuming human society as it can increase economic income, can provides people access to affordable and dietary food and with that helps to fight global food insecurity.

1.9 References

Adeyanju A, Little C, Yu J, Tesso T (2015) Genome-wide association study on resistance to stalk rot disease in grain sorghum. *G3* 5:1165-1175.

Addo-Quaye C, Tuinstra M, Carraro N, Weil C, Dilkes BP (2018) Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. *G3 (Bethesda)* 8: 1079-1094.

Addo-Quaye C, Buescher E, Best N, Chaikam V, Baxter I, Dilkes BP (2017) Forward genetics by sequencing EMS variation-induced inbred lines. *G3 (Bethesda)* 7:413-425.

Ahmed Z, Tetlow IJ, Ahmed R, Morell MK, Emes MJ (2015) Protein-protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heterometric enzyme complexes in the synthesis of A and B granules. *Plant Science* 233:95-106.

Akingbala JO, Gomez MH, Rooney LW, Sweat VE (1988) Thermal Properties of Sorghum Starch. *Starch/Stärke* 40:375-378.

BeMiller JN, Huber KC (2008) Carbohydrates. *Fennema's Food Chemistry*, eds Damodaran S, Parkin K, Fennema O, (CRC press, FL), pp. 83-154.

BeMiller J (2014) Carbohydrate analysis. *Food Analysis*, ed Nielsen S (Springer, New York), pp 147-180.

Beta T, Corke H, Rooney LW, Taylor JRN (2000) Starch properties as affected by sorghum grain chemistry. *J Sci Food Agric* 81:245-251.

Bhattacharya M, Hanna MA (1987) Kinetics of Starch Gelatinization During Extrusion Cooking. *J. Food Sci*. 52, 764-766.

Bhattacharya KR (1979) Gelatinization temperature of rice starch and its determination. *International Rice Research Institute, Proceedings of the workshop on chemical aspects of rice grain quality.* (International Rice Research Institute, Philippines) pp. 231-249.

Bhattacharya KR, Sowbhagya CM (1972) An improved alkali reaction test for rice quality. *J Food Sci Technol* 7:323-331.

Biliaderis CG (2009) Structural transitions and related physical properties of starch. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam) pp 293 – 372.

Billot C, Ramu P, Bouchet S, Chantereau J, Deu M, Gardes L, Noyer J-L, Rami J-F, Rivallan R, Li Y, Lu P, Wang T, Folkertsma RT, Arnaud E, Updadhyaya HD, Glaszman J-C, Hash CT (2013) Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLOS ONE* 8:1-16.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23:2633–2635.

Brenton ZW, Cooper EA, Myers MT, Boyles RE, Shakoor N, Zielinski KJ, Rauh BL, Bridges WC, Morris GP, Kresovich S (2016) A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics* 204:21-33.

Boyer CD (1985) Soluble starch synthases and starch branching enzymes from developing seeds of sorghum. *Phytochemistry* 24:15-18.

Boyles RE, Pfeiffer BK, Cooper EA, Rauh BL, Zielinski KJ, Myers MT, Brenton Z, Rooney WL, Kresovich S (2017) Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theor Appl Genet* 130:697-716.

Campbell BC, Gilding EK, Mace ES, Tai S, Tao Y, Prentis PJ, Thomelin P, Jordan DR, Godwin ID (2016) Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol J* 14:2240-2253.

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. *Crop Science* 48:30-40.

Cuevas RP, Daygon VD, Corpuz HM, Nora L, Reinke RF, Waters DLE, Fitzgerald MA (2010) Melting the secrets of gelatinization temperature in rice. *Funct. Plant Biol.* 37, 439-447.

Cuevas HE, Rosa-Valentin G, Hayes CM, Rooney WL, Hoffmann L (2017) Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics* 18:108 1-17.

Dahlberg JA (2000) Classification and characterization of sorghum. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 99-130.

Deu M, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorhgums using a core collection. *Genome* 49:168-180.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, 6:1-10.

Englyst KN, Liu S, Englyst HN (2007) Nutritional characterization and measurement of dietary carbohydrates. *Eur J Clin Nutr*, 61:S19-S39

FAO, 2019, http://www.fao.org/land-water/databases-and-software/crop-information/sorghum/en/). FAOSTAT, 2019, http://www.fao.org/faostat/en/) (Jan. 2019).

FAOSTAT, 2016, FAOSTAT Database. http://www.fao.org/faostat/en/ (Oct. 2016).

Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y,Yan M, Jiang H, Dong G, Huang Y, Han B, Li J, Qian Q (2011) *ALK*, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. *J Integr Plant Biol* 53: 756-765.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLOS ONE* 9:1-11.

Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164:731-740.

Griebel S, Webb MM, Campanella OH, Craig BA, Weil CF, Tuinstra M.R. (2019) The alkali spreading phenotype in *Sorghum bicolor* and its relationship to starch gelatinization. *J Cereal Sci* 86:41-47.

Griebel S, Westerman RP, Adeyanju A, Addo-Quaye C, Craig BA, Weil CF, Cunningham M S, Bhavesh P, Campanella OH, Tuinstra MR (2019b) Mutations in sorghum *SBEIIb* and *SSIIa* affect alkali spreading value, starch composition, thermal properties and flour viscosity. *In Review. Theor Appl Genet.*

Guan HP, Preiss J (1993) Differentiation of the properties of the branching isozymes from maize *(Zea Mays*). *Plant Physiology* 102:1269-1273.

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. *Crop Science* 12:172-176.

Hayes CM, Burow GB, Brown PJ, Thurber C, Xin Z, Burke JJ (2015) Natural variation in synthesis and catabolism genes influences dhurrin content in sorghum. *The Plant Genome* 8:1-9.

Hill H, Lee LS, Henry RJ (2012) Variation in sorghum starch synthesis genes associated with differences in starch phenotype. *Food Chemistry* 131:175-183.

House LR, Gomez M, Sun Y, Murty DS, Verma BN (2000) Development of some agricultural industries in several African and Asian countries. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 131-190.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z**,** Li M, Fan D, Guo Y, Wang A, Wang L**,** Deng L, Li W, Lu Y, Wenig Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F**,** Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42:961-969.

Huang X, ZhaoY, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44:32-41.

Jane J-l (2009). Structural features of starch granules II. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam), pp 193 – 236

Jiao Y, Burke J, Chopra R, Burow G, Chen J, Wang B, Hayes C, Emendack Y, Ware D, and Xin Z (2016) A sorghum mutant resource as an efficient platform for gene discovery in grasses. *The Plant Cell* 28:1551-1562.

Juliano BO, Blakeney AB, Butta I, Castillo D, Choudhury NH, Iwasaki T, Shibuya N, Kongseree N, Lapis ET, Murty VVS, Paule CM, Perez CM, Webb BD (1982) International Cooperative Testing of the Alkali Digestibility Values for Milled Rice. *Starch/Stärke* 34:21-26.

Juliano BO, Bautista GM, Lugay JC, Reyes AC (1964) Studies on the physicochemical properties of rice. *J Agric Food Chem* 12:131-138.

Kimber CT (2000) Origins of domesticated *Sorghum* and its early diffusion to India and China. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 3-98.

Kimber CT, Dahlberg JA, Kresovich S (2013) The gene pool of *Sorghum bicolor* and its improvement. *Genomics of the Saccharinae, Plant Genetics and Genomics: Crops and Models 11, ed Paterson AH* (Springer New York), pp 23-41.

Li L, Jiang H, Campbell M, Blanco M, Jane J-l (2008) Characterization of maize amylose-extender (*ae*) mutant starches. Part I: Relationship between resistant starch contents and molecular structures. *Carbohydr Polym* 74:396-404.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and predition integrated tool. *Bioinformatics* 28:2397-2399.

Little RR, Hilder GB, Dawson EH (1958) Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem*. 35:111-126.

Liu F, Ahmed Z, Lee EA, Donner E, Lui Q, Ahmed R, Morell MK, Emes MJ, Tetlow IJ (2012a) Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein-protein interactions. *J Exp Bot* 63:1167-1183.

Liu F, Romanova N, Lee EA, Ahmed R, Evans M, Gilbert EP, Morell MK, Emes MJ, Tetlow I. J (2012b) Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. *Biochem J* 448:373-387.

Loveless A (1958) Increased rate of plaque-type and host-range mutation following treatment of bacteriophage in vitro with ethyl methane sulphonate. *Nature* 181:1212–1213.

Luo J, Ahmed R, Kosar‐Hashemi B, Larroque O, Butardo (Jr) VM, Tanner GJ, Colgrave ML,·Upadhyaya NM, Tetlow IJ, Emes MJ, Millar A, Jobling SA, Morell MK, Li Z (2015) The different effects of starch synthase IIa mutations or variation on endosperm amylose content of barley, wheat and rice are determined by the distribution of starch synthase I and starch branching enzyme IIb between the starch granule and amyloplast stroma. *Theor Appl Genet* 128:1407-1419.

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa´s indigenous cereal crop sorghum. *Nat. Commun.* 1-9 DOI: 10.1038/ncomms3320.

Mariotti M, Fongaro L, Catenacci F (2010) Alkali spreading value and image analysis, *J Cereal Sci* 52:227-235.

McKneight MM (2015) Functional genomics of sorghum grain quality traits. *Master Thesis*, Purdue University, West Lafayette, Indiana, USA.

Mogga M, Sibiya J, Shimelis H, Lamo J, Yao N (2018) Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (Oryza sativa L.) using DArT markers. *PLOS ONE* 1-19 doi.org/10.1371/journal.pone.0198012.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum, *PNAS*, 110:453-458.

Mutisya J, Sathish P, Sun C., Andersson L, Ahlansberg S, Baguma Y, Palmqvist S, Odhiambo B, Åman P, Jansson C (2003) Starch branching enzymes in sorghum (*Sorghum bicolor*) and barley (*Hordeum vulgare*): Comparative analyses of enzyme structure and gene expression. *J Plant Physiol* 160:921-930.

Mutisya J, Sun C, Palmqyist S, Baguma Y, Odhiambo B, Jansson C (2006) Transcriptional regulation of the sbeIIb genes in sorghum (Sorghum bicolor) and barley (Hordeum vulgare): Importance of the barley sbeIIb second intron. *J. Plant Physiol* 163:770-780.

Nakamura Y (2015). Biosynthesis of reserve starch. *Starch metabolism and structure,* eds Nakamura Y, (Springer, Japan), pp 161-209.

Nakamura Y (2015b) Manipulation of rice starch properties for application. *Starch metabolism and structure*, eds Nakamura Y, (Springer, Japan), pp 335-369.

Nishi A, Nakamura Y, Tanaka N, Satoh H (2001) Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiol* 127:459-472.

NuLifeMarket, 2019. https://nulifemarket.com/what-is-sorghum/, accessed 2019.

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *PNAS* 101:9903-9908.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J and Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551-556.

Paterson AH (2013) The sorghum genome sequence: A core resource for *Saccharinae* genomics. *Genomics of the Saccharinae Plant Genetics and Genomics: Crops and Models 11,* ed Paterson AH (Springer, New York) pp 105-119.

Preiss J (2009) Biochemistry and molecular biology of starch biosynthesis. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam), pp 83 – 148.

Ragheb AA, Abd El-Thalouth I, Tawfik S (1995) Gelatinization of starch in aqueous alkaline solutions. *Starch/Stärke* 46:338-345.

Rooney LW (2000) Genetics and cytogenetics. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp 261-307.

Rooney LW and Waniska RD (2000) Sorghum food and industrial utilization. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp 689-729.

Rosenow DT, Dahlberg JA, Stephens JC, Miller FR, Barnes DK, Peterson GC, Johnson JW, Schertz KF (1997a) Registration of 63 converted sorghum germplasm lines from the sorghum conversion program. *Crop Science* 37:1399-1400.

Rosenow DT, Dahlberg JA, Peterson GC, Clark LE, Miller FR, Sotomayor-Rios A, Hamburger AJ, Madera-Torres P, Quiles-Belen A, Woodfin CA (1997b). Registration of fifty converted sorghums from the sorghum conversion program. *Crop Science* 37:1397-1398.

Sang Y, Bean S, Seib PA, Pedersen J, Shi Y-C (2008) Structure and functional properties of sorghum starches differing in amylose content. *J Agric Food Chem* 56:6680-6685.

Sham PC, Purcell SM (2014) Statistical Power and significance testing in large-scale genetic studies. *Nature reviews Genetics*, 15:335-346.

Shannon JC, Garwood DL, Boyer CD (2009). Genetics and physiology of starch development. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam), pp 23-82.

Shenstone E, Cooper J, Rice B, Bohn M, Jamann TM, Lipka AE (2018) An assessment of the performance of the logistic mixed model for analyzing binary traits in maize and sorghum diversity panels. *PLOS ONE* 1-20 doi.org/10.1371/journal.pone.0207752.

Song J-M, Arif M, Zhang M, Sze S-H, Zhang H-B (2019) Phenotypic and molecular dissection of grain quality using a USDA rice mini-core collection. *Food Chemistry* 284:312-322.

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghum to early combine genotypes. *Crop Science* 7:396.

Sukumaran S, Xiang W, Bean SR, Pedersen JF, Kresovich S, Tuinstra MR, Tesso TT, Hamblin MT, Yu J (2012) Association mapping for grain quality in a diverse sorghum collection. *The Plant Genome* 5:126-135.

Swigoňová Z, Lai J, Ma J, Ramakrishna W, Lllaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. *Genome Research* 14:1916-1923.

Tan Y, Corke H (2002) Factor analysis of physiochemical properties of 63 rice varieties. *J Sci Food Agr* 82:745-752.

Taylor JRN (1992) Mashing with malted grain sorghum. *J. Am. Soc. Brew. Chem*. 50:13-18.

Taylor JRN, Dewar J, Taylor J, Ascheraden RF (1997) Factors affecting the porridge-making quality of south african sorghums. *J. Sci. Food Agr*. 73:464-470.

Taylor J, Dewar J (2000) Fermented products: Beverages and porridges. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 751-796.

Torkamaneh D, Laroche J, Belzile F (2016) Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLOS ONE*, 1-14. DOI:10.1371/journal.pone.0161333.

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. *J Exp Bot* 55:2131-2145.

Tetlow IJ, Emes MJ (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life* 66:546-558.

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M, Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *PNAS* 106:21760-21765.

USDA ARS (2019) United States Department of Agriculture (USDA), Agriculture Research Service (ARS), Germplasm Resources Information Network (GRIN). Database https://www.ars-grin.gov/npgs/aboutgrin.html (Accessed 2019).

Wang LQ, Liu WJ, Xu Y, He YQ, Luo LJ, Xing YZ, Xu CG, Zhang Q (2007) Genetic basis of 17 traits and viscosity parameters characterizing the eating and cooking quality of rice grain. *Theor Appl Genet* 115:463-476.

Waniska RD (1976) Methods to assess quality of boiled sorghum, gruel and chapaties from sorghum with different Kernel Characteristics. *Master thesis*. Texas A&M University, College Station, Texas, USA.

Waniska RD, Rooney LW (2000). Structure and chemistry of the sorghum caryopsis. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp 649-688.

Westergaard M (1957) Chemical mutagenesis in relation to the concept of the gene. *Experientia* 13: 224–234.

Yu J, and Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin in Biotechnol 17:155-160.

Zhao K, Tung C-W , Eizenga GC , Wright MH , Liakat Ali M , Price AH , Norton GJ, Rafiqul Islam M , Reynolds A , Mezey J , McClung AM , Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat. Commun*. 1- 10 DOI: 10.1038/ncomms1467.

Zhang X, Colleoni C, Ratushna V, Sirghie-Colleoni M, James MG, Myers AM (2004) Molecular characterization demonstrates that the *Zea mays* gene *sugary2* codes for the starch synthase isoform SSIIa. *Plant Mol Biol* 54:865-879.

CHAPTER 2. THE ALKALI SPREADING PHENOTYPE IN *SORGHUM BICOLOR* AND ITS RELATIONSHIP TO STARCH GELATINIZATION

A version of this chapter has been published at the Journal of Cereal Science 86 (2019) 41–47

https://doi.org/10.1016/j.jcs.2019.01.002

Stefanie Griebel[+], Molly M. Webb[§], Osvaldo H. Campanella[#], Bruce A. Craig[++], Clifford F. Weil[+], and Mitchell R. Tuinstra*[+]

[+] Department of Agronomy, Purdue University, West Lafayette, Indiana 47907;
[++] Department of Statistics, Purdue University, West Lafayette, Indiana 47907
[#] Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana 47907
[§] formerly Purdue University

2.1 Declaration of Author Contributions to this Chapter

Stefanie Griebel wrote the paper, created all figures and tables, performed and designed research. The other author contributions in this chapter are the followings: Clifford F. Weil, Mitchell R. Tuinstra: created and provided sorghum EMS population; Osvaldo Campanella: provided the DSC instrument and food scientific guidance; Molly Webb: initial alkali test on SbEMS4308 in her master thesis; Bruce Craig: provided statistical guidance; Mitchell R. Tuinstra: major advisor and project lead. All co-authors reviewed the paper and edited if needed.

2.2 Abstract

Sorghum, a staple food crop for millions of people in Africa and South Asia, is mainly consumed for the calorific value of its starch content, while the protein, mineral, lipid and fiber contents are important for flavor and nutritional quality. The starch structure largely controls cooking properties, processing quality, and digestibility. In this study, the Alkali Spreading Value (ASV)

assay was modified to identify variants in sorghum starch quality. The modified ASV test was then used to identify sorghum genotypes with contrasting phenotypes for the trait. Sorghum EMS mutants with strong ASV phenotypes (ASV+) exhibited a range of starch GT that were lower or higher than the wild-type Tx623 or Sepon82. Significant point biserial correlations were found between ASV and two starch-related traits: enthalpy ($r = -0.53$) and range of starch GT (Tc-To) ($r = 0.60$). ASV scores (ASV+ in mutants vs ASV- in controls) correlated with lower enthalpies and greater starch GT ranges (T_c-T_o). These results suggest that the ASV mutants have mutations in genes involved in starch biosynthesis or processing that cause changes in starch GT.

2.3 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a self-pollinated, cereal (FAOSTAT, 2016; Kimber, 2000) grown as an annual crop. It is commercially important in semiarid tropical and dry temperate environments around the world. Most sorghum (78.6%) is grown in low-income, food-deficient countries, with the world's largest sorghum growing areas in Africa (65.3%); 52.1% is grown in net food importing developing countries (FAOSTAT, 2016). Sorghum serves as a staple food for millions of people, especially in Africa and South Asia. Sorghum grown in Africa is utilized by smallholders for many purposes with the major focus on food and beverage production (House et al., 2000). Sorghum flour is less expensive than imported wheat and barley flour in these environments (Taylor and Dewar, 2000). Additionally, sorghum flour is valued as a gluten free ingredient for making cereal-based foods.

The sorghum grain is a caryopsis consisting of pericarp and germ, and a floury and corneous endosperm that contain the carbohydrates, protein matrix and protein bodies (Waniska and Rooney, 2000). Carbohydrates contribute to key textural and nutritional properties of foods (BeMiller, 2014; BeMiller and Huber, 2008; Biliaderis, 2009), especially starch, because of its

unique chemical and physical properties and nutritional value. Starch consists of glucose units that can form two key polymers, amylose and amylopectin (BeMiller and Huber, 2008; Jane, 2009). Unmodified starches are important in the production of food products, like extruded cereals, snack food products, baby foods, dry soup mixes, cakes (BeMiller and Huber, 2008) and in the beer brewing industry. Native starches have physical and chemical properties that change depending on the plant species, genotype, composition of the starch granule and environmental factors (Biliaderis, 2009).

Starch gelatinization is the process by which starch granules swell and lose their native structure. Gelatinization can occur either by adding water in the presence of heat (hot gelatinization) or adding an alkali solution (cold gelatinization) (Bhattacharya, 1979). Starches also can be transformed using food extrusion, which is a high shear process (Bhattacharya and Hanna, 1987) at low moisture content where starch granules are subjected to a melting process rather than the typical gelatinization occurring at high moisture contents. Cereal starches can also be gelatinized by adding a potassium hydroxide (KOH) or sodium hydroxide (NaOH) solution, which may promote the absorption of water in the starch granules and their swelling (Ragheb et al., 1995).

Starch GT is a species-specific property (Shannon et al., 2009), with sorghum starch GT reported as being higher than that of some other cereals (Akingbala et al., 1988; Beta et al., 2000; Taylor, 1992; Taylor et al., 1997; Waniska and Rooney, 2000). Starch gelatinization changes the rheological properties, behavior, and functionality of starchy products, which are important for end-product quality (Biliaderis, 2009). Furthermore, starch GT impacts malting and brewing qualities of sorghum (House et al., 2000). Starches with high GT makes simultaneous starch gelatinization and saccharification of starch into fermentable sugars a challenge (Taylor and Dewar, 2000; Taylor, 1992). Saccharification at high mashing temperature often results in a large fraction

of the starch being gelatinized but incompletely hydrolysed due to heat inactivation of alpha-and beta amylases (Taylor and Dewar, 2000; Taylor, 1992). Conversely, mashing at low temperatures may result in incomplete starch gelatinization and a low level of fermentable sugars. The decantation mashing process can be used to enhance starch gelatinization and saccharification but is very time consuming (Taylor and Dewar, 2000; Taylor, 1992). Thus, it would be of great advantage to the sorghum malting industry to develop varieties with lower starch GT. Low starch GT varieties could also enhance functionality of sorghum in extrusion cooking, allowing for reduced extrusion temperatures and improved energy efficiencies (Bhattacharya and Hanna, 1987).

An alkali test is a standard assay to classify rice varieties into those with high, intermediate or low starch GT (Mutters and Thompson, 2009). The alkali test was developed for milled white rice varieties using 1.7% KOH treatment (Little et al., 1958). The alkali spreading value (ASV) and alkali clearing value of the grain was scored on a scale of 1 to 7 (Little et al., 1958). A modified alkali test showed that a KOH concentration of 1.4% gave better results when studying rice varieties with a wide range of ASV scores (Bhattacharya and Sowbhagya, 1972). ASV was negatively correlated with starch GT in rice, with low starch GT varieties showing high ASV scores (Bhattacharya, 1979; Bhattacharya and Sowbhagya, 1972; Juliano et al., 1964; Mariotti et al., 2010; Tan and Corke, 2002).

Little information is available on the alkali spreading phenotype in sorghum. Early studies by Waniska (1976) described treatments of whole, pearled sorghum kernels with 5N NaOH (Waniska, 1976). The alkali disintegration and dispersion phenotypes were scored on a scale of 1 to 6 and could be used to distinguish between waxy and non-waxy genotypes. The alkali test widely used in rice breeding has not been adapted for use in sorghum. The objective of the current

study was to adapt and modify the ASV test as a rapid screening tool for identifying sorghum genotypes with altered starch composition and properties.

2.4 Experimental

2.4.1 Alkali Spreading Test

The alkali spreading test was adapted/modified from the protocol described by Little et al. (1958). For rice, six dehulled seeds are tested for each genotype with three seeds cut transversely and three seeds uncut. The rice seeds are placed in a petri dish (Little et al., 1958) and treated with 10ml 1.8% KOH for 24h. Images are taken using a flatbed scanner at 600dpi. For sorghum, seeds were cut longitudinally, each half placed in a well of a 96 well plate and treated with 250µl of either a 1.5% KOH or a 1.8% KOH solution. Each half seed sample was evaluated by visual inspection after 24h and again after 48h at room temperature. ASV was expressed as a binary trait with little variation in intermediate expression. Given the lack of intermediate phenotypes when scoring ASV in sorghum seeds, binary scores consisting of ASV- = no swelling, no disintegration and ASV+ = complete gelatinization (Juliano, 1980; Little et al., 1958, Waniska, 1976) were used in the present work.

2.4.2 Plant Material

Ethyl methanesulfonate (EMS) treatment of BTx623 seeds was used to create a mutant population of sorghum (Addo-Quaye et al., 2018). A population of approximately 5,700 lines (approximately 582 M4 and approximately 5116 M3 lines) was generated by self-pollination and screened using the modified ASV phenotype test at 1.5% KOH concentration. Mutants were validated by retesting at 1.5% and 1.8% KOH. The wild type BTx623 was used as a control. Ten EMS mutants were identified as exhibiting a score of ASV+. These mutants were self-pollinated

to the M5 to M7 generation with selection for the ASV phenotype at each generation. In West Lafayette 2016 (temperate climate), seeds from three panicles (biological replicates) for each mutant were harvested and screened for ASV and verified as having a score of ASV+; only two panicles were tested for mutants SbEMS3568 and SbEMS4308. The offspring of the West Lafayette 2016 panicles were grown in field trials in Guayanilla, Puerto Rico during the off-season. Three bulk seed samples were produced for six mutants and two bulked seed samples were produced for mutants SbEMS3403, SbEMS3920, SbEMS4565 and SbEMS3568, all from panicles exhibiting a score of ASV+.

Twelve rice varieties with contrasting ASV phenotypes were obtained from the Germplasm Resource Information Network (GRIN) of the USDA-ARS (Table S.2.1; USDA ARS, 2017).

2.4.3 Flour Preparation

Sorghum seeds from three panicles of eight mutants (only two panicles for SbEMS3568 and SbEMS4308) produced in West Lafayette 2016 were prepared as separate samples and milled into a fine flour using a ball mill (Retsch, Haan, Germany). The bulked samples described above were also milled into a fine flour. Seeds were milled in 30s intervals to avoid starch damage. Each of the sorghum checks were represented as two biological replicates. Seeds of the rice checks were obtained from the National Plant Germplasm System (USDA ARS, 2017) because these samples could not be produced in our field environments.

2.4.4 Starch Extraction

Starch was extracted from each flour sample (Benmoussa and Hamaker, 2011). Each flour sample (100mg) was mixed in 1ml NaCl (0.05M) with 1% sodium metabisulfite. The suspensions were inverted frequently for 30 min. Each sample was sonicated twice (1s pulse with 1s interval, 20% amplitude for 15s) using an ultrasonic processor (Model GE130, General Electric, Cleveland,

OH). The samples were then filtered through a 230 mesh-screen (Standard Testing Sieve, USA) into a new Eppendorf tube. The recovered slurries were sonicated for 15s and centrifuged for 5 min at 13,000 rpm (15871 x g). The supernatant was discarded and the pellets were solubilized in 250µL 0.05M NaCl. New Eppendorf tubes containing a 500µl sucrose density solution were prepared. The sucrose density solution has a concentration 65% w/v, which was prepared by dissolving 65g of sucrose in a 0.05M NaCl solution and adjusted to 100ml. The re-suspended pellets were layered on top of the sucrose density solution and then centrifuged for 5min at 500rpm (23 x g). The remaining sucrose density solution and suspension in the top layer was discarded. The remaining pellets were washed two times with 70% ethanol. The samples were allowed to dry for 1h at 40°C followed by drying overnight and then ground into a fine powder with a rod.

2.4.5 Starch Thermal Properties

Each mutant starch sample was analysed in duplicate for its thermal properties. Differential scanning calorimetry (DSC, Q2000-1223, TA instruments, New Castle, DE, USA) was used to determine starch GT. The starch was weighed into hermetic aluminium pans (Tzero) with a starch: deionized water ratio of 1:3. The pans were sealed with hermetic aluminum lids (Tzero) and stored at 4°C overnight for moisture homogenization in the sample. DSC was run using a heating profile from 20°C to 120°C at a rate of 5°C/min. DSC analysis was conducted using the TA instruments universal analysis software 2000, version 4.5A (TA instruments, https://www.tainstruments.com/support/software-downloads-support/downloads/), integrating the peak linear. Onset (T_o), peak (T_p) and completion (T_c) starch GT, range ($T_c - T_o$), endothermic enthalpy and melting point were determined.

2.4.6 Statistical Analyses

SAS version 9.4 was used for statistical analyses. Linear mixed model analysis of variance (ANOVA) with the Tukey-Kramer multiple comparison adjustment was performed using the GLIMMIX procedure to compare GT properties across genotypes and environments. Point biserial correlations were calculated using the CORR procedure.

2.5 Results and Discussion

2.5.1 Alkali Test

Previous studies of ASV in rice used whole, dehulled seeds and a scoring system of 1 to 7 (Little et al. 1958) whereas scales of 1 to 6 were used for whole, pearled sorghum (Waniska, 1976). Since non-pearled sorghum seeds do not exhibit an ASV phenotype in our study and decortication is slow and laborious, the modified ASV assay used cut seeds and a binary scoring system.

The ASV phenotypes of contrasting rice varieties were compared to sorghum to determine if the alkali test could be used to evaluate starch properties of sorghum. Rice varieties with known low, intermediate, and high starch GT were evaluated for variation in the ASV phenotype (Table S2.1). Our study confirmed previous results on varieties like Zenith, Vialone Nano, Taichung 65, Magnolia, Kaohsiung 68, Nato, and Colusa, reported as low starch GT types with ASV ratings above 5, and Century Patna 231, Bluebonnet, and Early Prolific reported as high starch GT types with ASV ratings of 1 to 2 (Table S2.1; Figure 2.1). Fortuna exhibited low ASV scores in our

study and mixed ASV scores by Little et al. (1958) (Table S2.1, Figure 2.1). However, it is reported as a low GT type in the USDA database (Table S2.1).

The ASV phenotypes of contrasting rice varieties were also evaluated in cut and dehulled seeds to determine whether cutting the seeds produced differences in the ASV phenotype. The cut-seed samples of the high starch GT varieties did not exhibit any swelling or disintegration, while low and intermediate starch GT types exhibited varying degrees of swelling or disintegration (Figure 2.1).

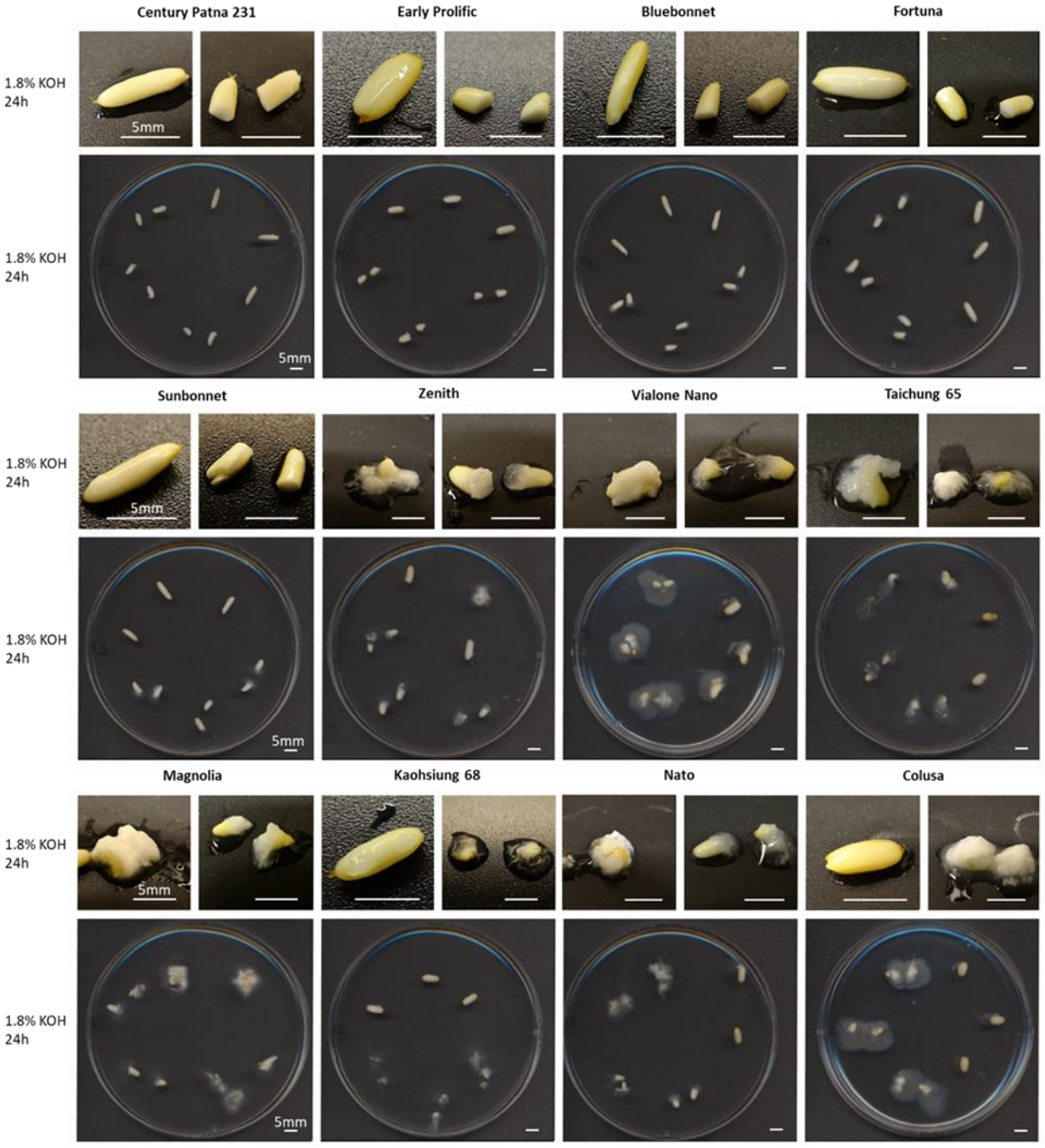


Figure 2.1 ***Oryza sativa*** **varieties** with contrasting starch GT phenotypes tested for ASV as whole and cut seed samples of dehulled rice seeds after 24 h treatment with 1.8%KOH.

Given the results of ASV testing on rice, cut sorghum seeds were screened with 1.5% KOH and 1.8% KOH for 24h. Our initial analyses of ASV in elite U.S. sorghum breeding lines indicated little variation. Seed samples from BTx623 and Sepon82 did not show any swelling or disintegration at 1.5% and 1.8% KOH concentrations (Figure 2.2). The screening program was expanded to evaluate approximately 5,700 sorghum EMS mutants (SbEMS mutants). Ten SbEMS mutants were identified with strong ASV phenotypes (Figure 2.2) similar to observations of ASV in rice (Figure 2.1). These mutants exhibited strong swelling phenotypes suggesting potential changes in starch content, composition, or physical properties (BeMiller and Huber, 2008; Jane, 2009). Literature suggests that adjusted and increasing KOH concentration in the ASV test better differentiates rice varieties (Juliano et al., 1982). Although the rice literature suggests a range of KOH concentrations for different varieties and studies (Bhattacharya, 1979; Bhattacharya and Sowbhagya, 1972; Juliano et al., 1982; Little et al., 1958), mutants were tested using concentrations 1.5% KOH and 1.8% KOH as they were created from the same genetic background Tx623, white sorghum. The ASV phenotypes for most of the sorghum genotypes were similar using 1.5% KOH and 1.8% KOH. The ASV phenotype varied for SbEMS3920, which was inconsistent at 1.5% but expressed more consistently at 1.8% KOH (data not shown). Therefore, it was decided to use 1.8% KOH as the optimal concentration. Incubation periods of 24h and 48h were evaluated and no significant differences in the appearance of seeds showing ASV+ (data not shown) were observed.

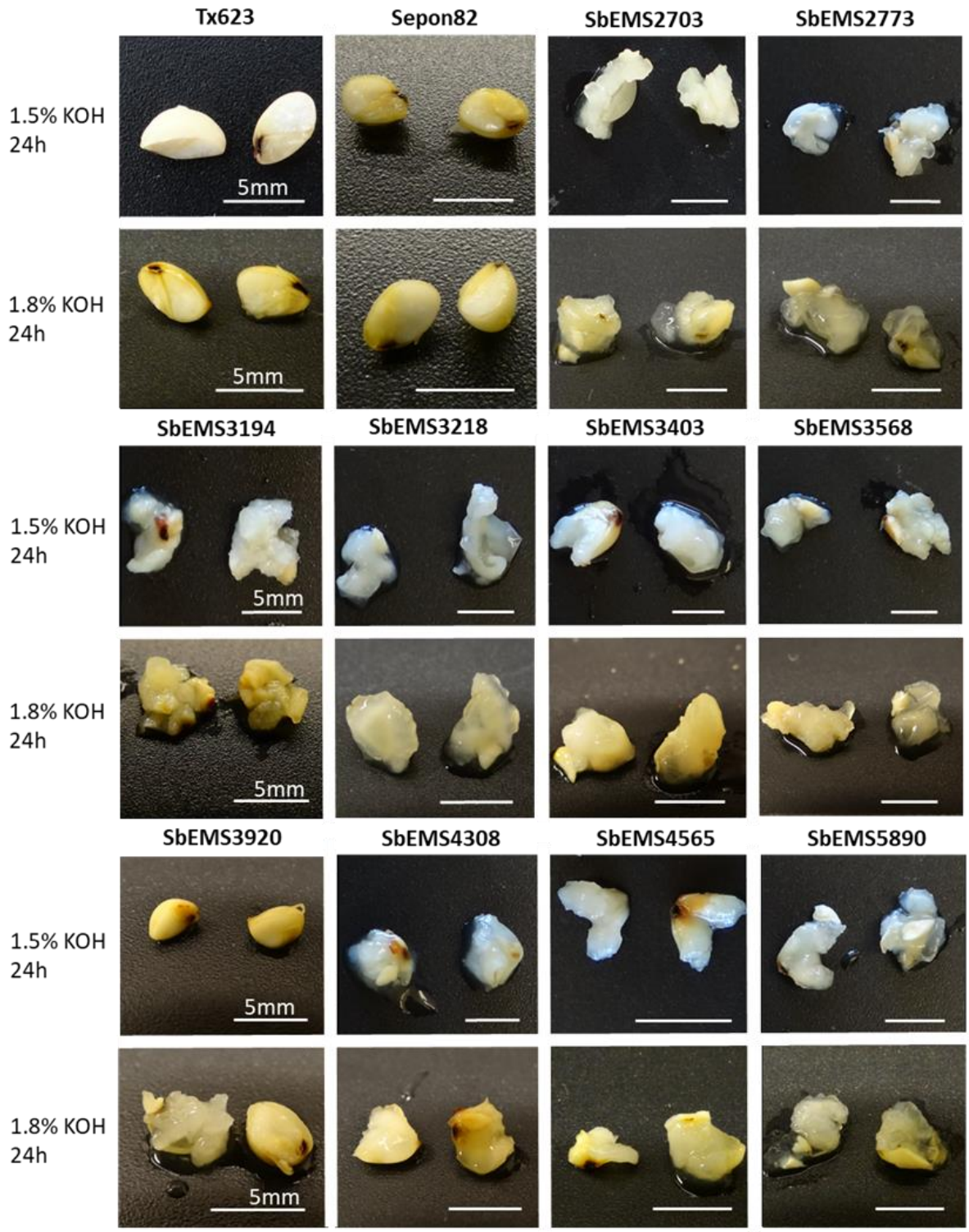


Figure 2.2 *Sorghum bicolor* EMS mutants exhibit ASV phenotypes of cut, not decorticated seeds after 24 h treatment with 1.5%KOH and 1.8%KOH. Note: Figs. 1 and 2 should be displayed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Native starches have physical and chemical properties that change depending on the plant species, genotype, composition of the starch granule and environmental factors (Biliaderis, 2009). A visual analysis of ASV phenotypes for sorghum grain produced in West Lafayette, IN and Guayanilla, PR demonstrated stable expression of ASV for some mutants and less stable expression for other mutants (Table 2.1).

Table 2.1 Alkali spreading values in sorghum and rice genotypes.

			West Lafayette 2016			Puerto Rico 2016/2017		
Genotype	**Type**	**ASV score[1]**	**# Plots**	**Overall # of Panicles**	**# Seeds ASV+ (average)** [1]	**# Plots**	**Overall # of Panicles**	**# Seeds ASV+ (average)** [1]
SbEMS 2703	EMS mutant	+	3	18	2.5	3	14	6.7
SbEMS 2773	EMS mutant	+	3	25	6.1	3	12	5.4
SbEMS 3194	EMS mutant	+	3	37	8	3	11	8
SbEMS 3218	EMS mutant	+	3	11	7.7	3	14	6.9
SbEMS 3403	EMS mutant	+	3	18	6.8	2	2	6.5
SbEMS 3568	EMS mutant	+	1	8	8	2	8	4.3
SbEMS 3920	EMS mutant	+	2	10	4	2	6	3.3
SbEMS 4308	EMS mutant	+	2	17	6.9	3	15	6.1
SbEMS 4565	EMS mutant	+	1	10	6.6	3	15	3
SbEMS 5890	EMS mutant	+	3	14	7.4	3	10	5.5
Sepon82	Check	-	1	Bulk	0	1	Bulk	0
Tx623	Check	-	1	Bulk	0	1	Bulk	0
CenturyPatna231	Rice check	-		1	0			
Magnolia	Rice check	+		1	6			

[1]ASV scores based on 1.8%KOH; 8 sorghum seeds tested per panicle; 6 rice seeds from GRIN samples were teste
Not all biological replicates tested are part of DSC analysis

The expression of ASV was somewhat variable within a sorghum panicle and between panicles of the same mutant as well as between SbEMS mutants (Table 2.1, Figure 2.2). This variation seems to be influenced by the environment. Kernel to kernel variation of up to 3 ASV points within a variety, variation in gelatinization and swelling between and within varieties, and variation between years tested have also been reported in rice (Juliano et al., 1982; Little et al., 1958). Non-starch components may also affect the ASV phenotypes. Waniska (1976) reported that the pericarp and pericarp thickness as well as the endosperm colour did not influence ASV

expression, but endosperm texture and endosperm type were important factors determining the strength of the ASV phenotype in sorghum. The mutants described in this study were created in the Tx623 background, and non-starch components such as seed storage protein are not expected to vary as much between genotypes as might be observed in the standing variation of sorghum germplasm.

2.5.2 Gelatinization Properties

Starch was extracted from grain samples using the protocol from Benmoussa and Hamaker (2011). Benmoussa and Hamaker (2011) reported that the methodology successful separates starch from proteins with very little protein is left in the final extracted starch sample. The amount of non-starch components like protein were therefore assumed to have no significant effects on the starch gelatinization using DSC.

Sample DSC thermograms of gelatinized sorghum starches are shown in Figure S.2.1. The sample from Tx623 exhibited a smaller T_c-T_o and a larger enthalpy than representative mutants SbEMS2703, SbEMS4565, SbEMS3920 and SbEMS4308.

Table 2.2. ANOVA of fixed effects of DSC data from West Lafayette 2016 and Puerto Rico 2016/2017

	Type III Test of Fixed Effects											
	T_o		**T_p**		**T_c**		**T_c-T_o**		**Enthalpy**		**Melting Point**	
Effect	F-value	Pr > F	F-value	Pr > F	F-value	Pr > F	F-value	Pr > F	F-value	Pr > F	F-value	Pr > F
Environment	0.83	0.38	0.75	0.40	3.88	0.07	3.46	0.08	24.11	0.00	0.49	0.50
Pedigree	109.9	<.0001	366.5	<.0001	95.27	<.0001	10.02	<.0001	10.26	<.0001	3.23	0.02
Environment*Pedigree	2.37	0.08	1.33	0.29	2.91	0.03	3.53	0.01	2.16	0.08	0.94	0.53

To= onset starch GT, Tp= peak starch GT, Tc= conclusion starch GT

The sorghum genotypes showed no significant genotype by environment interactions for the traits T_o, T_p, enthalpy and melting point (Table 2.2). A significant genotype by environment interaction was observed for the traits T_c and T_c-T_o; however, the environment was not significant and the pedigree had a very large effect. The data is therefore presented as a combined analysis

(Table 2.3). Sorghum starches exhibit starch GT in the range 69.7 °C to 85.2 °C for BTx623, 67.2 °C to 84.6 °C for Sepon82 (Table 2.3). The SbEMS mutants showed a broad range of starch GT from 56.0°C as the lowest T_o to 90.9 °C as the highest T_c, implying that each mutant could be used for specific food processes that require specific starch GTs and starch structures. All mutants exhibited a wider range of starch gelatinization (T_c-T_o) than the check BTx623 (Table 2.3), suggesting that it takes a longer time to gelatinize the mutant starch molecules. The enthalpies of the mutants, except SbEMS2773, were significantly lower than BTx623 (Table 2.3), implying that less energy is needed to gelatinize the mutant starches. All mutants, except SbEMS3403, exhibited significantly different T_p values than BTx623 samples (Table 2.3).

Table 2.3. Starch gelatinization characteristics (starch:water 1:3 ratio) of sorghum genotypes produced in West Lafayette 2016 and Puerto Rico 2016/2017 and rice checks.

			Starch Gelatinization Characteristics					
Genotype	**Type**	**Bio. Rep[1]**	**T_o [°C] lsmean**	**T_p [°C] lsmean**	**T_c [°C] lsmean**	**T_c-T_o [°C] lsmean**	**Enthalpy [J/g]**	**Melting Point [°C]**
SbEMS 2703	Mutant	6	68.3 ab	78.9 a	90.3 ab	22.0 ab	3.0 cd	98.2 a
SbEMS 2773	Mutant	6	68.5 a	78.7 a	90.9 a	22.5 a	4.2 abc	98.0 a
SbEMS 3194	Mutant	6	66.2 c	77.0 bc	87.1 cde	20.8 ab	2.0 d	97.3 a
SbEMS 3218	Mutant	6	67.8 abc	78.0 ab	88.3 bcd	20.4 abc	2.4 d	98.2 a
SbEMS 3403	Mutant	5	66.5 abc	76.3 cd	88.6 abcd	21.9 ab	3.3 bcd	98.4 a
SbEMS 3568	Mutant	4	66.3 bc	77.1 bc	87.6 cde	21.4 ab	2.1 d	98.0 a
SbEMS 3920	Mutant	5	58.7 d	66.7 f	78.9 f	20.0 abc	3.4 bcd	96.2 a
SbEMS 4308	Mutant	5	56.0 e	63.5 g	76.1 g	20.2 abc	3.1 cd	96.0 a
SbEMS 4565	Mutant	5	68.8 a	77.2 bc	88.4 bcd	19.6 bc	3.5 bcd	97.0 a
SbEMS 5890	Mutant	6	68.4 ab	78.7 a	89.2 abc	20.8 ab	2.2 d	98.1 a
Sepon82	Check	2	67.2 abc	72.7 e	84.6 e	17.4 cd	5.4 ab	98.0 a
Tx623	Check	2	69.7 a	73.7 de	85.2 de	15.5 d	5.8 a	97.6 a
Century Patna 231	Rice check	1	73.6 A	78.9 A	89.2 A	19.3 A	6.4 A	100.4 A
Magnolia	Rice check	1	60.8 B	69.3 B	80.1 B	15.6 B	2.8 B	97.7 A

T_o= onset starch GT, T_p= peak starch GT, T_c= conclusion starch GT

Values followed by the same letters in the same column are not significantly different ($p<0.05$).

[1] Two technical replications for a biological replicate; Sepon82 (one season 3 technical reps) and Tx623 three technical

[2] Tx623, Sepon82 N=4; SbEMS2773 N=11; SbEMS4565, SbEMS4308 N=9; SbEMS3403 N=7; SbEMS3920 N=8; Century Patna 231 N=1.

Even though all SbEMS mutants exhibited ASV+ phenotypes, there seem to be different types of SbEMS mutants based on starch gelatinization properties (Table 2.3). Previous studies on

rice have shown that a low ASV (1 to 2) indicates a high starch GT (> 74°C), while an ASV of 6 to 7 indicates low starch GT (< 70°C) (Table S.2.1) (Bhattacharya, 1979; Juliano, 1980; Juliano et al., 1964; Mariotti et al., 2010; Tan and Corke, 2002). Similarly, in this study, Century Patna 231 exhibited an ASV of 1 and had a starch GT of 73.6-89.2°C, while Magnolia exhibited an ASV of 7 and had a starch GT of 60.8 to 80.1°C (Table S.2.1, Table 2.3). The relationship between ASV and starch GT was less clear in sorghum. Some sorghum mutants with ASV+ phenotypes had lower starch GT values and others exhibited higher starch GT values compared to the controls (Table 2.3). Thus, it appears that ASV can be used to identify genotypes with modified starches but these genotypes need to be screened for starch GT to determine final food processing applications and to define breeding strategies for sorghum starch improvement. The mutants SbEMS3920 and SbEMS4308 differed significantly from the checks and other mutants and exhibited the lowest T_o, T_P, and T_C starch GTs (Table 2.3). SbEMS2703, SbEMS2773 and SbEMS5890 exhibited the highest T_c (Table 2.3). These SbEMS mutants and SbEMS3194, SbEMS3218, SbEMS3568, SbEMS4565 also exhibited higher T_p (Table 2.3).

Table 2.4. Correlations between ASV score and starch gelatinization characteristics in sorghum from West Lafayette 2016 and Puerto Rico 2016/2017.

Comparison	West Lafayette-2016	Puerto Rico-2016/2017	Combined
ASV Score vs. T_o	-0.24	-0.11	-0.16
ASV Score vs. T_p	0.10	0.11	0.11
ASV Score vs. T_c	0.04	0.13	0.09
ASV Score vs. $T_c$minusT_o	0.62**	0.62**	0.60*
ASV Score vs. Enthalphy	-0.74*	-0.41	-0.53*
ASV Score vs. Melting Point	-0.10	0.00	-0.05

**Pr<0.0001; * Pr<0.01

ASV score was not correlated with T_o, T_p, or T_c (Table 2.4). However, a negative relationship was observed between ASV score and enthalpy, although the correlation varied with environment conditions (-0.74 in West Lafayette and -0.41 in Puerto Rico; Table 2.4). Puerto Rico

is a tropical environment with shorter days, while West Lafayette is in the temperate region with longer days. The enthalpies are slightly larger in the Puerto Rico environment; however, the directionality of the correlations is similar across environments and the strength of association does not differ statistically across environments using Fishers' Z test. Sorghum ASV scores also exhibited a positive relationship with T_c-T_o (0.62 in West Lafayette and 0.62 in Puerto Rico; Table 2.4). Thus, a score of ASV+ indicated lower enthalpies and larger T_c-T_o values as compared to the control, BTx623.

The SbEMS mutants with altered ASV values could have functional mutations in genes that cause lower starch GTs, a larger range of starch GT or changed enthalpies. Two DSC peaks for high ASV samples were also reported for rice (Cuevas et al., 2010) together with a weak correlation between ASV and DSC parameters. In that study, ASV+ was not only associated with a lower starch GT but also higher DSC parameters, which is similar to the results obtained in this study on sorghum. The starch GT in other studies was not always determined using the DSC instrument, and these differences in methodology could explain why the literature remains unclear on whether ASV is correlated with DSC parameters (Cuevas et al., 2010; Juliano et al., 1964; Mariotti et al., 2010; Shu et al., 2006).

2.6 Conclusions

The modified Alkali Spreading Value (ASV) assay can be used to screen for sorghum genotypes with modified starch quality attributes. In this study, mutants with strong ASV phenotypes exhibited starch GT that ranged from lower to higher than the wild-type Tx623. Mutants with ASV+ phenotypes exhibited lower enthalpies and greater starch GT ranges. Changes in starch GT can influence crumb structure and mouth-feel of cooked products. Changes in starch GT can also impact starch digestibility and nutritional characteristics. The identification of

sorghum genotypes with modified starch properties would be of great advantage to the sorghum malting industry to develop varieties with lower starch GT. Low starch GT varieties could also enhance functionality of sorghum in extrusion cooking, allowing for reduced extrusion temperatures and improved energy efficiencies (Bhattacharya and Hanna, 1987). The development of the low-cost ASV assay will help to identify sorghum genes and genotypes that contribute to improved starch quality and structures influencing food product quality.

2.7 Acknowledgments/Funding Sources

This research was 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). The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. Program activities are funded by the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A-13-00047.

This research was also partially supported by the Bill & Melinda Gates Foundation.

2.8 Supplementals

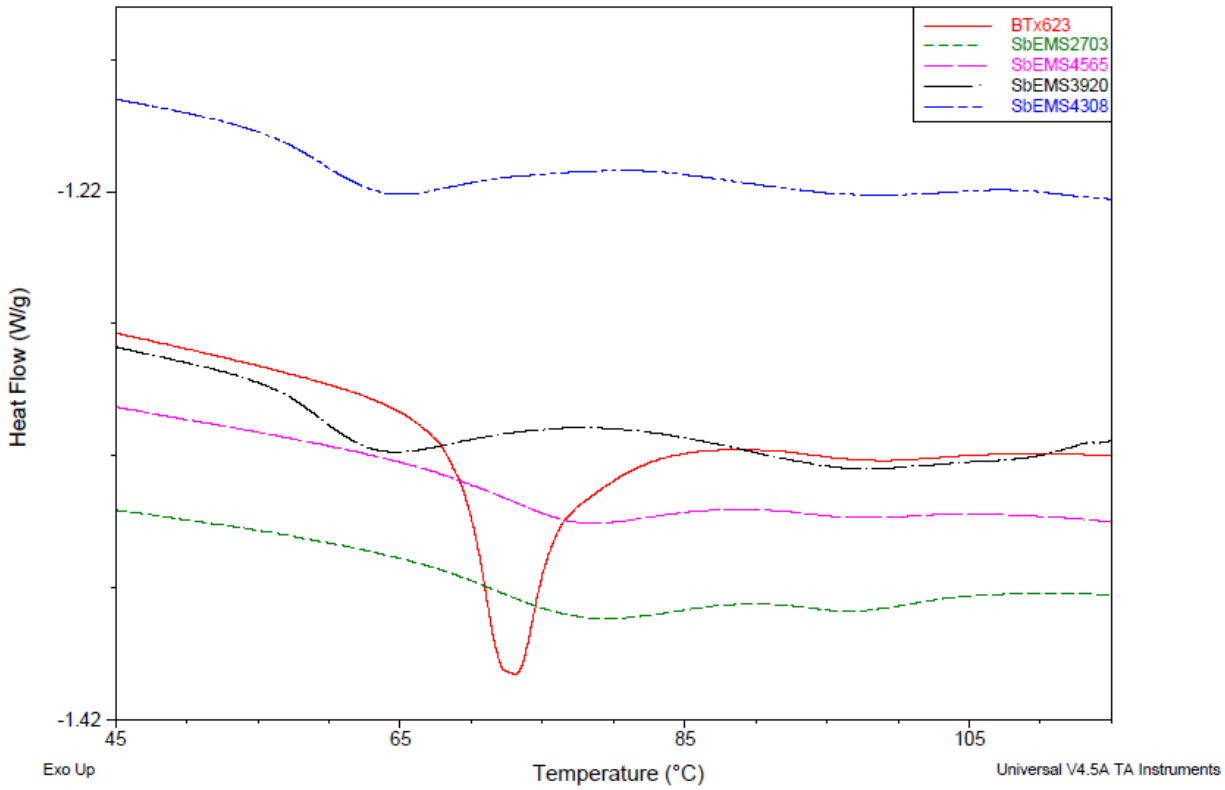


Figure S2.1. DSC thermograms of representative sorghum genotypes.

Table S2.1: Oryza sativa varieties with known alkali test scores and starch gelatinization temperatures.

PI Number	Variety	Origin	Endosperm	ASV Score	ASV-spreading 1.7% KOH Mar/Sept [1]	ASV-clearing 1.7% KOH Mar/Sept [1]	ASV-spreading 1.5% / 1.7% KOH [2]	ASV-clearing 1.5%/ 1.7% KOH [2]	ASV-spreading 1.5% KOH [3]	GT Information	GT (°C) Visco Amylograph-Brabender
PI389996	Zenith	USA	Non-waxy	7.0	5.9/6.0 6.0/6.0 6.0/7.0 5.7/7.0 5.0/5.8	6.1/5.6 6.2/5.1 6.6/5.8 6.4/6.8 6.6/5.0				Low	
PI215486	Vialone Nano	Italy	Non-waxy	6.5					6	Low	MVA: 64.2 [3]
PI210983	Early Prolific	USA	Non-waxy	2.5	2.0/2.0 2.7/2.2	1.0/1.0 1.7/1.2				Intermediate	
PI422509	Taichung 65	Taiwan	N/A	7.0			6.7/6.6	4.8/5.2		no data	BEPT: 62-67/59-67 [2]
CIor8989	Sunbonnet	USA	Non-waxy	3.5	4.7/3.1 4.3/3.0 4.8/2.9	3.1/2.2 2.6/2.1 2.0/2.2				Intermediate	
CIor8322	Bluebonnet	USA	Non-waxy	4	2.9/2.7	2.1/1.7				Intermediate	
CIor8318	Magnolia	USA	Non-waxy	6.0	5.8/6.0 6.0/6.0	5.4/4.8 5.0/4.0				Low	
PI313561	Kaohsiung 68	Taiwan	Non-waxy	6.7			6.7/6.6	4.7/5.3		Low	BEPT: 57-66/60-64 [2]
PI389997	Nato	USA	Non-waxy	7.0	6.0/7.0 6.0/6.7	6.0/5.9 6.0/5.3				Low	
PI366134	Century Patna 231	USA	N/A	N/A	2.0/2.0-2.1 6 diff lots	1.0/1.0-1.1 6 diff lots	2.2/2.7	1.2/1.7		no data	BEPT: 68-75/72.5-76.5 [2]
CIor1600	Colusa	USA	Non-waxy	7.0	6.0/6.0 6.5/7.0 6.0/5.8	6.0/5.0 7.0/7.0 5.7/5.0				Low	
PI388605	Fortuna	USA	Non-waxy	5.7	2.8/3.0 5.0/3.5	2.2/1.9 2.2/2.1				Low	

All values are for polished rice, whole seeds. ASV values for reference 22 are from crop 1, crop 2. ASV value for reference 20 are from two different month and different years. All Data obtained from USDA ARS if not otherwise indicated: 1 = Little et al., 1958; 2 = Juliano et al., 1964; 3= Mariotti et al., 2010.

2.9 References

Addo-Quaye, C., Tuinstra, M., Carraro, N., Weil, C., Dilkes, B.P., 2018. Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. G3 (Bethesda) 8, 1079-1094.

Akingbala, J. O., Gomez, M. H., Rooney, L. W., Sweat, V. E., 1988. Thermal properties of sorghum starch. Starch/Stärke 40, 375-378.

BeMiller, J. N., Huber, K. C., 2008. Carbohydrates. In: Damodaran, S., Parkin, K., Fennema, O. (Eds.), Fennema's Food Chemistry. CRC press, FL, pp. 83-154.

BeMiller, J., 2014. Carbohydrate Analysis. In: Nielsen, S. (ed.), Food Analysis. Springer, New York, pp. 147-180.

Benmoussa, M., Hamaker, B. R., 2011. Rapid small-scale starch isolation using a combination of ultrasonic sonication and sucrose density separation. Starch/Stärke 63, 333-339.

Beta, T., Corke, H., Rooney, L. W., Taylor, J. R.N., 2000. Starch properties as affected by sorghum grain chemistry. J. Sci. Food Agric. 81, 245-251.

Bhattacharya, K. R., Sowbhagya, C. M., 1972. An improved alkali reaction test for rice quality. J. Food Sci. Technol. 7, 323-331.

Bhattacharya, K. R., 1979. Gelatinization temperature of rice starch and its determination. In: International Rice Research Institute, Proceedings of the workshop on chemical aspects of rice grain quality. International Rice Research Institute, Philippines, pp. 231-249.

Bhattacharya, M., Hanna, M. A., 1987. Kinetics of starch gelatinization during extrusion cooking. J. Food Sci. 52, 764-766.

Biliaderis, C. G., 2009. Structural transitions and related physical properties of starch. In: BeMiller, J., Whistler, R. (Eds.), Starch Chemistry and Technology. Elsevier, Academic Press, Netherlands, pp. 293 – 372.

Cuevas, R. P., Daygon, V. D., Corpuz, H. M., Nora, L., Reinke, R. F., Waters, D. L. E., Fitzgerald M. A., 2010. Melting the secrets of gelatinization temperature in rice. Funct. Plant Biol. 37, 439-447.

FAOSTAT, 2016. Food and Agriculture Organisation of the United Nations. FAOSTAT Database. http://www.fao.org/faostat/en/ (Accessed October 2016).

House, L. R., Gomez, M., Murty, D. S., Sun, Y., Verma, B. N., 2000. Development of Some Agricultural Industries in Several African and Asian Countries. In: Smith, C. W., Frederiksen, R. A. (Eds.), Sorghum Origin, History, Technology and Production. John Wiley & Sons, Inc., New York, pp. 131-190.

Jane, J., 2009. Structural Features of Starch Granules II. In: BeMiller, J., Whistler, R. (Eds.), Starch Chemistry and Technology. Elsevier, Academic Press, Netherlands, pp. 193 – 236.

Juliano, B. O., Bautista, G. M., Lugay, J. C., 1964. Reyes, A. C. Studies on the physicochemical properties of rice. J. Agric. Food Chem. 12, 131-138.

Juliano, B. O., 1980. Rice: Recent Progress in Chemistry and Nutrition. In: Inglett, G. E., Munck, L., (Eds.), Cereals for Food and Beverages: Recent Progress in Cereal Chemistry. Academic Press, New York, 409-428.

Juliano, B. O., Blakeney, A. B., Butta, I., Castillo, D., Choudhury, N. H., Iwasaki, T., Shibuya, N., Kongseree, N., Lapis, E. T., Murty, V. V. S., Paule, C. M., Perez, C. M., Webb, B. D., 1982. International Cooperative Testing of the Alkali Digestibility Values for Milled Rice. Starch/Stärke 34, 21 – 26.

Kimber, C. T., 2000. Origins of Domesticated Sorghum and Its Early Diffusion to India and China. In: Smith, C. W., Frederiksen, R. A. (Eds.), Sorghum Origin, History, Technology and Production. John Wiley & Sons, Inc., New York, pp. 3-98.

Little, R. R., Hilder, G.B., Dawson, E.H., 1958. Differential effect of dilute alkali on 25 varieties of milled white rice. Cereal Chem. 35, 111-126.

Mariotti, M., Fongaro, L., Catenacci, F., 2010. Alkali spreading value and Image analysis, Journal of Cereal Science 52, 227-235.

Mutters, R.G., Thompson, J.F., 2009. Rice quality handbook (Vol. 3514). UCANR Publications.

Ragheb, A. A., Abd El-Thalouth, I., Tawfik, S., 1995. Gelatinization of Starch in Aqueous Alkaline Solutions. Starch/Stärke 46, 338-345.

Shannon, J. C.; Garwood, D. L.; Boyer, C. D., 2009. Genetics and Physiology of Starch Development. In: BeMiller, J., Whistler, R. (Eds.), Starch Chemistry and Technology. Elsevier, Academic Press, Netherlands, 23-82.

Shu, X, Shen, S., Bao, J., Wu, D., Nakamura Y., Shu, Q., 2006. Molecular and biochemical analysis of the gelatinization temperature characteristics of rice (*Oryza sativa* L.) starch granules. J. Cereal Sci. 44, 40-48.

TA instruments. https://www.tainstruments.com/support/software-downloads-support/downloads/. TA instruments universal analysis software, 2000, version 4.5A. Accessed, 2017.

Tan, Y., Corke, H., 2002. Factor analysis of physiochemical properties of 63 rice varieties. J. Sci. Food Agr. 82, 745-752.

Taylor, J. R. N., 1992. Mashing with malted grain sorghum. J. Am. Soc. Brew. Chem. 50, 13-18.

Taylor, J. R. N., Dewar, J., Taylor, J., Ascheraden, R.F., 1997. Factors affecting the porridge-making quality of south african sorghums. J. Sci. Food Agr. 73, 464-470.

Taylor, J., Dewar, J., 2000. Fermented Products: Beverages and Porridges. In: Smith, C. W., Frederiksen, R. A. (Eds.), Sorghum Origin, History, Technology and Production. John Wiley & Sons, Inc., New York, pp. 751-796.

USDA ARS, 2017. United States Department of Agriculture (USDA), Agriculture Research Service (ARS), Germplasm Resources Information Network (GRIN). Database https://www.ars-grin.gov/npgs/aboutgrin.html (Accessed 2017).

Waniska, R. D., Rooney, L. W., 2000. Structure and Chemistry of the Sorghum Caryopsis. In: Smith, C. W., Frederiksen, R. A. (Eds.), Sorghum Origin, History, Technology and Production. John Wiley & Sons, Inc., New York, 2000, pp. 649-688.

Waniska, R. D. (1976). Methods to assess quality of boiled sorghum, gruel and chapaties from sorghum with different kernel characteristics. M.Sc. thesis. Texas A&M University, College Station, Texas, USA.

CHAPTER 3. MUTATIONS IN SORGHUM *SBEIIB* AND *SSIIA* AFFECT ALKALI SPREADING VALUE, STARCH COMPOSITION, THERMAL PROPERTIES AND FLOUR VISCOSITY

A version of this chapter has been submitted to a peer reviewed academic journal for publication. In Review at the Journal of Theoretical and Applied Genetics. Will be published as open access (CC BY) with open choice.

Stefanie Griebel[1], Richard P. Westerman[2], Adedayo Adeyanju[1], Charles Addo-Quaye[3], Bruce A. Craig[4], Clifford F. Weil[1], Suzanne M. Cunningham[1], Bavesh Patel[5], Osvaldo H. Campanella[56], and Mitchell R. Tuinstra[1]

[1] Department of Agronomy, Purdue University, West Lafayette, IN 47907
[2] College of Agriculture Administration, Purdue University, West Lafayette, IN 47907
[3] Division Natural Sciences and Mathematics, Lewis-Clark State College, Lewiston, ID 83501
[4] Department of Statistics, Purdue University, West Lafayette, IN 47907
[5] Whistler Carbohydrate Research Center, Purdue University, West Lafayette, IN 47907
[6] Department of Food Science and Technology, The Ohio State University, Columbus, OH, 43210-1007

3.1 Key message

Seven novel alleles of *SBEIIb* and two alleles of *SSIIa* co-segregate with the ASV phenotype and contribute to distinct starch quality traits important for food processing industries.

3.2 Declaration of Author Contributions to this Chapter

Stefanie Griebel: wrote paper, designed research, performed research, analysed data, created tables, figures, graphs. Richard P. Westerman: support on next-generation sequencing analysis; Adedayo Adeyanju: provided technical support on DNA marker development (TetraPrimer), helped with genotyping; Charles Addo-Quaye: next-generation sequencing analysis of a mutant; Clifford F. Weil, Mitchell R. Tuinstra: provided new tools; Osvaldo Campanella: provided DSC and RVA instrument; Patel Bhavesh: provided technical support using the Rapid

Visco Analyzer; Suzanne Cunningham: helped on amylose analyses; Mitchell R. Tuinstra, Bruce A. Craig, Clifford F. Weil, Osvaldo H. Campanella guided research. All co-authors reviewed and edited the paper. M.R.T. major advisor and project lead.

3.3 Acknowledgments

This research was 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). The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. Program activities are funded by the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A-13-00047

3.4 Abstract

Sorghum is an important food crop for millions of people in Africa and Asia. Whole genome re-sequencing of sorghum EMS mutants exhibiting an alkali spreading value (ASV) phenotype revealed candidate SNPs in *Sobic.004G163700* and *Sobic.010G093400*. Comparative genomics identified *Sobic.010G093400* as a starch synthase IIa, and *Sobic.004G163700* as a *starch branching enzyme IIb*. Segregation analyses showed that mutations in *Sobic.010G093400* and *Sobic.004G163700* co-segregated with the ASV phenotype. Mutants in *SSIIa* exhibited no change in amylose content but expressed lower final viscosity and lower starch gelatinization temperature (GT) than starches from non-mutant plants. The *sbeIIb* mutants exhibited significantly higher amylose levels and starch GT and lower viscosity compared to non-mutant starches and ssIIa mutants. Mutations in *SBEIIb* had a dosage dependent effect on amylose content. Double mutants of *sbeIIb* and *ssIIa* resembled their *sbeIIb* parent in amylose content, starch thermal properties and

viscosity profiles. These variants will provide opportunities to produce sorghum varieties with modified starch end-use qualities likely important for the beer brewing and baking industries and specialty foods for humans with diabetes.

3.5 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is important in drought-prone environments (FAO 2019; Kimber 2000). It is a staple crop in Africa and Asia and is utilized to produce traditional foods such as stiff porridge and beverages (House et al. 2000). Indigenous to Africa, it is less expensive than other imported flours (Taylor and Dewar 2000); thus, it has a great economic value for smallholder farmers.

Comparative genomics is a valuable tool to compare gene sequences and putative functions across grass species (Paterson et al. 2004; Paterson et al. 2009). *Sorghum bicolor* is diploid with 2n=20 and a genome size of approximately 730Mb (Paterson et al. 2009; Rooney 2000). Sorghum and rice shared a whole genome duplication event approximately 70 Mya before both diverged (Paterson et al. 2004). Sorghum and maize diverged only approximately 11.9 Mya (Swigoňová et al. 2004).

BTx623, the sorghum reference genome, has been chemically mutagenized to generate genetic mutant resources (Addo-Quaye et al. 2017; Addo-Quaye et al. 2018; Jiao et al. 2016). EMS (ethyl methane sulfonate) creates random mutations across the genome (Greene et al. 2003; Loveless 1958; Westergaard 1957). Such populations are an important tool for gene identification in plants using forward or reverse genetic approaches (Addo-Quaye et al. 2017; Addo-Quaye et al. 2018; Greene et al. 2003; Jiao et al. 2016; Krothapalli et al. 2013). A common strategy to link identified SNP mutations to a phenotype is a co-segregation analysis (Addo-Quaye et al. 2017; Jiao et al. 2016; Krothapalli et al. 2013; Zhang et al. 2004).

Starch is a major component of sorghum flour and provides unique texture and nutritional values into foods (BeMiller 2014; BeMiller and Huber 2008; Biliaderis 2009). The two major starch fractions amylose and amylopectin and their ratio influence starch functionality (BeMiller and Huber 2008; Jane 2009; Tetlow et al. 2004), where amorphous and crystalline regions play an important role (Tetlow and Emes 2014). Both polymers and their ratio impact characteristics such as paste viscosity, gel consistency, thickening power, and starch GT (BeMiller and Huber 2008; Bouvier and Campanella 2014; Shannon et al. 2009). The melting of the crystalline regions, mainly formed by amylopectin, are critical for starch granule swelling, loss of crystallinity and starch gelatinization (BeMiller and Huber 2008; Bouvier and Campanella 2014). Some of the granule physio-chemical and structural characteristics are directly related to granule swelling, starch pasting properties (Desam et al. 2018) and is positively correlated with the peak viscosity (Bouvier and Campanella 2014). In sorghum, amylose levels are positively correlated with final viscosity and setback, and negatively with breakdown (Hill et al. 2012) and peak viscosity (Beta et al. 2000; Hill et al. 2012). A sorghum study reported a positive correlation between amylose content and peak starch GT (T_p) (Beta et al. 2000). Thus, end-product quality depends on the functionality of starches, which is influenced by starch characteristics and its GT (BeMiller and Huber 2008; Biliaderis 2009).

The relative amounts and structural characteristics of these two polymers are under genetic control (Tetlow et al. 2004). The starch synthesis pathway (SSP) is a conserved pathway across cereal species (Tetlow et al. 2004). Located in the endosperm of the seed, adenosine 5’ diphosphate glucose pyrophosphorylase (AGPase), granule bound starch synthases (GBSSs), starch synthases (SSs), starch branching enzymes (SBEs), and two types of debranching enzymes, isoamylases and pullulanases, are the key group of enzymes involved in starch biosynthesis (Tetlow et al. 2004).

The different enzymes vary in their catalytic activities (Guan and Preiss 1993; Nakamura 2015; Tetlow et al. 2004; Tetlow and Emes 2014). A soluble SS and different SBE's in sorghum showed high similarity to maize soluble SS and SBE's (Boyer 1985). The SBEIIb of sorghum showed high sequence similarity with other cereals (Mutisya et al. 2003).

Amylose extender (*ae*) is a mutant starch type resulting in high amylose starches in maize (Li et al. 2008; Liu et al. 2012a; Nakamura 2015b; Shannon et al. 2009; Tetlow and Emes 2014) and rice (Nakamura 2015b; Nishi et al. 2001; Tetlow and Emes 2014). *Amylose extender* is regulated by *SBEIIb* (Jane 2009; Nakamura 2015b; Nishi et al. 2001; Tetlow and Emes 2014). A recent sorghum study on natural genetic variation reported that mutations in *SSIIa* and *SBEIIb* result in increased amylose values (Hill et al. 2012). The *Su2* gene codes for SSIIa in corn (Liu et al. 2012b; Zhang et al. 2004) and contributes to low starch GT (Liu et al. 2012b; Preiss 2009; Shannon et al. 2009; Zhang et al. 2004). In rice, the *ALK* (SSSII-3) gene controls starch GT (Gao et al. 2011). A sorghum study on *SBEIIb* and *SSIIa* reported alleles in both genes that result in both low and high starch GT (Hill et al. 2012). High amylose in sorghum was associated with higher starch GT (Hill et al. 2012). The *Su2* wild type allele is completely dominant over the *su2* allele (Shannon et al. 2009).

Hot starch gelatinization occurs when a starch water dispersion is heated (BeMiller and Huber 2008; Bhattacharya 1979), while cold gelatinization occurs by adding an alkali solution like potassium hydroxide (KOH) or sodium hydroxide (NaOH) to starch (Bhattacharya 1979; Ragheb et al. 1995). The alkali spreading value (ASV) of rice is determined using KOH on rice seeds, (Little et al. 1958). A high ASV score is negatively correlated with starch GT in rice (Bhattacharya 1979; Bhattacharya and Sowbhagya 1972; Juliano et al. 1964; Mariotti et al. 2010; Tan and Corke 2002). In a recent study, the ASV test was adapted and modified to identify sorghum genotypes

with modified starch thermal properties exhibited lower to higher starch GTs compared to BTx623, thus providing new opportunities to develop new food products (Griebel et al. 2019).

In this study, gene sequencing and genetic segregation studies were conducted to identify the genes controlling the alkali spreading phenotype in sorghum. Grain samples produced from these mutants were characterized for variation in starch quality and related food processing parameters.

3.6 Materials and methods

3.6.1 Plant material

A sorghum EMS population was developed as described before (Addo-Quaye et al. 2018). The seed and flour samples from PR16/17 described in Griebel et al. (2019) were used for further analyses. Ten SbEMS mutants and controls BTx623, Macia, and Sepon82 were also grown in West Lafayette 2017 (WL17) and 2018 (WL18). Two biological replicates per genotype and year were chosen (WL18 three replicates for BTx623) and tested for ASV, amylose content and paste viscosity profiles. A-Lines were developed by crossing each mutant as pollen donor to ATx623 followed by backcrossing over several generations. The B-Line was used as a recurrent parent in each generation. The A-lines were phenotyped and genotyped.

The mutants will be registered as genetic stocks in the USDA National Plant Germplasm System.

3.6.2 Allele dosage test

The A-lines for each mutant described above were part of the allele dosage test. In 2017, the A-line for SbEMS2703 was not available so hybrids were made by hand emasculation. The SbEMS3920 mutant was not included as no co-segregation data were available at that point. To obtain zero allele dosage of the mutant allele BTx623 was used as pollen donator and crossed to ATx623. A single dosage of the mutant allele was created by crossing the mutants onto ATx623.

Two dosages of the mutant allele were created by crossing BTx623 onto the A-Line (or hand emasculated panicle) of each EMS mutant. Three doses were created by self-pollinating the B-line (West Lafayette 2017) or the B Line was crossed to the corresponding A-Line (West Lafayette 2018) except a self-pollinated B-line was used when seed was limited (SbEMS 2703, SbEMS 3218, SbEMS 3403). The samples were analysed for ASV with 32 seeds and for amylose content. The experiment was set up to obtain two biological replicates per genotype and allele dosage, which was influenced by crossing success and seed set in the field.

3.6.3 Double mutant creation

Selected B-lines of EMS mutants were hand emasculated in 2017 and crossed with each other. The harvested F_1 seeds were grown in the greenhouse and genotyped for both parental alleles. The verified F_1 plants were self-pollinated and the F_2 population was grown in the greenhouse and genotyped to identify homozygous double mutant F2 plants. The homozygous double mutants were transplanted in the field in 2018, self-pollinated, and harvested as single panicles. The parental genotypes were also grown in the field in 2018. The seeds of double mutants and two biological replicates of the parents were subject to further phenotypic analyses.

3.6.4 Alkali spreading test

The alkali spreading test was conducted as described in Griebel et al. (2019) with 1.8% KOH for 24h. 32 seeds were tested and in rare cases fewer seeds if seed quantity was limited. The seeds for the allele dosage test were cut horizontally so that only the triploid endosperm was evaluated for ASV with 32 seeds per panicle.

3.6.5 DNA preparation

EMS mutants exhibiting an ASV+ phenotype in Griebel et al. (2019) were subjected to DNA extraction. Leaf segments of single plants were sampled and later pooled for NGS. A modification of the standard CTAB DNA protocol (Murray and Thompson 1980) was used. A cold pestle and mortar were used to grind the frozen leaf tissue into a fine powder. For each sample, 0.8µL BME was added to 800µL Extraction Buffer (1.2 M NaCl, 100 mM Tris pH 8, 20 mM EDTA pH8, 2% CTAB (1% = 1 g / 100 mL), 0.1% BME (day of use)) in a 2mL Eppendorf tube. The samples were heated to 60°C and the tissue was transferred to a 2ml Eppendorf tube with 800µL Extraction Buffer and BMR. The samples were vortexed for 30 sec and incubated at 60°C for >45 min with mixing every 15 min. The samples were cooled to room temperature for 10 min and 800µL Chloroform Mixture (24 parts chloroform, 1 part isoamyl alcohol) was added and mixed by inversion. The samples were centrifuged at 12,000 rpm (3220 x g) for 10 minutes. The aqueous layer was transferred to new 2mL (50mL) Eppendorf tube with 1070µL Dilution Buffer (100 mM Tris pH 8, 20 mM EDTA pH 8, 2% CTAB) and mixed by swirling. The samples were incubated at 60°C for 30 minutes. The samples were centrifuged at 12,000 rpm (3220 x g) for 10 minutes and the supernatant was discarded. 530µL Wash Buffer (3 parts TE Buffer, 7 parts ethanol) was added to each sample and gently mixed. The samples were incubated at room temperature for 15 minutes, centrifuged at 12,000 (3220 x g) for 10 minutes, and the supernatant was discarded. 1µL of RNaseA (10 mg/mL) was added to 53 µL high-salt TE (1 M NaCl, 10 mM Tris pH 8, 2 mM EDTA pH 8, RNaseA 50 µg/mL (day of use)) and used to re-suspend the pellet. The samples were transferred to a 1.5ml Eppendorf tube and incubated at 60°C for 15 minutes. 212 µL H_2O was added to each sample and mixed by inversion. 530 µL pure EtOH were added to each sample and mixed by inversion. The samples were incubated at -80°C for 30 minutes in a metal rack. The samples were centrifuged 12,000 (3220 x g) for 10 minutes and the supernatant was discarded.

The pellet was washed with 500 µL 70% ethanol. The ethanol was poured off each sample. The samples were air-dried for 20 to 30 minutes in the fume hood until the DNA turned clear. The pellets were re-suspended in 25 µL TE Buffer (10 mM Tris pH 8, 1 mM EDTA pH 8).

The fast DNA extraction method described by Xin et al. (2003) was used for DNA extractions for co-segregation analyses.

3.6.6 Next Generation Sequencing (NGS) and analyses

Whole-genome re-sequencing data for the sorghum samples were generated at the Purdue Genomic Core facility. Sorghum mutants SbEMS 3218, SbEMS 3568 and SbEMS 4308 had undergone next-generation sequencing and genomic analyses as described in reference (Addo-Quaye et al. 2018). The other EMS mutants SbEMS 2703, SbEMS 2773, SbEMS 3194, SbEMS 3403, SbEMS 3920, SbEMS 5890, SbEMS 4565 and as quality control SbEMS 3568 and SbEMS 4308 were re-sequenced using Illumina HiSeq2500 with 100bp paired-end reads.

The SbEMS 3218 sample was processed as described in reference (Addo-Quaye et al. 2018). All other mutant samples were processed as follows.

The filtered NGS R1 and R2 files obtained from the Purdue Genomics Core were cleaned of contaminants like the viral PhiX174 genome. Therefore, the R1 and R2 files were aligned to the indexed Phix174 genome using *Bowtie 2* short reads aligner (Langmead and Salzberg 2012). Next, NGS reads with origins in the Sorghum mitochondrial genome were filtered by using *Bowtie 2* to detect evidence of alignment to the mitochondrial genome (NC 008360.1: https://www.ncbi.nlm.nih.gov/nuccore /NC_008360.1) downloaded (NCBI sequence repository 2016).

A reference genome based read mapping approach was used for SNP calling and quality filtering. The *Sorghum bicolor* reference genome assembly and annotation sequences (version 3.0;

Sbicolor_313_v3.1) were downloaded from *Phytozome* (Goodstein et al. 2012, version 11; Phytozome 11 2016, version 11). The reference genome assembly sequences were indexed using Bowtie2 *build*. The filtered NGS paired-end reads were aligned to the indexed sorghum reference genome using *Bowtie 2 (*Langmead and Salzberg 2012). The reference genome sequences were indexed separately using the *faidx* and *CreateSequenceDictionary* programs in the SAMtools alignment manipulation software (Li et al. 2009) and the GATK genome analysis toolkit packages (McKenna et al. 2010; DePristo et al. 2011), respectively. The tool GATK was used for SNP calling (McKenna et al. 2010; DePristo et al. 2011). The SAM-formatted output files containing the aligned sorghum NGS reads were sorted and indexed using the *sort* and *index* programs in the *SAMtools* software package. NGS read duplicated were tagged using the *MarkDuplicates* tool in the *Picard Tools* package (Picard Toolkit 2016). Variant detection in the resulting sorted BAM files was performed using the *GATK HalotypeCaller* tool (DePristo et al. 2011; McKenna et al. 2010) for variant discovery.

We will submit the SNP data (vcf files) to a repository for genetic polymorphisms such as dbSNP, dbVar or European Variation Archive (EVA).

3.6.7 Functional annotations of SNP variation

The effects of identified SNPs on the sorghum gene function were identified using *snpEff* variant effect annotation tool (Cingolani et al. 2012). Using the *Sorghum bicolor* reference genome sequences in a FASTA-formatted file, and the genome annotation information available in the *.gff3* file obtained from *Phytozome* (Goodstein et al. 2012, version 11), the default *snpEff* configuration file was updated with a new database name for the sorghum genome. These files were needed to create a new database using *snpEff build-gff3* command. A snpEff-annotated variant call *.vcf*

output file containing the variant annotations was created using as input *.vcf* files containing the GATK Haplotype Caller variants.

3.6.8 Designation of starch-synthesis genes (candidate gene approach)

A list of candidate genes involved in sorghum starch biosynthesis in the amyloplast was obtained from Campbell et al. (2016 table 1). The genes *Sobic.007G101500* and *Sobic.009G127500* were also added to the list (Goodstein et al. 2012, version 11; Phytozome 11 2016, version 11).

3.6.9 Multiple sequence alignment - protein sequence comparison

The protein sequence from *Sobic.004G163700* and *Sobic.010G093400* was downloaded from *Phytozome.* The sequence was blasted against all green plants, viridiplantae, in NCBI using tBLASTn (Altschul et al. 1990; Camacho et al. 2009). More and less related species were selected and the protein sequence downloaded in fasta format. Multiple sequence alignment was conducted using Clustal Omega (https://www.ebi.ac.uk/Tools/ msa/clustalo/ , Chojnacki et al. 2017). Clustal Omega results were imported into Simple Phylogeny from the ClustalW2 package (https://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=simple_ phylogeny&sequence= clustalo-I20181231-144824-0717-67809619-p1m) and distance was determined using the UPGMA clustering method (Chojnacki et al. 2017).

3.6.10 Co-segregation analyses

The SbEMS mutants were crossed to Tx623bmr6 (except SbEMS3920 was un-successful in crossing). F_1 plants were selected and self-pollinated to create biparental F_2 populations. Panicles of F_2 plants were self-pollinated and $F_{2:3}$ seeds were screened for ASV with 32 seeds per panicle and in rare cases less if limited by seed number. The number of panicles per population screened varied from 30 for SbEMS3403 to 112 for SbEMS2773. F_3 plants of those $F_{2:3}$ panicles were grown

on a sand bench, tissue sampled, and genotyped. Each panicles' ASV phenotype and the corresponding SNP genotype (M=mutant, H=heterozygous, W=wild type) was determined. A Chi-Square test was conducted to determine goodness of fit for the 3:1 phenotypic ratio as well as the 1:2:1 genotypic ratio.

3.6.11 Primer design, PCR and SNP genotyping (CAPS, dCAPS, tetraprimer)

CAPS (Cleaved Amplified Polymorphic Sequences) markers were created for the mutant alleles *sbeIIb-2773-2, sbeIIb-5890-7, sbeIIb-3568-6*. The sequence around the SNP was obtained from *Phytozome* (Goodstein et al. 2012, version 11; Phytozome 11 2016, version 11). The software dCAPS Finder 2.0 (Neff et al. 2002, http://helix.wustl.edu/dcaps/) was used with 0 mismatches to find restriction enzymes cutting both forward and reverse strand and only the mutant or wildtype, not both. NCBI Primer-BLAST (Altschul et al. 1990; Camacho et al. 2009; NCBI Primer BLAST 2016) was used to create primers.

dCAPS (Derived Cleaved Amplified Polymorphic Sequences) were created for the mutant allele *sbeIIb-3194-3*. The sequence around the SNP was obtained from *Phytozome* (Goodstein et al. 2012, version 11, Phytozome 11 2016, version 11). The dCAPS Finder 2.0 (Neff et al. 2002, http://helix.wustl.edu/dcaps/) was used to create dCAPS with maximum of 2 mismatches. Restriction enzymes and primers were determined using this software.

TetraPrimer for ARMS-PCR were created for the mutant alleles *ssIIa-4308-2, sbeIIb-2703-1, sbeIIb-3403-5, sbeIIb-3218-4*. The sequence around the SNP was obtained from the reference genome used. The software PRIMER1: primer design for tetra-primer ARMS-PCR (http://primer1.soton.ac.uk/primer1.html) was used to create tetra-primer (Collins and Ke 2012; Ye et al. 2001).

Primers used for PCR amplification are described in **Table 3.1**. The PCR reaction was adjusted to primer specific annealing temperatures for each mutant. The ratio and amount of primers used were adjusted for tetra-primer types. PVP 20% and BSA were added to improve PCR amplification from fast and dirty DNA extraction (Xin et al. 2003).

Table 3.1. PCR conditions, primer sequences and restriction enzymes used for the genotyping of sorghum EMS mutants

PCR primer ID/ Type	Gene/ SNP-position /allele	Primer sequence	Temp. (°C)	Enzyme cut	Product size
2703-3_ S12_51294121 Tetra primer	*Sobic.004G163700* 51294121 *sbeIIb-2703-1*	InF: TTGCATTGCCTGATCAAACTCGTA; InR: CGGACTATCTTAGGTATCGTGGTAGGC; OuF: GACATTACAAGAAGAATCCCCACCAA; OuR: GTTTGGTCAATAATTGATCATTGTCGG	65.2	none	Product size T allele: 110 C allele: 165 outer primers: 224
2773_ 04_SNP_51302688 CAPS	*Sobic.004G163700* 51302688 *sbeIIb-2773-2*	F: CACGGTAAAGAGTACCTGCGA; R: TGAGCATGAAGGAGGCTTGG	57.1	ApoI HF cuts Mutant (M)	~ 78 (cut ~ 28, ~50bp)
3194_ 04_SNP_51298115 dCAPS	*Sobic.004G163700* 51298115 *sbeIIb-3194-3*	F: ATCATGGAGGTCACACCATCAATGC; R: TCCAATGCTAGATGGTGGCTTGAG	59.1	HpyCH4V cuts M	~71 (cut ~ 26, ~45)
3218 S_TET Tetra primer	*Sobic.004G163700* 51295327 *sbeIIb-3218-4*	InF: GTGTGTGCACAATATCACCCATCGTT InR: CAGGCAAAGTGATGAAGCTGGG OuF: ACCTTGTCCATCAACCAAAATGCA OuR: CGTGTGCGTTCACTTTGAGCTATG	61.0	none	Product size G allele: 150 A allele: 100 outer primers: 202
3403(1) Tetra primer	*Sobic.004G163700* 51295906 *sbeIIb-3403-5*	InF: AGTTCAATCCATTTGTCAGCCACATCT InR: GTAGGTTTTGACTATCGGATGCACCTG OuF: CAGCCCATGCAATTAAACATTAGTGTATG OuR: CCTCCACATCATTGGCTTACATAAACC	60.0	none	Product size A allele: 201 G allele: 182 outer primers: 329
3568_ 04_SNP_51292780 CAPS	*Sobic.004G163700* 51292780 *sbeIIb-3568-6*	F: ACATCTGTGTACCAAAGGCGAT R: CAGGATCCATCACGCAGCA	57.2	HpyCH4III cuts WT	~79 (cut ~ 37, ~42)
4308_ S13_8302675 Tetra primer	*Sobic.010G093400* 8302675 *ssIIa-4308-2*	InF: TTGTTCTGCAAGGTTGCTGGTA InR: TGACAGTATAGTTCAGGGGGATACATC OuF: TGATGCACCTCTCTTCCGG; OuR: TTCGCTAGTGCAAAAGTTGATCC	61.6	none	Product size A allele: 181 G allele: 224 two outer primers: 356
5890_ 04_SNP_51301340 CAPS	*Sobic.004G163700* 51301340 *sbeIIb-5890-7*	F: TGCTAGCCCAAAGTAGGAACAA R: CAGGCTCCAGGAGAAATACCA	56.4	MboI cut WT 1x, M 2x	~121 (cut: W:~78, 43; M: ~67,11,43)

Each PCR product from CAPS and dCAPS was digested with enzymes from New England Biolabs with digests were carried out under recommended conditions. The alleles *sbeIIb-3568-6* (3568_04_SNP_51292780) was digested with *Hpy*CH4III, *sbeIIb-2773-2* (2773_04_SNP_51302688) with *Apo*I-HF, *sbeIIb-5890-7* (5890_04_SNP_51301340) with *Mbo*I, *sbeIIb-3194-3* (3194_04_SNP_51298115) with *Cvi*RI=*Hpy*CH4V.

All samples from enzyme digest or tetra-primers were examined in 2.5-3% TAE high-resolution agarose gels, depending on the fragment size.

3.6.12 Flour preparation

The seed samples from PR16/17, WL17 and WL18 described above were milled into a fine flour using a ball mill (Retsch, Haan, Germany). Flour preparation was conducted as reported in Griebel et al. (2019).

3.6.13 Moisture determination

Moisture was determined using the standard oven method (Bradley 2014). Aluminum pans were oven dried before use. For moisture determination, 1g of flour sample was weight into pre-weight aluminum pans and dried over night at 105°C. In rare cases, flour was limited and less than 1g was sampled. The samples were placed in a desiccator after drying and before the final weight was recorded. The weights before and after drying were recorded and moisture was calculated based on water loss.

3.6.14 Starch extraction

Starch was extracted as described before (Griebel et al. 2019; Benmoussa and Hamaker 2011), from flour samples of double mutants and parents.

3.6.15 Starch thermal properties

The flour samples of double mutants and parents were analysed in duplicate for their starch thermal properties using Differential Scanning Calorimetry (DSC) as described before (Griebel et al. 2019).

3.6.16 Past viscosity profiles

Paste viscosity analysis was performed on whole grain flour samples using the Rapid Visco Analyzer (RTE-100), model 4 (serial number 970835) from Newport Scientific, Australia using the manuals' RVA standard method 1 and the results were analysed using standard analysis 1. A water-flour-dispersions of 11.86% was prepared (Bouvier and Campanella 2014). For the SbEMS mutants, two biological replicates in each season WL17 and WL18 were analysed. Each biological replicate was sampled in duplicate. The double mutants from WL18 did not have enough flour sample, so in some cases the biological replicate was not analysed in duplicate.

3.6.17 Amylose/Amylopectin analysis

Amylose and amylopectin analyses were performed on whole grain flour samples following the Megazyme K-AMYL kit manual. Each flour sample from PR16/17, WL17 and WL18 were analysed in duplicate. The spectrophotometer (Spectronic Genesys 10 Bio, Thermo Electron Corporation, Model: 970S0008) was used at 510nm.

3.6.18 Statistical analyses

Part of the data analysis for this paper was generated using SAS software, Version 9.4 of the SAS System for Windows. Copyright © 2002-2012. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

The GLIMMIX procedure was used to perform a linear mixed model analysis of variance (ANOVA) with the Tukey-Kramer multiple comparison adjustment to compare GT, amylose and viscosity properties across genotypes and environments.

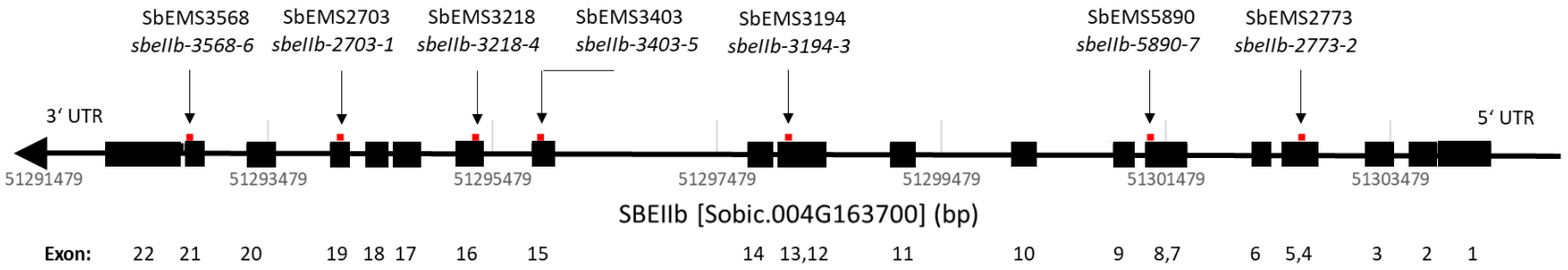
3.7 Results

3.7.1 Whole genome re-sequencing of ASV mutants reveals SNPs in *SBEII* and *SSII*

Whole genome re-sequencing studies characterizing the sorghum mutants described by Griebel et al. (2019) revealed homozygous SNPs in two genes associated with starch biosynthesis (**Fig. 3.1**). SbEMS 2703, SbEMS 2773, SbEMS 3194, SbEMS 3218, SbEMS 3403, SbEMS 3568, SbEMS 5890 exhibited homozygous SNPs in *Sobic.004G163700,* similar to a 1-4-alpha glucan starch branching enzyme II (Phytozome 11 2016, version 11; Goodstein et al. 2012, version 11; Campbell et al. 2016), producing mutant alleles *sbeIIb-2703-1, sbeIIb-2773-2, sbeIIb-3194-3, sbeIIb-3218-4, sbeIIb-3403-5, sbeIIb-3568-6 and sbeIIb-5890-7*. SbEMS 3920 and SbEMS 4308 exhibited homozygous SNPs in *Sobic.010G093400,* similar to a starch synthase zSTSII-1 (Campbell et al. 2016), resulting in mutant alleles *ssIIa-3920-1, ssIIa-4308-2*.

The snpEff tool (Cingolani et al. 2012) was used to predict the effects of each SNP on *Sobic.004G163700* and *Sobic.010G093400* influencing transcription and translation (**Fig. 3.1, B and D**). In *Sobic.004G163700*, alleles *sbeIIb-3568-6* and *sbeIIb-5890-7* are predicted to alter splice site donor sites and *sbeIIb-2703-1* and *sbeIIb-3218-4* created premature stop codons. Moderate effects on the proteins are predicted as missense variants resulting in amino acid changes in *sbeIIb-2773-2*, *sbeIIb-3403-5*, and *sbeIIb-3194-3*. In *Sobic.010G093400*, moderate SNP effects were predicted for the alleles *ssIIa-3920-1* as a missense variant and for *ssIIa-4308-2* as a missense and splice region variant.

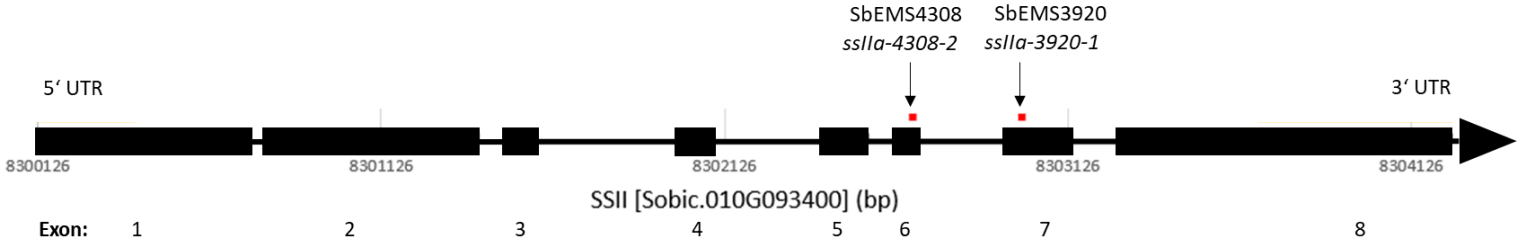
A



B

SbEMS	Gene	Allele	#CHROM	Effect	Impact	Codon Change	Amino Acid Change
2703	*Sobic.004G163700*	*sbeIIb-2703-1*	Chr04	STOP_GAINED	HIGH	c.2053C>T	p.Q685*
2773	*Sobic.004G163700*	*sbeIIb-2773-2*	Chr04	MISSENSE_VARIANT	MODERATE	c.512G>A	p.G171E
3194	*Sobic.004G163700*	*sbeIIb-3194-3*	Chr04	MISSENSE_VARIANT	MODERATE	c.1345C>T	p.R449C
3218	*Sobic.004G163700*	*sbeIIb-3218-4*	Chr04	STOP_GAINED	HIGH	tgG/tgA	p.W548*
3403	*Sobic.004G163700*	*sbeIIb-3403-5*	Chr04	MISSENSE_VARIANT	MODERATE	c.1593G>A	p.M531I
3568	*Sobic.004G163700*	*sbeIIb-3568-6*	Chr04	SPLICE_DONOR & INTRON_VARIANT	HIGH	c.2328+1G>A	
5890	*Sobic.004G163700*	*sbeIIb-5890-7*	Chr04	SPLICE_DONOR & INTRON_VARIANT	HIGH	c.825+1G>A	

C



D

SbEMS	Gene	Allele	#CHROM	Effect	Impact	Codon Change	Amino Acid Change
3920	*Sobic.010G093400*	*ssIIa-3920-1*	Chr10	MISSENSE_VARIANT	MODERATE	c.1222G>A	p.D408N
4308	*Sobic.010G093400*	*ssIIa-4308-2*	Chr10	MISSENSE &SPLICE_REGION_VARIANT	MODERATE	c.1180G>A	p.E394K

Fig. 3.1. Mutations in *SBEIIb* and *SSIIa* produce changes in alkali spreading value – (a) Mutants and positions of SNPs in *sbeIIb* (*Sobic.004G163700*), **(b)** *SBEIIb* mutations, effects, and impacts predicted by snpEff on transcription and translation, **(c)** mutants and positions of SNPs in *ssIIa* (*Sobic.010G093400*), and **(d)** *ssIIa* mutations, effects, and impacts predicted by snpEff on transcription and translation.

3.7.2 Comparative genomics identifies *SSIIa* and *SBEIIb*

Multiple sequence alignment and blastp (Altschul et al. 1990; Camacho et al. 2009) showed that *Sobic.004G163700* is likely the previously reported sorghum *SBEIIb* (AY304540 (Mutisya et al. 2003; Figueiredo et al. 2008) as both show 98.9% sequence similarity (Goodstein et al. 2012, version 12). Comparisons of the protein sequence and a phylogram (Chojnacki et al. 2017) of *Sobic.004G163700* showed that the new alleles are predicted to affect regions of the protein highly conserved across species like *Zea mays, Oryza sativa, Hordeum vulgare, Triticum aestivum, Lens culinaris, Pisum sativum, Phaseolus vulgaris* and *Colocasia esculenta* (**Fig. 3.2**).

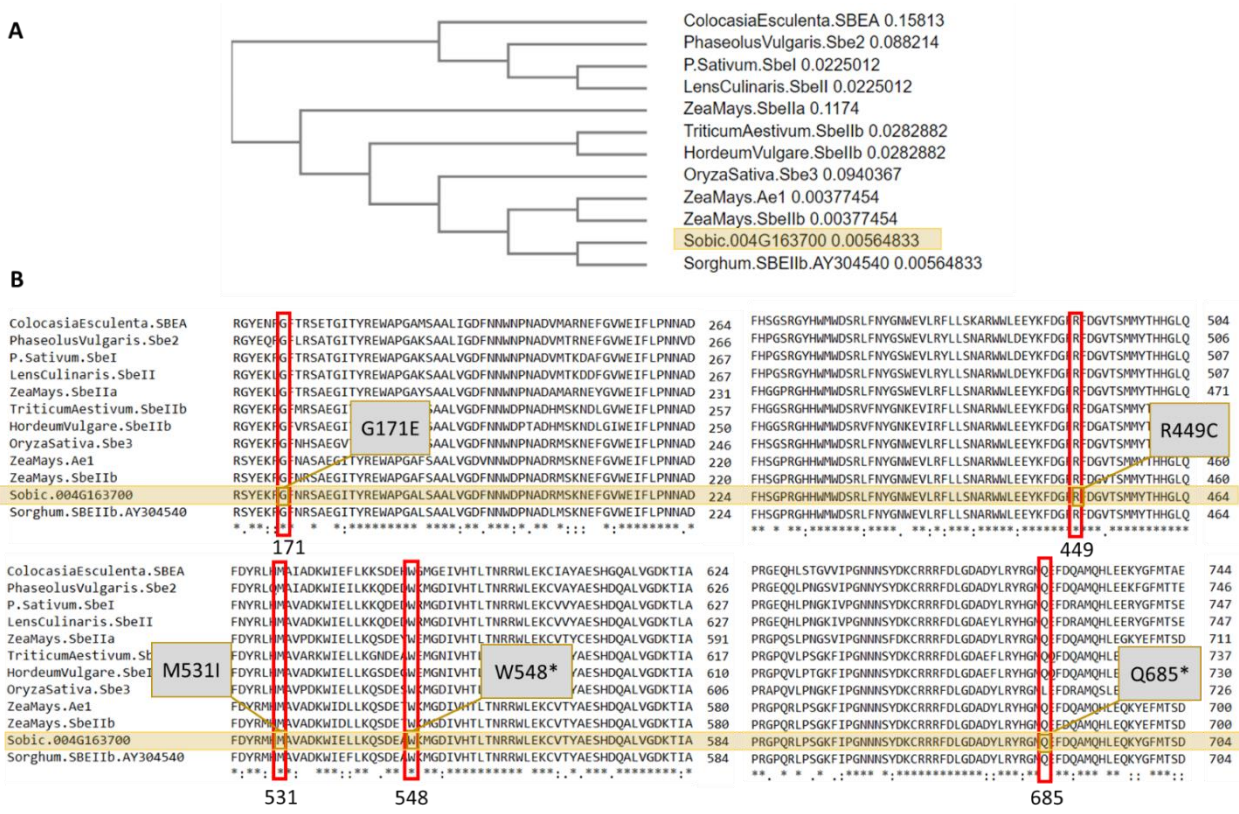


Fig. 3.2. Multiple sequence alignment of protein sequences from different sources with sorghum starch branching enzyme: (**A**) Phylogram using UPGMA and distance correction (**B**) SNPs in protein regions of *Sobic.004G163700* compared to other species.

The similarity of the SBEIIb in sorghum, followed by maize, rice (sbe3), barley, and wheat suggests that the identified gene is a starch branching enzyme of the isoform SBEIIb. This suggested that *Sobic.004G163700* encodes a *SBEIIb*.

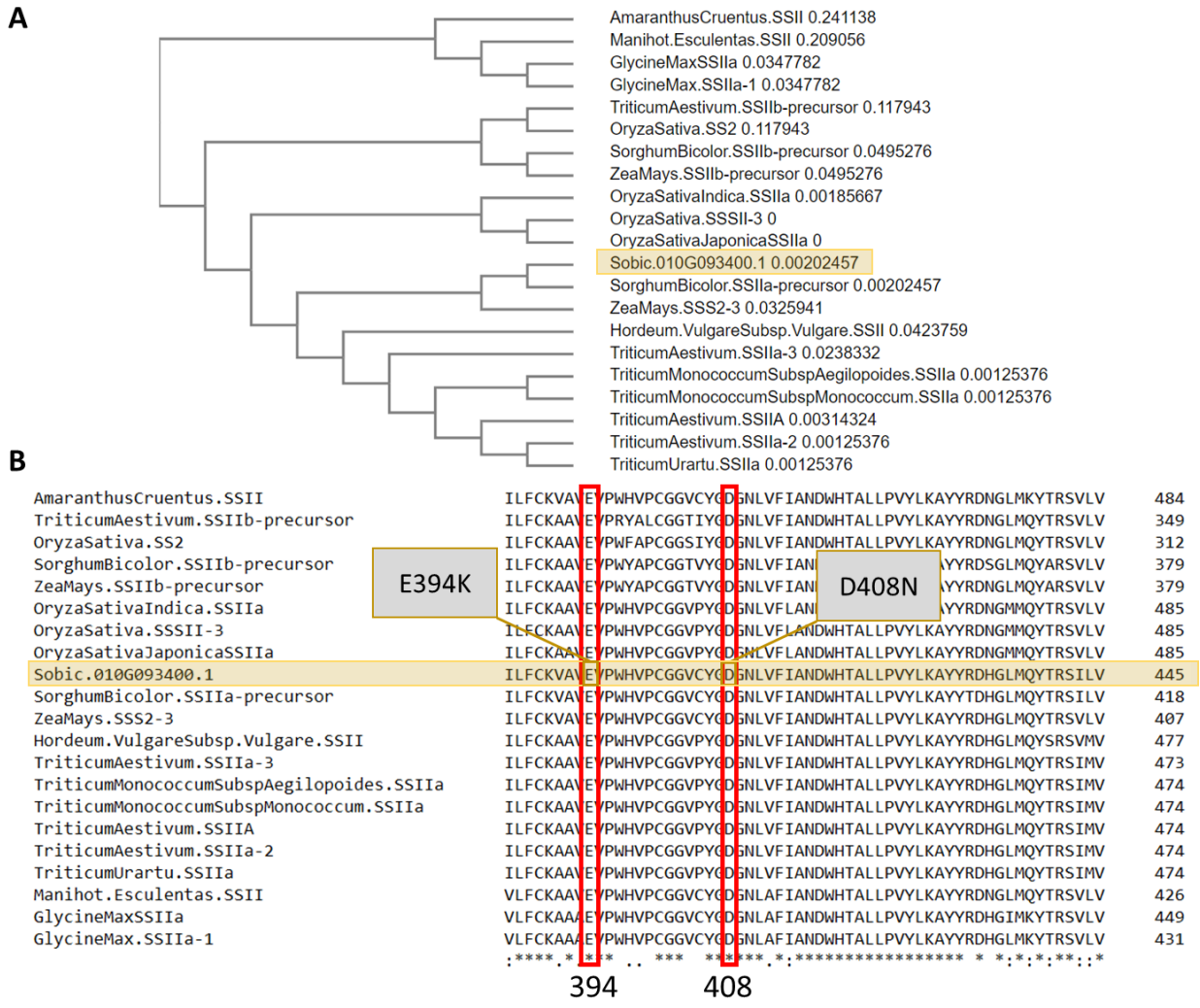


Fig. 3.3. Multiple sequence alignment of protein sequences from different sources with sorghum starch synthase enzyme: (**A**) Phylogram using UPGMA and distance correction (**B**) SNPs in protein regions of *Sobic.010G093400* compared to other species.

Multiple sequence comparison of protein sequences (Chojnacki et al. 2017) similar to that encoded by *Sobic.010G093400* showed that the SNPs of *ssIIa-3920-1 and ssIIa-4308-2* are in regions highly conserved across species of *Glycine max, Triticum aestivum, T. monococcum* and

T. urartu, Hordeum vulgare, Zea mays, Oryza sativa and *Amaranthus cruentus* (**Fig. 3.3**). The phylogenetic tree showed the sorghum *Sobic.010G93400* is most closely related to a sorghum SSIIa followed by SSS2-3 from maize, and several SSIIa from other genotypes of the *Triticum* species. These analyses provided evidence that *Sobic.010G93400* encodes a *SSIIa*.

3.7.3 ASV phenotype co-segregates with SNPs in *SSIIa* and *SBEIIb*

Eight of the SbEMS mutants were crossed to BTx623 and evaluated in $F_{2:3}$ populations for their ASV phenotype and associated SNP genotype (**Fig. 3.4**). Phenotypic evaluations and chi-square tests provided evidence that the mutant alleles *sbeIIb-2773-2*, *sbeIIb-3218-4*, *sbeIIb-3403-5* and *ssIIa-4308-2* segregated for ASV in a 3:1 ratio (**Fig. 3.4, A**). The 3:1 ratio did not fit (chi-square *p*-value) for the alleles *sbeIIb-2703-1, sbeIIb-3194-3, sbeIIb-3568-6* and *sbeIIb-5890-7*. Genotyping studies showed that the *sbeIIb-2703-1, sbeIIb-2773-2, sbeIIb-3218-4,* and *ssIIa-4308-2* alleles segregated in a 1:2:1 genotypic ratio but *sbeIIb-3194-3*, *sbeIIb-3403-5*, *sbeIIb-3568-6*, and *sbeIIb-5890-7* alleles exhibited distorted ratios (**Fig. 3.4, B**). This may be due to low viability of the homozygous mutants.

Linkage analyses demonstrated perfect co-segregation between the putative causal SNP markers and the ASV mutant phenotype in bi-parental populations of Tx623 with SbEMS 2703, SbEMS 3218, SbEMS 3403, SbEMS 3568, SbEMS 4308, and SbEMS 5890 (**Fig. 3.4**). Genetic analyses in populations developed using SbEMS 2773 and SbEMS 3194 also exhibited strong linkage between the SNP markers and the ASV phenotype but three progenies for the variant *sbeIIb-2773-2* and two progenies for *sbeIIb-3194-3* exhibited a heterogenous ASV phenotype but wild type genotype. It is possible to conclude that each mutant's SNP is either in the causal gene or linked to a gene causing ASV that is in a larger chromosome block. The co-segregation of seven *sbeIIb* alleles in *Sobic.004G163700* with the ASV phenotype provides strong evidence that this

gene is one of the genes controlling ASV in sorghum. Similarly, *ssIIa* also co-segregated with the ASV phenotype in one population supporting the conclusion that *Sobic.010G093400* also controls ASV.

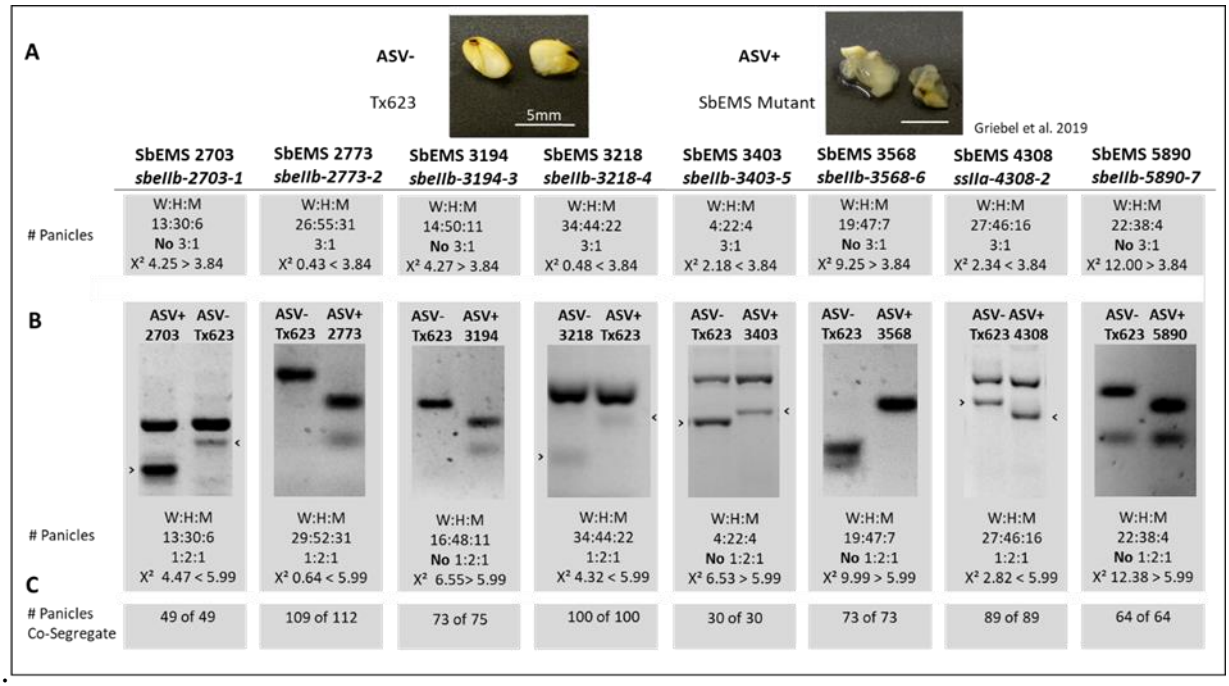


Fig. 3.4. Co-segregation analyses – (**A**) contrasting ASV phenotypes (1.8% KOH, 24h), ASV segregation ratios for each SbEMS mutant, and statistical tests for segregation of ASV, (**B**) SNP markers for each SbEMS mutant, genotype segregation ratios, and statistical tests for SNP segregation, and (**C**) co-segregation of SNP markers and ASV (N=# of panicles, tested with 32 seeds each) (Seed pictures from Griebel et al. 2019).

3.7.4 SSIIa and SBEIIb regulate distinct lower and higher starch GT

The mutants SbEMS 3920 and SbEMS 4308 showed significantly lower starch GT compared to the controls and the other mutants (Griebel et al. 2019). In the current study, we identified that these genotypes have mutations in *Sobic.010G093400* as represented by *ssIIa-3920-1* and *ssIIa-4308-2*. SbEMS 2703, SbEMS 2773, SbEMS 3194, SbEMS 3218, SbEMS 3403, SbEMS 3568, SbEMS 5890 exhibit the highest GT approximately 5°C higher T_p than the control samples (Griebel et al. 2019). These genotypes were identified in the current study as having

mutations in *Sobic.004G163700* as represented by the alleles *sbeIIb-2703-1*, *sbeIIb-2773-2*, *sbeIIb-3194-3, sbeIIb-3218-4, sbeIIb-3403-5, sbeIIb-3568-6* and *sbeIIb-5890-7.*

3.7.5 Sorghum sbeIIb, behaves similar to *amylose extender* and produces increased amylose values

Comparative studies showed that *Sobic.004G163700* is similar to SBEIIb (Mutisya et al. 2003) in sorghum and closely related to *Zea mays* amylose extender gene *Ae1*; hence, we hypothesized that mutants with SNP mutations in this gene could have increased amylose values in comparison to controls and *SSIIa* variants (**Table 3.2**). There was no significant genotype by environment interaction observed across three seasons (Puerto Rico 2016/2017, West Lafayette 2017 and 2018), so amylose values are presented for each environment and in a combined analysis. Macia was not evaluated in Puerto Rico 2016/17 but exhibited amylose values similar to the other controls in the other environments.

Table 3.2. Amylose values for *Sorghum bicolor* EMS mutants and controls from Puerto Rico 2016/2017 (PR16/17), West Lafayette 2017 (WL17) and West Lafayette 2018 (WL18).

		PR 16/17	WL 2017	WL 2018	Combined	PR 16/17	WL 2017	WL 2018
Pedigree	**Allele**	**Amylose % - LSMeans**				**Rank**		
EMS2703	*sbeIIb-2703-1*	43.54 c	43.60 bc	38.14 bc	41.76 d	2	2	3
EMS2773	*sbeIIb-2773-2*	38.16 c	37.89 bc	35.79 bc	37.28 c	5	6	6
EMS3194	*sbeIIb-3194-3*	44.02 c	48.15 c	45.10 c	45.76 e	1	1	1
EMS3218	*sbeIIb-3218-4*	41.40 c	40.54 bc	38.18 bc	40.04 cd	3	3	2
EMS3403	*sbeIIb-3403-5*	36.70 bc	34.63 ab	29.26 ab	33.53 b	6	7	7
EMS3568	*sbeIIb-3568-6*	36.41 bc	38.45 bc	37.94 bc	37.6 c	7	5	4
EMS5890	*sbeIIb-5890-7*	39.88 c	40.15 bc	36.42 bc	38.82 cd	4	4	5
EMS3920	*ssIIa-3920-1*	26.20 a	24.20 a	24.76 a	25.06 a	10	12	8
EMS4308	*ssIIa-4308-2*	28.36 ab	25.41 a	23.27 a	25.68 a	8	10	11
Macia	control	NA	27.15 a	23.28 a	NA		8	10
Sepon82	control	24.63 a	26.05 a	22.71 a	24.46 a	11	9	12
Tx623	control	26.90 ab	24.47 a	23.94 a	25.11 a	9	11	9

Macia not evaluated in season Puerto Rico 16/17, so not part of combined analysis. Values followed by the same letter (A-E) in the same column are not significantly different.

The *sbeIIb*-mutants *sbeIIb-2703-1, sbeIIb-2773-2, sbeIIb-3194-3, sbeIIb-3218-4, sbeIIb-3403-5, sbeIIb-3568-6* and *sbeIIb-5890-7* produced significantly higher amylose values (>30% amylose) compared to the *ssIIa*-mutants *ssIIa-3920-1* and *ssIIa-4308-2* and the controls. In all three seasons, SbEMS3194 had the highest amylose values. The SbEMS 2703, SbEMS 3218, and SbEMS 5890 were ranked 2nd to 4th in two environments, respectively. These data support that *Sobic.004G163700* is similar in sequence and in function to *Ae1* in maize. Mutations of *SBEIIb* could be used to produce high-amylose sorghum varieties.

The *SSIIa* mutations represented by *ssIIa-3920-1* and *ssIIa-4308-2* produced similar amylose values and did not show higher amylose values compared to the controls.

3.7.6 ASV and *sbeIIb* mutations alter amylose values in an allele dependent manner

Based on the higher amylose values of *sbeIIb-2703-1, sbeIIb-2773-2, sbeIIb-3194-3, sbeIIb-3218-4, sbeIIb-3403-5, sbeIIb-3568-6* and *sbeIIb-5890-7* allelic variants (**Table 3.2**) and sequence similarity to *ae* of maize, we propose that *Sobic.004G163700* could be a sorghum homolog of amylose extender. Based on this proposal, the amylose level in sorghum may depend on the number of alleles in the triploid endosperm. To test this hypothesis, each of the *sbeIIb*-mutants was sterilized in A1-cytoplasm and the A-lines and B-lines of mutants and Tx623 were used in crosses to produce seeds carrying zero, one, two and three mutant alleles in the triploid endosperm. Over the course of the study, the amylose values increased in a dosage dependent and mutant allele dependent manner for *sbeIIb*-mutants and did not increase in *ssIIa*-mutants. In most cases, amylose content increased by 1% to 2% with each additional dose of *sbeIIb* in the endosperm with values greater than 30% when three mutant *sbeIIb* alleles were present (**Table 3.3**). This shift in amylose content was not observed in samples with varying doses of *ssIIa* (SbEMS 4308). Analyses of these samples for ASV in 2018 showed an increase in percentage of seeds showing an ASV depending

on whether one, two or three mutant *sbeIIb* or *ssIIa* alleles were present in the endosperm. We did not observe any ASV swelling in genotypes carrying one dose of the allele.

Table 3.3. Amylose contents and ASV phenotypes of grain samples from an allele dosage test in West Lafayette (WL) 2017 and 2018.

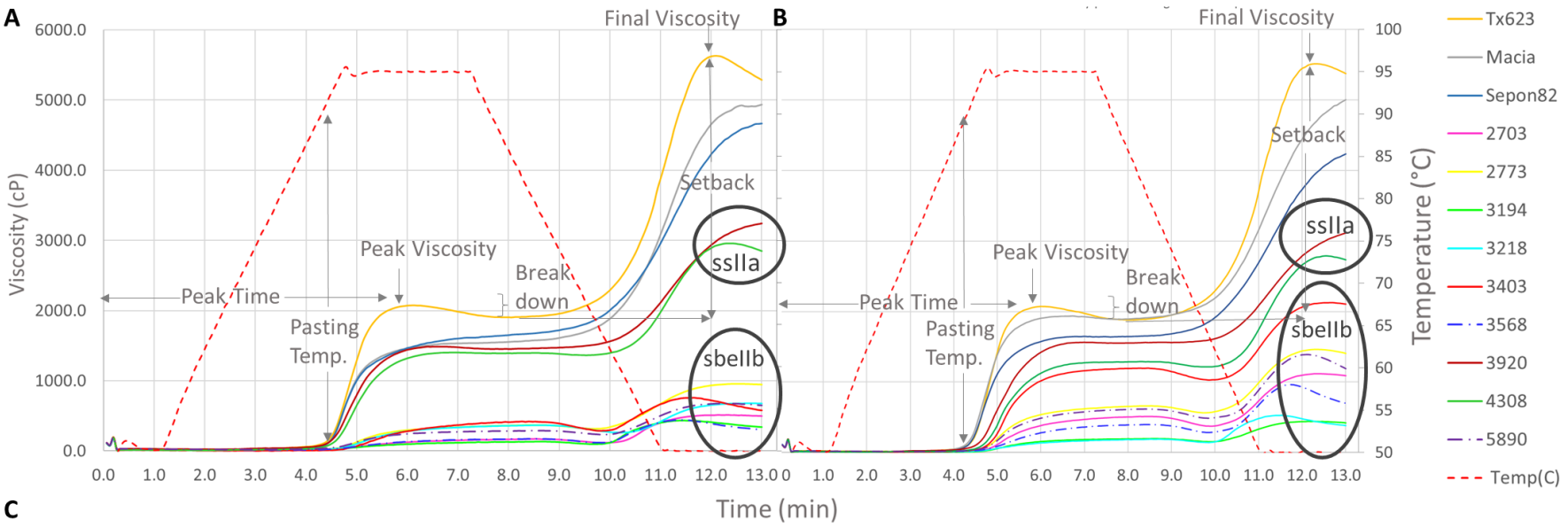
		WL2017				WL2018							
		Amylose %				Amylose %				% Seeds with ASV+			
Pedigree		0 Dose	1 Dose	2 Dose	3 Dose	0 Dose	1 Dose	2 Dose	3 Dose	0 Dose	1 Dose	2 Dose	3 Dose
ATx623/ BTx623	wild type	24.79				22.80 e				0			
EMS 2703	*sbeIIb-2703-1*		26.65	28.60[c]	43.10		23.46 e	26.11[c] cde	36.62 a		0[b]	34[c]	100[a]
EMS 2773	*sbeIIb-2773-2*		26.93	26.94[c]	38.35		24.84 de	25.04[c] cde	35.76 ab		0[b]	0[bc]	100[a]
EMS 3194	*sbeIIb-3194-3*		26.13	28.76[c]	47.29		24.72 de	26.34 cde	39.22 a		0[b]	8	96[a]
EMS 3218	*sbeIIb-3218-4*		26.75	28.48[c]	40.64		no seed	24.49 de	38.28 a		no seed	38	83[a]
EMS 3403	*sbeIIb-3403-5*		no seed	no seed	34.63		25.29 cde	21.33[c] e	29.26 cd		0[b]	6[c]	88[a]
EMS 3568	*sbeIIb-3568-6*		26.18	25.79[c]	38.45		25.57 cde	26.09[c] cde	31.06 cb		0	0[c]	100[a]
EMS 4308	*ssIIa-4308-2*		24.57	25.51[c]	25.47		24.04 e	23.28 e	25.28 cde		0[b]	8	100[a]
EMS 5890	*sbeIIb-5890-7*		26.04	29.09[c]	39.64		24.71 de	26.02 cde	39.45 a		0[b]	38	100[a]

WL2018 Amylose: Values followed by the same letters horizontal and vertical are not significantly different. WL17: 2703 hand emasculated for 2 Doses. ASV tested on 32 seeds (resuts in %), [a] fewer than 32 seeds tested due to seed limitation; [b] no ASV swelling but leaching, very litle surface swelling. [c] one biological replicate, all crosses were made in duplicate or triplicate but some A lines (2 doses) did not set seeds.

3.7.7 The *sbeIIb* and *ssIIa* alleles control paste viscosity profiles and fall in distinct classes

Pasting viscosity is an important property of starch related to its gelatinization. The paste viscosity profiles are very distinct for *ssIIa* and *sbeIIb* mutants in both environments. There was a significant genotype by environment effect for the pasting viscosity profiles; therefore, data are presented by individual environments (**Fig. 3.5**). The *sbeIIb* mutants exhibited lower peak and

final viscosity compared to *ssIIa-4308-2, ssIIa-3920-1,* and controls (**Fig. 3.5 A, B**). The viscosity profiles of BTx623, Sepon82, Macia, *ssIIa-3920-1* and *ssIIa-4308-2* followed the same trend in both environments. The *sbeIIb* mutants slightly overlap but cluster together across environments. The SNP mutations in *Sobic.010G093400*, encoding SSIIa, resulted in reduced viscosity in comparison to the controls but higher viscosity in comparison to *sbeIIb*-mutants. BTx623 showed an earlier and larger breakdown than all other starches (**Fig. 3.5 A, B, C**). The mutant starches demonstrated minimal breakdown and later peak viscosity or a plateau from peak viscosity. The setback viscosity, formation of a gel as the long chain amylose molecules begin to realign, was approximately 10× lower in the *sbeIIb*-mutants compared to wild type starches.



C

Pedigree	Allele	WL2017 Peak 1 [cP]		WL2018 Peak 1 [cP]		WL2017 Trough 1 [cP]		WL2018 Trough 1 [cP]		WL2017 Breakdown [cP]		WL2018 Breakdown [cP]		WL2017 Final Viscosity [cP]		WL2018 Final Viscosity [cP]		WL2017 Setback [cP]		WL2018 Setback [cP]		WL2017 Peak Time [min]		WL2018 Peak Time [min]		WL2017 Pasting Temp. [°C]		WL2018 Pasting Temp. [°C]	
2703	*sbeIIb-2703-1*	155.0	A	471.0	A	119.5	A	372.5	A	35.5	A	98.5	ABC	497.75	A	1091	AB	378.3	A	718	AB	6.9	BC	7	B	NA		95	D
2773	*sbeIIb-2773-2*	338.5	A	620.3	A	299.8	A	546	AB	38.8	A	74.3	AB	943.75	A	1404	AB	644	A	857.5	ABC	7	C	7	B	95	C	95	D
3194	*sbeIIb-3194-3*	120.8	A	188.0	A	98.5	A	153.8	A	22.3	A	34.3	AB	340.5	A	421.3	A	242	A	267.5	A	6.9	BC	6.9	B	NA		95	D
3218	*sbeIIb-3218-4*	338.3	A	165.4	A	262.5	A	134.1	A	75.8	AB	31.6	A	680.5	A	393.3	A	418	A	259.1	A	6.9	BC	7	B	95	C	NA	
3403	*sbeIIb-3403-5*	384.5	A	1161.5	B	307.5	A	1017	BC	71.8	AB	144	BC	664.16	A	2107	BC	358.5	A	1089	BCD	7	BC	7	B	95	C	94	CD
3568	*sbeIIb-3568-6*	158.0	A	362.5	A	114.9	A	281.3	A	43.2	A	81.3	AB	301.5	A	700	A	186.6	A	418.8	AB	7	BC	7	B	NA		95	D
3920	*sslIa-3920-1*	1489.3	B	1565.0	BC	1445	BC	1446	CDE	44	A	119	ABC	3239.5	B	3111	CD	1794	B	1666	D	6.6	B	7	B	91	AB	92	B
4308	*sslIa-4308-2*	1405.5	B	1274.5	B	1335	B	1147	CD	70.5	A	128	ABC	2849.5	B	2733	C	1515	B	1586	CD	6.9	BC	7	B	91	B	93	BC
5890	*sbeIIb-5890-7*	225.0	A	575.5	A	187.3	A	484	A	37.8	A	91.5	AB	591	A	1190	AB	403.8	A	705.8	AB	6.9	BC	7	B	95	C	95	D
Macia	control	1535.5	B	1943.0	CD	1472	BC	1878	E	63.3	A	65	AB	4938.25	C	5003	EF	3466	C	3125	EF	6.9	BC	6.7	B	90	AB	88	A
Sepon82	control	1604.0	BC	1651.3	BCD	1479	BC	1587	DE	125	AB	64.3	AB	4667	C	4231	DE	3188	C	2644	E	7	BC	6.9	B	90	AB	89	A
Tx623	control	2073.3	C	2075.6	D	1893	C	1880	E	184	B	197	C	5325.27	C	5366	F	3445	C	3484	F	6.2	A	6.1	A	89	A	89	A

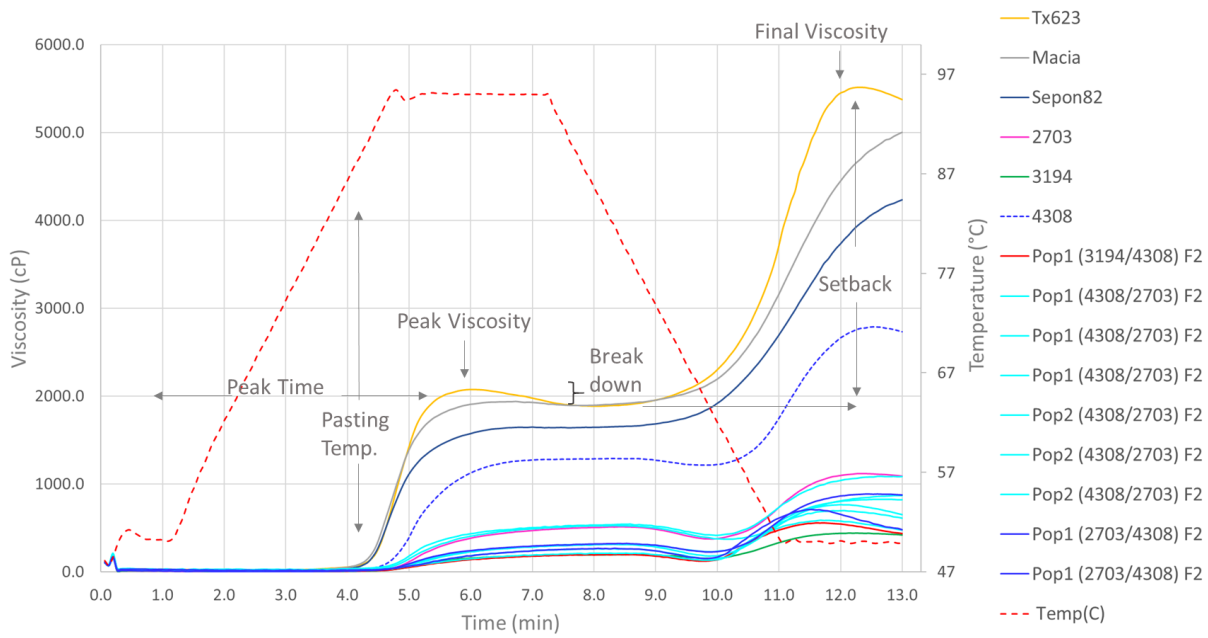
Values followed by the same letter (A–D) in the same column are not significantly different ($P<0.05$)

Fig. 3.5. RVA paste viscosity profiles of *Sorghum* EMS mutants and controls– lsmean values presented from two seasons (**A**) West Lafayette 2017 and (**B**) 2018 and (**C**) both seasons. N=2 biological replicates per genotype and three for BTx623 in WL2018.

3.7.8 Double Mutants of SSIIa and SBEIIb resemble SBEIIb parent in their phenotype

SbEMS mutant B-lines were hand emasculated and intercrossed to develop F_2 populations *sbeIIb-3194-3* × *ssIIa-4308-2* and *sbeIIb-2703-1* × *ssIIa-4308-2* (**Fig. 3.6**).

A



B

Pedigree	Allele	Reps	Amylose [%]	starch gelatinization characteristics					
				To [°C]	Tp [°C]	Tc [°C]	Tc-To [°C]	Enthalpy [J/g]	Melting Point [°C]
			lsmeans						
2703	*sbeIIb-2703-1*	2	37.55 AB	67.6 A	78.3 A	91.4 AB	23.7 A	4.3 A	99.6 AB
3194	*sbeIIb-3194-3*	2	44.24 A	67.3 A	80.0 A	97.5 A	30.1 A	2.5 A	100.0 A
4308	*ssIIa-4308-2*	2	23.34 B	55.7 B	63.3 B	76.7 B	21.0 A	4.3 A	97.7 B
Pop1 (2703/4308) F2	*sbeIIb-2703-1/ssIIa-4308-2*	2	38.82 AB	65.4 A	77.3 A	88.8 AB	23.3 A	2.9 A	100.2 A
Pop1 (3194/4308) F2	*sbeIIb-3194-3/ssIIa-4308-2*	1	37.65 AB	65.1 A	77.1 A	87.8 AB	22.7 A	2.4 A	99.7 AB
Pop1 (4308/2703) F2	*ssIIa-4308-2/sbeIIb-2703-1*	3	39.19 A	65.0 A	77.6 A	88.8 AB	23.8 A	3.0 A	99.4 AB
Pop2 (4308/2703) F2	*ssIIa-4308-2/sbeIIb-2703-1*	3	37.40 AB	64.2 A	76.1 A	88.6 AB	24.4 A	3.4 A	99.7 AB

Values followed by the same letter (A to B) in the same column are not significantly different (P<0.05).

Fig. 3.6. Double mutants of *sbeIIb* and *ssIIa* – (**A**) RVA paste viscosity profiles (**B**) starch thermal properties and amylose content. Pop = F2 Population, numbers (Pop1, Pop2) indicate different populations. (N="Reps"=replications listed in **B**).

Homozygous double mutants were identified in each population. Each double mutant has one parent with the *ssIIa-4308-2* allele and a second parent with either the *sbeIIb-2703-1* or *sbeIIb-3194-3* alleles. The lowest amylose values were measured in the parent carrying *ssIIa-4308-2* with 23% and higher amylose values in parent *sbeIIb-2703-1* with 38% and *sbeIIb-3194-3* with 44%.

The amylose values in the double mutants resembled the *sbeIIb*-mutant parent and ranged from 37% to 39%. The lowest starch GT values were measured in the parent carrying *ssIIa-4308-2* and the higher values in parent *sbeIIb-2703-1* and *sbeIIb-3194-3*. The starch thermal properties of the double mutants also resembled the *sbeIIb* mutant parent. The paste viscosity profiles varied depending on the mutant alleles involved and resulted in higher peak and final viscosity in *ssIIa-4308-2* than in *sbeIIb-2703-1* or *sbeIIb-3194-3*. The viscosity profiles for the double mutants were similar to their *sbeIIb*-mutant parents.

3.8 Discussion

Seven new alleles of a starch branching enzyme proposed as *SBEIIb* (*Sobic.004G163700*) and two new alleles of the putative *SSIIa (Sobic.010G093400)* were identified in sorghum in this study. The new alleles of *SBEIIb* represent SNP mutations spread across the entire gene and therefore provide a great resource for understanding the gene function. A previous study on sorghum *SBEIIb* (AY304540) focused only on a subset of exons evaluated in natural variation (Hill et al. 2012).

Comparative genomics suggests that *Sobic.004G163700* likely codes for a SBEIIb with high sequence similarity to the one reported previously (Mutisya et al. 2003) and is closely related to amylose extender (SBEIIb) in maize (**Fig. 3.2)**. The *Sobic.010G093400* product is most similar to SSII-3 in rice, SSIIa in sorghum and other crop species. A BLAST search of the NCBI protein sequence of maize amylose extender and sorghum *SBEIIb*, AY304540, against sorghum in *Phytozome* (Goodstein et al. 2012, version 12.1), exhibited the best hit for *Sobic.010G273800*, followed by *Sobic.003G213800* and *Sobic.004G163700*. All three genes are reported as similar to branching enzymes (Campbell et al. 2016), while *Sobic.003G213800* is listed as a putative 1-4-alpha glucan branching enzyme, *Sobic.010G273800* as similar to a starch branching enzyme I

precursor and and *Sobic.004G163700* is similar to a 1-4-alpha glucan branching enzyme 2 (Campbell et al. 2016). No SNPs were found in *Sobic.010G273800* or *Sobic.003G213800* in ASV mutants. The gene *Sobic.004G163700* appears to be a sorghum homolog of the maize gene *amylose extender1 (ae1)* encoding a SBEIIb.

The central catalytic A-domain of SBEIIb in maize and rice has four highly conserved regions important for the catalytic activity with key amino acid positions and a catalytic triad (Tetlow and Emes 2014). The mutant allele *sbeIIb-3194-3* results in the substitution R449C, which is the position reported in SBEIIb of sorghum (Mutisya et al. 2003) and the same residue in maize SBEIIb (Tetlow and Emes 2014; **Fig. 3.2**). The amino acid changes from alleles *sbeIIb-3403-5* (M531I) and *sbeIIb-3218-4* (W548*) might be also in the catalytic domain of SBEIIb, as similar amino acid positions are reported in maize to be in that region (Tetlow and Emes, 2014).

None of the nine mutants with new allelic variants of *SSIIa* and *SBEIIb* have a SNP mutation in the *GBSS* (*Sobic.002G116000*, *Sobic.010G022600*; Campbell et al. 2016), known to act on amylose (Nakamura 2015), we therefore conclude that our mutants predominantly affect amylopectin synthesis. The endosperm specific SBEIIb transfers shorter chains with a degree of polymerization (DP) of approximately 6-7 in corn, rice and wheat (Tetlow and Emes 2014; Nakamura 2015). Therefore, the *sbeIIb*-mutant alleles discovered in our study might lack the enzyme function or show reduced activity, resulting in amylopectins with shorter glucan chains DP less than 8 and a less branched molecule. An increase in short chains of DP 12-16 in amylopectin is associated with increased starch GT (Nakamura 2015b) and may explain the high starch GTs of our *sbeIIb* mutants. In monocots, SSIIa is involved in the elongation of short chains into intermediate size chains (DP 12-24) (Tetlow et al. 2004; Nakamura 2015b), explaining why *ssII* mutants in other species result in amylopectin with less intermediate chains and more short

chains (Nakamura 2015). An increase in amylopectin short chains of DP<12 and decrease of chains with DP 13 to 24 decrease starch GT (Nakamura 2015b), which could explain the low starch GT of our *ssIIa* mutants. The branch chain length structure needs evaluation as our mutants might behave like amylose extenders with longer internal branch chains of amylopectin (Jane 2009; Tetlow and Emes 2014; Nishi et al. 2001), less branched in outer chains (Jane 2009; Tetlow and Emes 2014) and have less short chains (Jane 2009; Tetlow and Emes 2014; Nishi et al. 2001).

Amylose extender mutant starches are known in maize and rice (Shannon et al. 2009; Li et al. 2008; Liu et al. 2012a; Nishi et al. 2001). We observed that different *sbeIIb* alleles result in varying amylose content with higher values overall in comparison to the *ssIIa* mutants, which is similar to *ae* maize and rice starches (Shannon et al. 2009; Li et al. 2008; Nishi et al. 2001). The amylose levels were >30% in our study while >50 % in maize (Li et al. 2008; Liu et al. 2012a). Different *ae* alleles in maize are reported to have different starch thermal properties (Shannon et al. 2009; Li et al. 2008; Liu et al. 2012a) such as observed in our study with *sbeIIb* mutants resulting in higher T_p and T_c and in a wider starch gelatinization range compared to the wild type (Griebel et al. 2019). A study of sorghum *SSIIa* and *SBEIIb* reported no distinct functions with both genes varying in starch GTs (Hill et al. 2012). The starch GT increased with increased amylose content (Hill et al. 2012). We report that *sbeIIb* mutations are recessive with an allele dosage dependent influence on the amylose content in the triploid endosperm, as reported in other studies for *ae* (Shannon et al. 2009; Nishi et al. 2001). We assume that the wild type allele is not completely dominant over the mutant allele *ae* as reported in another species (Shannon et al. 2009).

We show that the ASV in sorghum is not solely controlled by a SSII as reported in previous studies in sorghum and rice (Gao et al. 2011; McKneight 2015; Wang et al. 2007) but instead was controlled by two genes, *Sobic.004G163700 and Sobic.010G093400*. A comparative genomics

approach showed that sorghum SSIIa is related to rice SSSII-3. The *SSSII-3 (ALK*) gene is the major gene that regulates starch GT (starch GT measured by ASV) in rice (Tian et al. 2009). The proposed *ssIIa* alleles in sorghum lowers starch GT by approximately 10°C similar to maize (Zhang et al. 2004; Liu et al. 2012b) and sorghum (McKneight 2015), while a study of natural variation (Hill et al. 2012) showed no distinct reduction by gene in starch GT. The lower starch GT is reported to be related to amylopectin structure of *ssII* mutants (Preiss 2009). The *sbeIIb* allelic variants behave conversely and confirm earlier reports that higher amylose contents are observed together with an increase in starch GT (starch GT measured by ASV) (Tian et al. 2009). In transgenic rice, the position of mutations in the *ALK* gene influenced whether ASV scores were correlated with starch GT (Gao et al. 2011). The expression of ASV in sorghum is similar but distinct for a *ssIIa* or *sbeIIb* allelic variant. We did not find a consistent relationship between ASV with onset or peak starch GT (Griebel et al. 2019); however, this may be due to fact that the ASV phenotype was caused by SNP mutations in two different genes, *Sobic.004G163700* and *Sobic.010G093400*, having different effects on starch functional properties.

The amylose levels for the controls were similar to previous reports (Beta et al. 2000; Waniska and Rooney 2000; Sang et al. 2008). In rice, the gene *Wx* controls amylose content and gel consistency (GC), both traits being negatively correlated (Tian et al. 2009). We observed that high amylose sorghum mutants also result in a low gel consistency but have mutations in *SBEIIb*. In rice, amylose content is also regulated by minor genes such as the *ALK (SSII-3)* gene, *SSIII-2*, *SSI*, *PUL* (Tian et al. 2009). Minor genes controlling GC in rice are *ISA*, *SBE3* (Tian et al. 2009) and the *ALK (SSII-3)* gene (Gao et al. 2011; Tian et al. 2009). The *ssII*-mutants showed higher amylose values in other cereals (Zhang et al. 2004; Liu et al. 2012b; Luo et al. 2015), and nearly no change in amylose content in rice (Luo et al. 2015). In this study, no change in amylose content

was observed for *ssIIa-3920-1* and *ssIIa-4308-2* in sorghum, similar to reports in rice (Luo et al. 2015). A study (Hill et al. 2012) reported lower and higher amylose contents for *ssIIa* allelic variants in the standing variation of sorghum.

The ASV mutants fall into distinct classes regarding functional properties and effects on starch thermal characteristics, amylose levels and pasting behavior. The *sbeIIb* genetic variants exhibited higher T_p and T_c, a wider range of starch GT (Griebel et al. 2019), higher amylose levels, and lower peak and final viscosity compared to the controls. Peak viscosity was positively correlated with gel consistency and negatively correlated with starch GT as previously shown in rice (Gao et al. 2011). Low amylose content together with high gel consistency, and low starch GT (Griebel et al. 2019), were observed in the sorghum *ssIIa* mutants carrying the alleles *ssIIa-3920-1* and *ssIIa-4308-2,* which are desired cooking qualities in rice (Wang et al. 2007). The *ssIIa* mutants exhibit a lower peak and final viscosity than the controls but higher peak and final viscosity than the high amylose *sbeIIb* allelic variants. Hill et al. (2012) reported an increase in amylose content by genetic variants of *ssIIa* and *sbeIIb* resulting in higher starch GT, higher final viscosity, high setback, low peak viscosity and low breakdown. In the current study, a clear distinction was observed between *ssIIa* and *sbeIIb* genes on amylose content, starch GT, and pasting behaviour with associated changes in lower final viscosity, small breakdown, smaller setback, lower peak viscosity. The different results might be due to the differences between allelic variation represented in the standing variation of sorghum or derived by chemical mutagenesis.

The peak viscosity is related to the maximum swelling capacity of starch granules (BeMiller and Huber 2008; Bouvier and Campanella 2014), and is mostly determined by the amylopectin content of the starch (Bouvier and Campanella 2014). This property is closely associated to the thickening power of a starch (Biliaderis 2009). The swelling capacity and

thickening power of the high amylose *sbeIIb* mutants in the present study is lower than in controls and *ssIIa* mutants, which implies that the *sbeIIb* mutants are less suitable to be used as thickening agents. The breakdown is reported to depend on the starch structure, with high amylopectin starches breaking down faster (Biliaderis 2009), as observed in the results for BTx623 but not for *ssIIa-3920-1* and *ssIIa-4308-2*. During cooling the controls and *ssIIa-4308-2* and *ssIIa-3920-1* formed stronger gels, while the *sbeIIb*-variants were unstable liquid dispersions and did not form a gel, which is likely related to the extent of molecule retrogradation and re-association influencing gel formation (BeMiller and Huber 2008; Biliaderis 2009; Bouvier and Campanella 2014). Higher amylopectin starches such as those produced by the *ssIIa*-mutants and the controls, resulted in better gel stability, gel consistency and gel volume development than *sbeIIb*-mutants, as amylopectin and amylose amylopectin interactions are the important drivers to control those characteristics (Biliaderis 2009).

Double mutants of amylose extender and *sugary2* (*su2*) have been reported with variation of amylose values and starch GT most similar to their *ae* parents in other crops (Shannon et al. 2009). The sorghum double mutants described in this study also resemble the *sbeIIb-2703-1* and *sbeIIb-3194-3* parents for amylose content, starch thermal properties and starch viscosity profiles. It is not clear if the *sbeIIb* influences *ssIIa* or is dominant over it. The *ALK (SSII*-3) gene in rice did not contribute strongly to changes in paste viscosity profiles as observed for the *wx* locus (Wang et al. 2007). Variants of the *wx* locus were not characterized in our study of sorghum mutants; however, it was observed that the allele *ssIIa-4308-3* did not contribute to the paste viscosity profiles of double mutants.

The branch chain length and branching frequency influences starch digestibility (Tetlow and Emes 2014). Therefore, it would be of great importance to further evaluate the newly identified

alleles for their branch chain length distribution and branching frequency. Genotypes with an *ae* mutation and high levels of amylose are reported as being correlated with resistant starch content (Jane 2009; Tetlow and Emes 2014; Li et al. 2008) and are less enzyme-digestible (Tetlow and Emes 2014; Li et al. 2008). It may be valuable to determine whether the *sbeIIb* mutants produce more resistant or slowly digested starches.

Nine novel alleles of a *SBEIIb* and *SSIIa* were identified and characterized in a *sorghum bicolor* EMS population. These nine alleles exhibit an ASV phenotype and varying starch functional properties. The ASV phenotype was controlled by the proposed *SSIIa* and *SBEIIb* with very distinct functionalities. The *sbeIIb* variants resulted in lower paste viscosity profiles, lower gel consistency, lower thickening power, higher amylose levels, and higher starch GT (especially Tc) than wild type samples or ssIIa mutants. The *ssIIa* mutant alleles resulted in a drop of final viscosity, less gel consistency, lower thickening power, same amylose values, and lower starch GT compared to wild-type samples from Tx623. These variants will provide opportunities to produce sorghum varieties with modified starch end-use qualities likely important for the beer brewing and baking industries and specialty foods for humans with diabetes.

3.9 References

Addo-Quaye C, Buescher E, Best N, Chaikam V, Baxter I, Dilkes BP (2017) Forward genetics by sequencing EMS variation-induced inbred lines. G3 (Bethesda) 7:413-425

Addo-Quaye C, Tuinstra M, Carraro N, Weil C, Dilkes BP (2018) Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. G3 (Bethesda) 8: 1079-1094

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–10

BeMiller J (2014) Carbohydrate analysis. Food Analysis. In: Nielsen S (ed), 4th edn. Springer, New York, pp 147-180

BeMiller JN, Huber KC (2008) Carbohydrates. In: Damodaran S, Parkin K, Fennema O (eds) Fennema's Food Chemistry, 4th edn. CRC press, Florida, pp 83-154

Benmoussa M, Hamaker BR (2011) Rapid small-scale starch isolation using a combination of ultrasonic sonication and sucrose density separation. Starch/Stärke 63:333-339

Beta T, Corke H, Rooney LW, Taylor JRN (2000). Starch properties as affected by sorghum grain chemistry. J Sci Food Agric 81:245-251

Bhattacharya KR (1979) Gelatinization temperature of rice starch and its determination. In: International Rice Research Institute, Proceedings of the workshop on chemical aspects of rice grain quality. International Rice Research Institute, Philippines, pp 231-249

Bhattacharya KR, Sowbhagya CM (1972) An improved alkali reaction test for rice quality. J Food Sci Technol 7:323-331

Biliaderis CG (2009) Structural transitions and related physical properties of starch. In: BeMiller J, Whistler R (eds) Starch Chemistry and Technology, 3rd edn. Academic Press, Amsterdam, pp 293 - 372

Bouvier JM, Campanella OH (2014) Quality analysis of extrusion-textured food products. In: Bouvier J-M, Campanella OH (eds) Extrusion processing technology: Food and non-food biomaterials, John Wiley and Sons, Oxford, pp 311-349

Boyer CD (1985) Soluble starch synthases and starch branching enzymes from developing seeds of sorghum. Phytochemistry 24:15-18

Bradley (Jr) RL (2014). Moisture and total solids analysis. In: Nielsen S (ed) Food Analysis, 4th edn. Springer, New York, pp 85-104

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009). BLAST+: architecture and applications. BMC Bioinformatics 10:421. doi:10.1186/1471-2105-10-421

Campbell BC, Gilding EK, Mace ES, Tai S, Tao Y, Prentis PJ, Thomelin P, Jordan DR, Godwin ID (2016) Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. Plant Biotechnol J 14:2240-2253

Chojnacki S, Cowley A, Lee J, Foix A, Lopez R (2017) Programmatic access to bioinformatics tool from EMBL-EBI update: 2017. Nucleic Acids Res 45:W550-W553. doi:10.1093/nar/gkx273

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L., Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80–92

Collins A, Ke X (2012) Primer1: Primer design web service for tetra-primer ARMS-PCR. The Open Bioinforma J 6:55-58

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A,Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly, M. J (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491–8

Desam GP, Narsimhan G, Chen G, Li J, Campanella O (2018) A mechanistic model for swelling kinetics of waxy maize starch suspension. J Food Eng 222:237-249

FAO, 2019. http://www.fao.org/land-water/databases-and-software/crop-information/%20sorghum/en FAOSTAT, 2019, http://www.fao.org/faostat/en/). Accessed 2019

Figueiredo LFdeA, Calatayud C, Dupuits C, Billot C, Rami J-F, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann J-C (2008) Phylogeographic evidence of crop neodiversity in Sorghum. Genetics 179:997–1008

Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y,Yan M, Jiang H, Dong G, Huang Y, Han B, Li J, Qian Q (2011) ALK, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. J Integr Plant Biol 53: 756-765

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putman N, Rokhsar DS (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40 (Database issue), D1178–86. doi:10.1093/nar/gkr944

Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164:731-740

Griebel S, Webb MM, Campanella OH, Craig BA, Weil CF, Tuinstra MR (2019) The alkali spreading phenotype in Sorghum bicolor and its relationship to starch gelatinization. J Cereal Sci 86:41-47

Guan HP, Preiss J (1993) Differentiation of the properties of the branching isozymes from maize (*Zea Mays*). Plant Physiology 102:1269-1273

Hill H, Lee LS, Henry RJ (2012) Variation in sorghum starch synthesis genes associated with differences in starch phenotype. Food Chemistry 131:175-183

House LR, Gomez M, Sun Y, Murty DS, Verma BN (2000) Development of some agricultural industries in several African and Asian countries. In: Smith CW, Frederiksen RA (eds) Sorghum Origin, History, Technology and Production, John Wiley & Sons, New York, pp. 131-190

Jane J-l (2009). Structural features of starch granules II. In: BeMiller J, Whistler R (eds) Starch Chemistry and Technology, Elsevier, Academic Press, Amsterdam, pp 193-236

Jiao Y, Burke J, Chopra R, Burow G, Chen J, Wang B, Hayes C, Emendack Y, Ware D, and Xin Z (2016) A sorghum mutant resource as an efficient platform for gene discovery in grasses. The Plant Cell 28:1551-1562

Juliano BO, Bautista GM, Lugay JC, Reyes AC (1964) Studies on the physicochemical properties of rice. J Agric Food Chem 12:131-138

Kimber CT (2000) Origins of domesticated Sorghum and its early diffusion to India and China. In: Smith CW, Frederiksen RA (eds) Sorghum Origin, History, Technology and Production, John Wiley & Sons, New York, pp 3-98

Krothapalli K, Buescher EM, Li X, Brown E, Chapple C, Dilkes BP, Tuinstra MR (2013) Forward genetics by genome sequencing reveals that rapid cyanide release deters insect herbivory of *Sorghum bicolor*. Genetics 195:309-318

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-9

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25:2078–9

Li L, Jiang H, Campbell M, Blanco M, Jane J-l (2008) Characterization of maize amylose-extender (ae) mutant starches. Part I: Relationship between resistant starch contents and molecular structures. Carbohydr Polym 74:396-404

Little RR, Hilder GB, Dawson EH (1958) Differential effect of dilute alkali on 25 varieties of milled white rice. Cereal Chem 35:111-126

Liu F, Ahmed Z, Lee EA, Donner E, Lui Q, Ahmed R, Morell MK, Emes MJ, Tetlow IJ (2012a) Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein-protein interactions. J Exp Bot 63:1167-1183

Liu F, Romanova N, Lee EA, Ahmed R, Evans M, Gilbert EP, Morell MK, Emes MJ, Tetlow I. J (2012b) Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. Biochem J 448:373-387

Loveless A (1958) Increased rate of plaque-type and host-range mutation following treatment of bacteriophage in vitro with ethyl methane sulphonate. Nature 181:1212–1213

Luo J, Ahmed R, Kosar-Hashemi B, Larroque O, Butardo (Jr) VM, Tanner GJ, Colgrave ML,·Upadhyaya NM, Tetlow IJ, Emes MJ, Millar A, Jobling SA, Morell MK, Li Z (2015) The different effects of starch synthase IIa mutations or variation on endosperm amylose content of barley, wheat and rice are determined by the distribution of starch synthase I and starch branching enzyme IIb between the starch granule and amyloplast stroma. Theor Appl Genet 128:1407-1419

Mariotti M, Fongaro L, Catenacci F (2010) Alkali spreading value and image analysis, J Cereal Sci 52:227-235

McKneight MM (2015) Functional genomics of sorghum grain quality traits. Master Thesis, Purdue University, West Lafayette

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**.** *Genome Research 20:1297-303*

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res 8:4321-4325

Mutisya J, Sathish P, Sun Chuanxin, Andersson L, Ahlansberg S, Baguma Y, Palmqvist S, Odhiambo B, Åman P, Jansson C (2003) Starch branching enzymes in sorghum (Sorghum bicolor) and barley (Hordeum vulgare): Comparative analyses of enzyme structure and gene expression. J Plant Physiol 160:921-930

Nakamura Y (2015). Biosynthesis of reserve starch. In: Nakamura Y (eds), Starch metabolism and structure, Springer, Japan, pp 161-209

Nakamura Y (2015b) Manipulation of rice starch properties for application. In: Nakamura Y (ed) Starch metabolism and structure, Springer, Japan, pp 335-369

NCBI Primer BLAST (2016) https://www.ncbi.nlm.nih.gov/tools/primer-blast/. Accessed 2016

NCBI sequence repository (2016). Sorghum bicolor mitochondrion, complete genome. NCBI Reference Sequence: NC_008360.1. https://www.ncbi.nlm.nih.gov/nuccore /NC_008360.1. Accessed 2016

Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 18:613-615

Nishi A, Nakamura Y, Tanaka N, Satoh H (2001) Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. Plant Physiol 127:459-472

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J and Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. PNAS 101:9903-9908

Phytozome 11 (2016) http://www.phytozome.net, https://phytozome.jgi.doe.gov/ pz/portal.html to download version 11. Accessed 2016

Picard Toolkit (2016) Broad Institute, GitHub Repository. http://broadinstitute.github.io/ picard/; Broad Institute. Accessed 2016

Preiss J (2009) Biochemistry and molecular biology of starch biosynthesis. In: BeMiller J, Whistler R (eds) Starch Chemistry and Technology, Elsevier, Academic Press, Amsterdam, pp 83 – 148

Ragheb AA, Abd El-Thalouth I, Tawfik S (1995) Gelatinization of starch in aqueous alkaline solutions. Starch/Stärke 46:338-345

Rooney LW (2000). Genetics and cytogenetics. In: Smith CW, Frederiksen RA (eds) Sorghum Origin, History, Technology and Production, John Wiley & Sons, New York, pp 261-307

Sang Y, Bean S, Seib PA, Pedersen J, Shi Y-C (2008) Structure and functional properties of sorghum starches differing in amylose content. J Agric Food Chem 56:6680-6685

Shannon JC, Garwood DL, Boyer CD (2009) Genetics and physiology of starch development. In: BeMiller J, Whistler R (eds) Starch Chemistry and Technology, Elsevier, Academic Press, Amsterdam, pp 23-82

Swigoňová Z, Lai J, Ma J, Ramakrishna W, Lllaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. *Genome Research* 14:1916-1923.

Tan Y, Corke H (2002) Factor analysis of physiochemical properties of 63 rice varieties. *J Sci Food Agr* 82:745-752.

Taylor J, Dewar J (2000) Fermented Products: Beverages and Porridges. In: Smith CW, Frederiksen RA (eds) Sorghum Origin, History, Technology and Production, John Wiley & Sons, New York, pp. 751-796.

Tetlow IJ, Emes MJ (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. IUBMB Life 66:546-558

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. J Exp Bot 55:2131-2145

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M, Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. PNAS 106:21760-21765

Westergaard M (1957) Chemical mutagenesis in relation to the concept of the gene. Experientia 13:224–234

Wang LQ, Liu WJ, Xu Y, He YQ, Luo LJ, Xing YZ, Xu CG, Zhang Q (2007) Genetic basis of 17 traits and viscosity parameters characterizing the eating and cooking quality of rice grain. Theor Appl Genet 115:463-476

Waniska RD, Rooney LW (2000). Structure and chemistry of the sorghum caryopsis. In: Smith CW, Frederiksen RA (eds) Sorghum Origin, History, Technology and Production, John Wiley & Sons, New York, pp 649-688

Xin Z, Velten JP, Oliver MJ, Burke JJ (2003) High-throughput DNA extraction method suitable for PCR. BioTechniques 34:820-4, 826

Ye S, Dhillon S, Ke X, Collins AR, Day IN (2001) An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res 29:E88-8

Zhang X, Colleoni C, Ratushna V, Sirghie-Colleoni M, James MG, Myers AM (2004) Molecular characterization demonstrates that the *Zea mays* gene sugary2 codes for the starch synthase isoform SSIIa. Plant Mol Biol 54:865-879

CHAPTER 4. GENETIC ANALYSIS OF THE ALKALI SPREADING PHENOTYPE IN A SORGHUM DIVERSITY PANEL

A version of this chapter has been submitted to a peer reviewed
academic journal for publication. Unpublished manuscript.
Will be published as open access (CC BY)

Stefanie Griebel[a], Adeyanju A[a] and Mitchell R. Tuinstra[a]

[a] Department of Agronomy, Purdue University, West Lafayette, IN 47907

4.1 Declaration of Author Contributions to this Chapter

Stefanie Griebel performed and designed research, wrote paper. Adeyanju A provided training on GWAS. Mitchell R. Tuinstra major advisor and project lead, guided research. All co-authors reviewed and edited the paper.

4.2 Abstract

Sorghum is an important cereal crop in Africa and South Asia. In this study, a diversity panel of sorghum breeding lines and 788 sorghum conversion (SC) lines representing the global germplasm diversity of the crop were evaluated for alkali spreading value (ASV), a qualitative measure of starch quality. A small set of genotypes with stable expression of ASV across seasons were identified; mostly representing the race *durra*, especially *Nandyal* types from India. Genetic studies showed the ASV+ phenotype was inherited as a recessive trait. Whole genome resequencing revealed SNPs in genes involved in starch biosynthesis, especially branching enzymes, that need to be further evaluated to identify the causal variation for ASV. A genome wide association study (GWAS) identified a significant SNP associated with ASV near *Sobic.010G273800,* a starch branching enzyme I precursor, and *Sobic.010G274800* and

Sobic.010G275001, both annotated as glucosyltransferases. Physiochemical analyses of accessions with known ASV phenotypes demonstrated an environment dependent lower starch gelatinization temperature (GT), amylose content of approximately 22%, and good gel consistency. The starch composition of these lines could be valuable for food products that require good gel consistency and viscosity.

4.3 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a cereal (FAOSTAT 2016; Kimber 2000) that serves as a staple food for millions of humans in Africa and South Asia (House et al. 2000). It is commercially important in many regions of the world but mostly grown in low-income, and food-deficient countries (FAOSTAT 2016, 2019). The utilization of sorghum is regionally and culturally dependent and of great economic value for smallholder farmers in Africa, because it is less expensive than other imported cereals (Taylor and Dewar 2000; House et al. 2000).

Sorghum bicolor is one of three sorghum species and has three subspecies (Dahlberg 2000). The cultivated sorghums belong to the subspecies *Sorghum bicolor subsp. bicolor* and are classified into five races based on plant/panicle/seed characteristics (Dahlberg 2000; Harlan and de Wet 1972). The five basic cultivated races are bicolor, guinea, kafir, caudatum and durra, which can combine to intermediate races (Dahlberg 2000; Harlan and de Wet 1972). Sorghums are classified based on Snowden's first classification in 1936, which was modified over decades by Murty and Govil, Harlan and de Wet, and Dahlberg (Dahlberg 2000; Harlan and de Wet 1972). Phenotypic evaluations and genomic studies reported that the five sorghum races originated from Africa and Asia (Morris et al. 2013; Kimber et al. 2013; Harlen and de Wet 1972, Dahlberg 2000, Deu et al. 2006; Brenton et al. 2016; Billot et al. 2013; Sukumaran et al. 2012); with northeast Africa as the center of diversity (Paterson et al. 2009).

The sorghum germplasm collections represent tremendous genetic diversity with more than 36,000 accessions available and stored at ICRISAT in India (Billot et al. 2013; Kimber et al. 2013) and more than 40,000 accessions at the USDA National Center for Genetic Resources Preservation (USDA-ARS-PGRCU) (Kimber et al. 2013). To make those accessions more readily available for research and breeding in the temperate zone, the Sorghum Conversion Program was initiated by Stephens et al. (1967). The program converted alien sorghums from the world collection into photoperiod insensitive, short, early-maturing sorghums (Stephens et al. 1967; Rosenow et al. 1997a, 1997b). Over 800 converted sorghum lines (SC lines) are currently available (Hayes et al. 2015; Kimber et al. 2013). The collection of SC lines embodies the genetic diversity of the crop and provides the basis for sorghum diversity panels such as the Sorghum Association Panel (SAP) (Casa et al. 2008; Morris et al. 2013; Cuevas et al. 2017; Boyles et al. 2017; Adeyanju et al. 2015; Mace et al. 2013; Sukumaran et al. 2012; Shenstone et al. 2018) and the Sorghum Diversity Panel (SDP) (Hayes et al. 2015).

The sorghum diversity panels are important resources used in genome wide association studies (GWAS). And in these studies genotype-by-sequencing (GBS) markers are associated with phenotypes to identify quantitative trait loci (QTLs) of interest (Morris et al. 2013; Boyles et al. 2017; Adeyanju et al. 2015). The sorghum LD decay depends on genotypes used and is reported to be approximately 15kb (Kimber et al. 2013), with slower LD decay in inbred lines (19.7kb) than in landraces (10.3kb) (Mace et al. 2013) and can be as fast as 1kb (Morris et al. 2013). Mixed linear models are commonly used for GWAS (Huang et al. 2010; Zhao et al. 2011; Morris et al. 2013; Adeyanju et al. 2015). The Logistic Mixed Model (LMM) has been suggested for analysis of binary traits in maize and sorghum diversity panels (Shenstone et al. 2018).

The alkali spreading phenotype (ASV) has been reported to identify sorghum genotypes with contrasting starch thermal properties (Griebel et al. 2019). The key gene for ASV in rice is the *ALK* (*SSII-3*) gene that also controls starch gelatinization temperature (GT) (Gao et al. 2011; Tian et al. 2009). GWAS for ASV has been conducted in rice with several genomic regions associated with the trait (Zhao et al. 2011; Mogga et al. 2018; Song et al. 2019; Huang et al. 2010; Huang et al. 2012). Only a few of the associations were close to candidate genes such as one on chromosome 7 that was close to amylase inhibitors (Huang et al. 2012), on chromosome 6 near the *ALK (SSII-3)* gene (Huang et al. 2010; Zhao et al. 2011), on chromosome 8 within 200kb of the *SSII-2* (Zhao et al. 2011), and on chromosome 5 (Song et al. 2019). The ASV hits were not confirmed across years (Zhao et al. 2011). Griebel et al. (2019b) showed that *SSIIa* and *SBEIIb* are key genes controlling the alkali spreading phenotype in sorghum mutants. These genes have contrasting effects on starch GT, amylose content, and paste viscosity profiles (Griebel et al. 2019b).

Our understanding of the genetic architecture for ASV in the standing varation of sorghum is limited. Thus, we aim to (1) conduct a GWAS for the ASV phenotype in a panel of more than 800 breeding and SC-lines of sorghum, (2) identify candidate genes for ASV based on genomic regions identified in the GWAS, and (3) describe the starch phenotypes for accessions with ASV phenotypes and determine whether ASV is inherited as a dominant or recessive trait.

4.4 Results

4.4.1 Racial and Geographic Classification of the Sorghum Diversity Panel

The sorghum diversity panel consists of approximately 788 SC lines from the sorghum conversion program and 51 sorghum breeding lines included as checks. The SC lines originated from a diverse set of countries before being converted in the sorghum conversion program (**Figure 4.1; Table S.4.1**; Stephens et al. 1967, Klein, personal communication, 2019). The countries of

origin are predominantly India (171 SC lines), Ethiopia (139 SC lines), Sudan (116 SC lines) and 305 SC lines from approximately 25 other African countries (**Table S.4.1**). The races represented in the diversity panel are predominant bicolor (37), caudatums (268), durra (171), durra-bicolor (61), guinea (124), kafir (21) and kafir-caudatum (22) (Klein, personal communication, 2019).

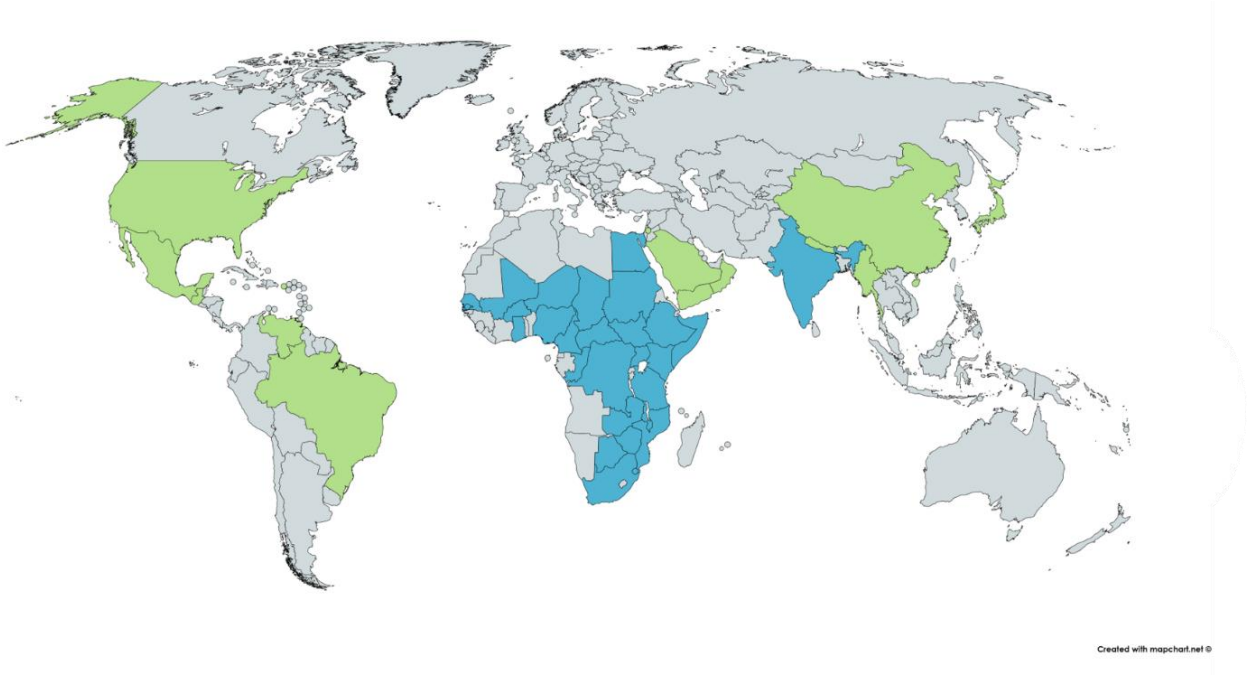


Figure 4.1. Sorghum diversity panel showing countries of origin in blue and green. The majority of the lines (blue) are from Africa and from India (Country of origin kindly provided by Klein, personal communication, 2019).

The GBS data (Brown, personal communication, 2019; Brown, 2017) consisting of 80,103 SNPs were used to conduct a principal component analysis to determine the covariates used for GWAS and to understand how the sorghum lines clustered in the diversity panel based on ASV, race, and origin. A scree plot from the principal component (PC) analysis showed that the first five PCs explained most of the variation from the SNP markers (**Figure 4.2**). The red line was drawn at the characteristic elbow as suggested by Wang et al. (2013) These five clusters may reflect the five basic races of sorghum.

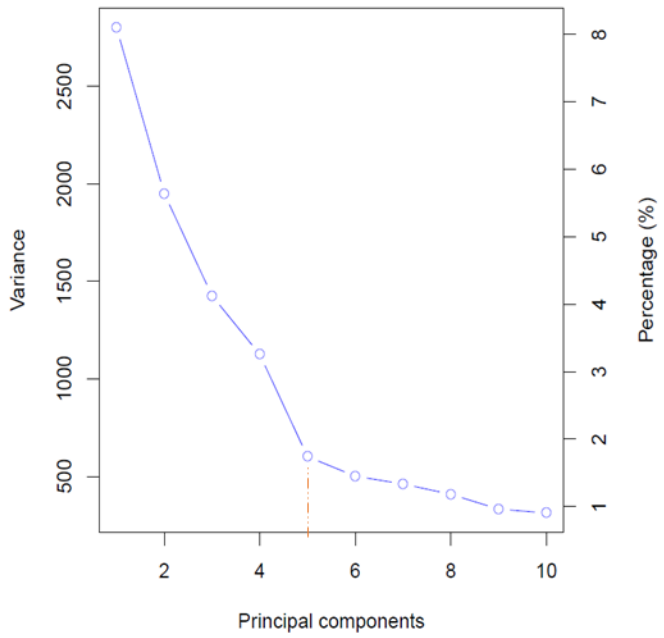


Figure 4.2. Scree plot of principal components (X-axis) and their contribution to the variance determined from GBS data. Threshold (red line) at characteristic elbow determines the optimal number of PCs.

The first 4 PCs were plotted against each other and colored by their race classification, region of origin and appearance of the ASV phenotype (**Figures 4.3 and 4.4**). PCA 5 explained the least variation and was excluded from **Figure 4.3 and 4.4**. The PC1 captures most of the variation and is represented by durra types from India and some durra-bicolors from eastern Africa (**Figure 4.3**). The PC2 was represented by two cluster, one showing variation in caudatum types originating from northern and eastern Africa and the second representing the guineas and some caudatums from West Africa. A few lines are from other African countries and India. The PC3 largely represents differentiation of the kafirs from eastern and southern Africa. The PC4 is dominated by the durra-bicolor cluster and some bicolors from eastern Africa (**Figure 4.4**). PC5 did not show a clear distinction between groups but consisted among others of a group of guineas from West Africa listed as a second group in PC2.

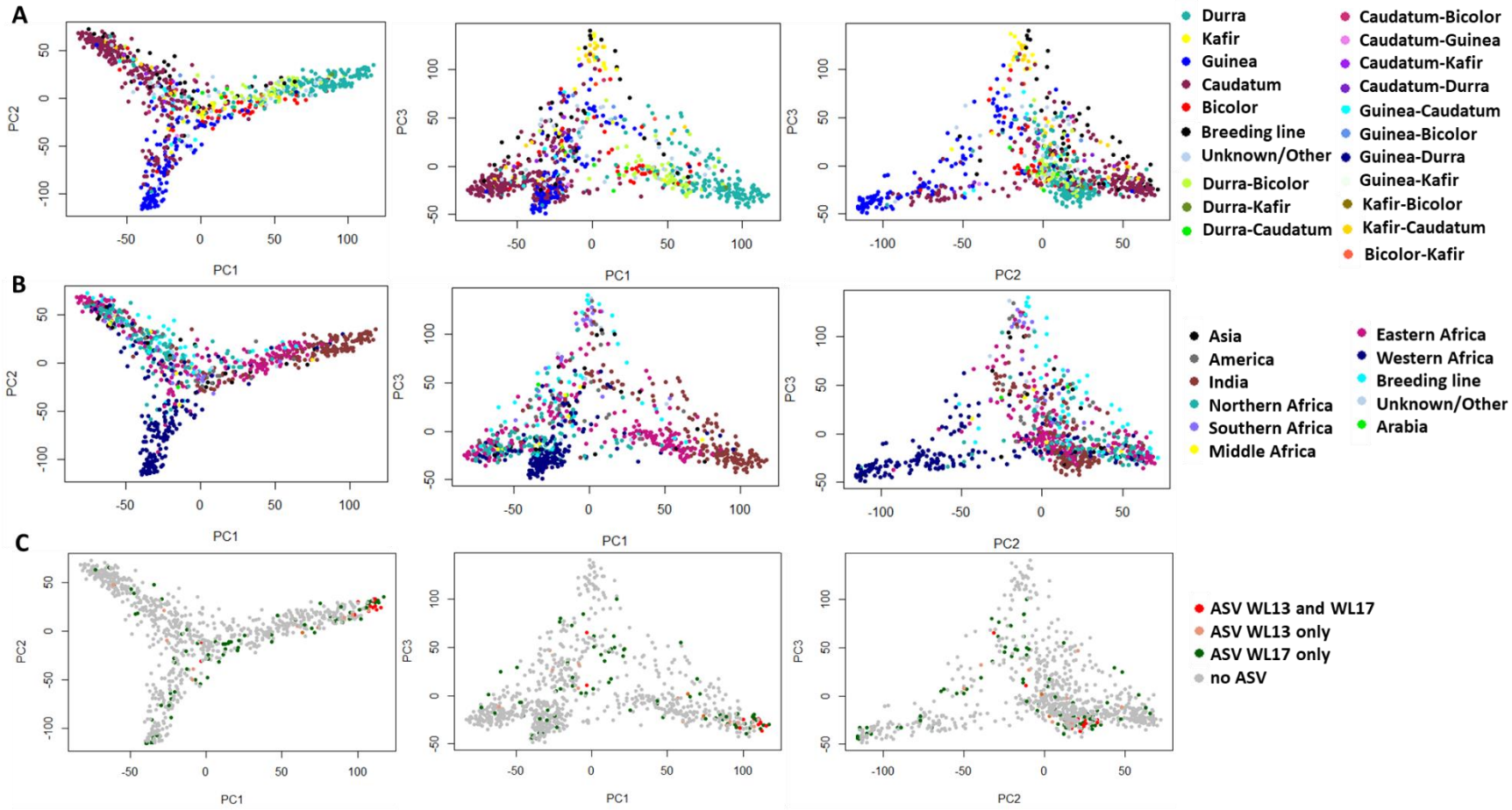


Figure 4.3. (and Table S4.1) Principal components showing clustering of the sorghum diversity panel by (A) race, (B) geographic origin, and (C) ASV phenotype – based on 5 race classification. 2D plots with principal components 1, 2 and 3. Genotypes were evaluated in the seasons West Lafayette 2013 (WL13) and 2017 (WL17).

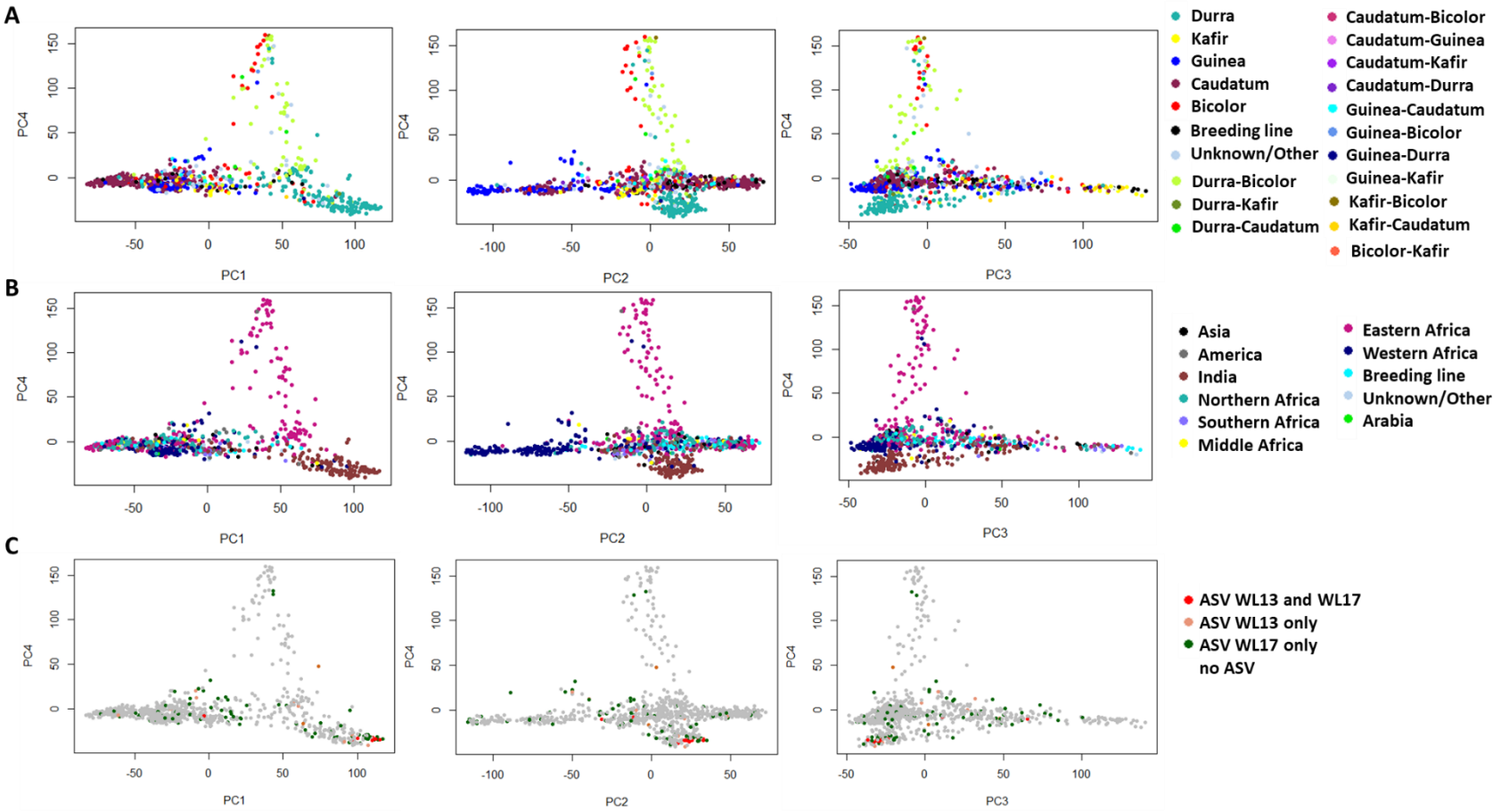


Figure 4.4. (and Table S4.1) Principal components showing clusters of the sorghum diversity panel by (A) race, (B) geographic origin, and (C) ASV phenotype – based on 5 race classification. 2D plots with principal components 1, 2, 3 against 4. Genotypes were evaluated in the seasons West Lafayette 2013 (WL13) and 2017 (WL17).

4.4.2 The ASV Phenotype Exists in Natural Variation

The ASV assay introduced in Griebel et al. (2019) used 1.8% KOH in a sorghum EMS population from BTx623. These conditions were also effective in distinguishing the standing variation of sorghum (**Figure 4.5**). Comparing concentrations of 1.5%, 1.8% and 2.0% KOH showed that BTx623 did not exhibit any ASV phenotype. Three SbEMS mutants of known ASV+ phenotype (Griebel et al. 2019) and four SC lines clearly showed the expected ASV+ phenotype at 1.8% and 2.0%, while 1.5%KOH was not effective in expressing the ASV phenotype. This gave confidence to continue screening with 1.8%KOH as described in Griebel et al. (2019).

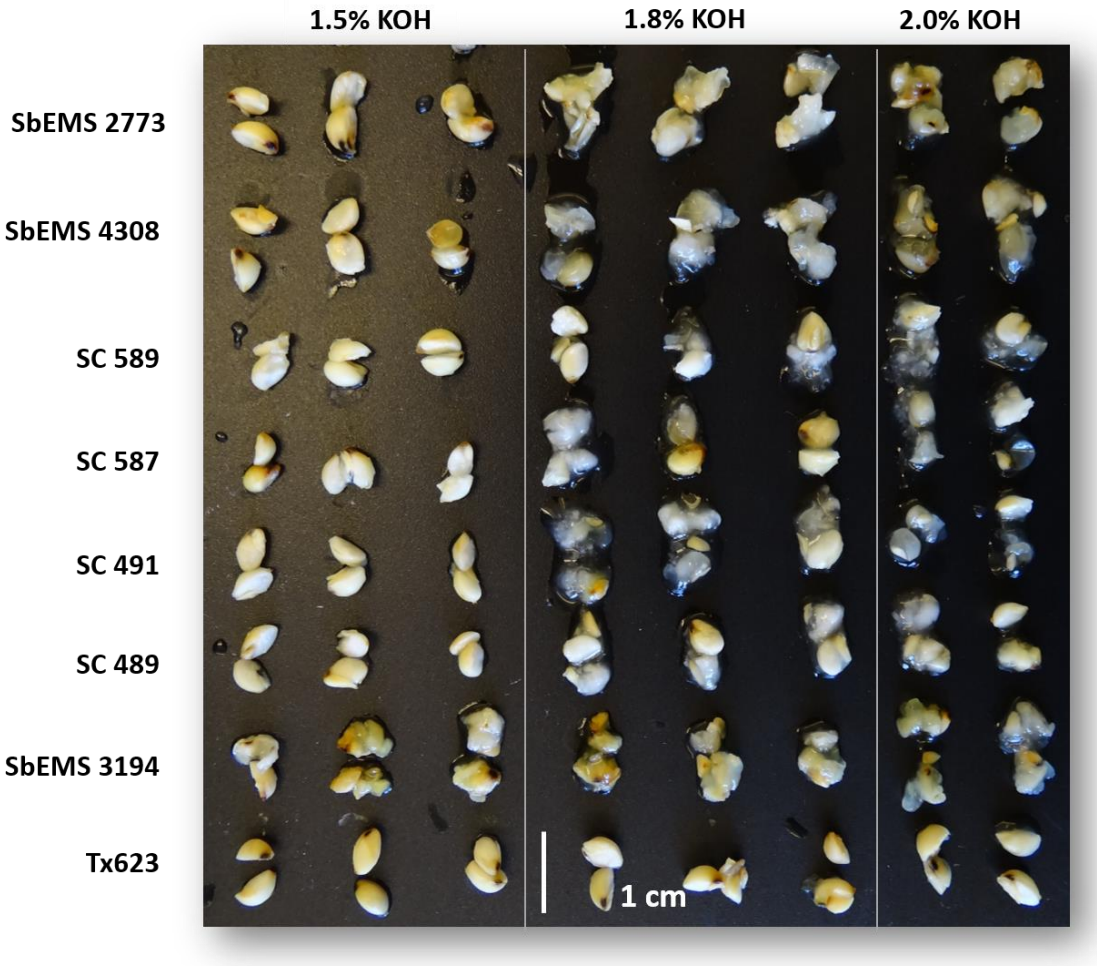
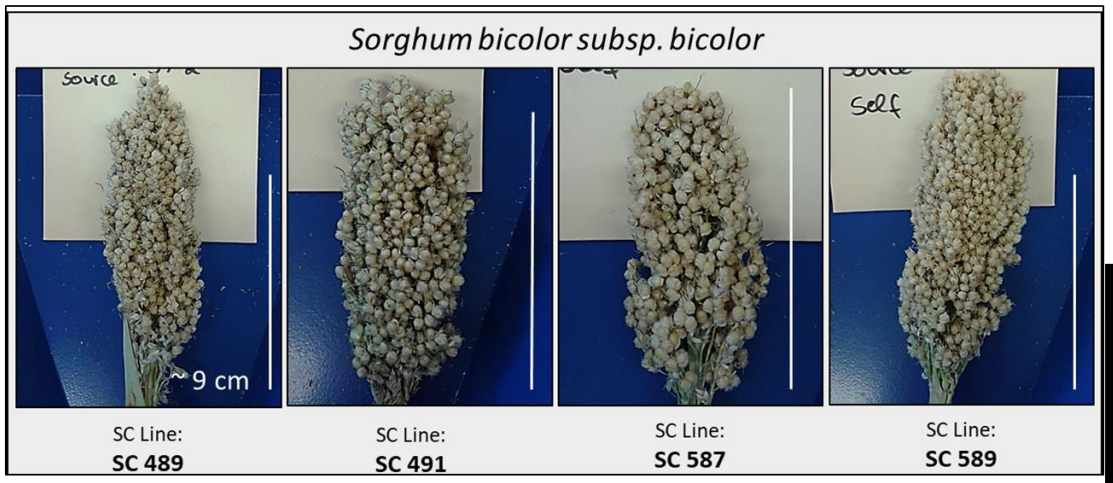


Figure 4.5. ASV phenotypes at 1.5%, 1.8% and 2.0% KOH for known SbEMS mutants and four SC lines having an ASV+ phenotype and BTx623 having an ASV- phenotype.

Variation in the ASV phenotype was found in the sorghum diversity panel in 2013 and 2017 but only a low trait correlation was observed between years (r=0.23). There were 24 lines that exhibited the ASV phenotype in WL2013 (salmon dots in **Figure 4.3 and 4.4; Table S4.1**) and 95 lines that showed the ASV in WL2017 (green dots in **Figure 4.3 and 4.4; Table S4.1)** with only 13 lines exhibiting the phenotype in both years. The principal component analysis helped to show that the 13 lines (red dots; **Figure 4.3 and 4.4**) clustered together at the tip of PC1 being *durras* from India (**Figure 4.3**). The genotypes exhibiting an ASV only in WL13 or only in WL17 have no clear race and geographic distinction (**Figure 4.3 and 4.4**).

The passport data of these 13 lines show that 11 belong to the working group 46(1) Nandyal (**Figure 4.6, B**). One other line belongs to the race *Guinea* and the working group 1: Roxburghii. One line with an ASV phenotype has no clear race classification (Klein 2019). The phylogenetic analysis reported as a neighbor-joining tree created in TASSEL (Tassel 2018; Bradbury et al. 2007) confirmed the PC results that the 11 lines cluster together and being very closely related (**Figure 4.6, C**). We selected four SC-lines, SC489, SC491, SC587 and SC589 for further genetic, genomic and physiochemical analysis (**Figure 4.6, A**).

A



B

SC line	Year	Donor	Recurrent	PI name	Local Name	Country	Race	Working Group
SC0199	1972	BTx406	IS1121C	533810	Karad 2-7-11	India	Durra	46(1): Nandyal
SC0242	1972	BTx406	IS3955C	533844	EC 18103	Nepal	Guinea	1: Roxburghii
SC0448	1997	BTx406	IS5843	NA	Safed Bedra Nayagaon	India	Durra	41:Durra
SC0477	1986	BTx406	IS6398C	534026	Jondla var. Gidgap	India	Durra	46(1): Nandyal
SC0479	NA	BTx406	IS6403	NA	Jowar Kalgundi	India	Durra	46(1): Nandyal
SC0483	1986	BTx406	IS6445C	534034	Karad Local	India	Durra	46(1): Nandyal
SC0489	1972	BTx406	IS6389C	533856	Jola Nandyal	India	Durra	46(1): Nandyal
SC0491	1986	BTx406	IS6394C	534025	Jondla var. Gidgap	India	Durra	46(1): Nandyal
SC0499	1972	BTx406	IS6439C	533858	Jowar Shenoli 4-2	India	Durra	46(1): Nandyal
SC0586	1986	BTx406	IS6446C	534035	Karad Local	India	Durra	46(1): Nandyal
SC0587	1986	BTx406	IS6356C	534021	Jola Nandyal	India	Durra	46(1): Nandyal
SC0589	1986	BTx406	IS6388C	534023	Jola Nandyal	India	Durra	46(1): Nandyal
SC0971	1989	BTx406	Millo blanco	PI 656111	Local variety	Puerto Rico (USA)	?	50(1): NA

Kindly provided by Robert R. Klein (USDA-ARS, personal communication); and USDA GRIN database

C

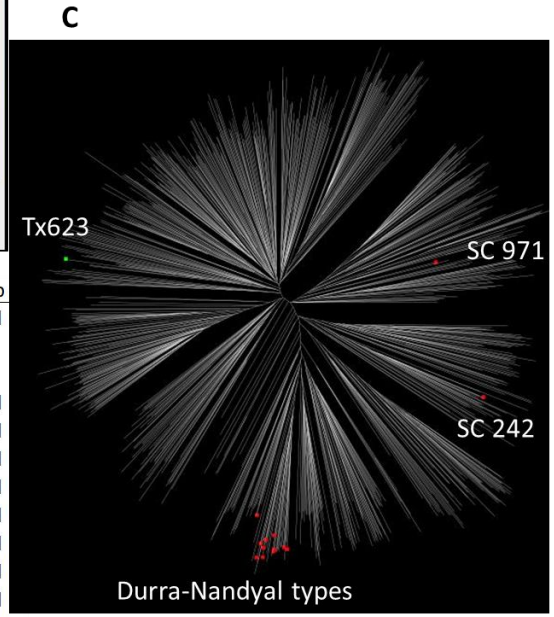


Figure 4.6. Passport data of a subset of sorghum conversion lines – **(A)** Panicles from a subset of lines exhibiting an ASV+ phenotype – SC489, SC491, SC587 and SC589. **(B)** SC lines exhibiting an ASV+ phenotype in West Lafayette 2013 and 2017. **(C)** Phylogenetic analysis: neighbor-joining tree – ASV genotypes across both environments in red, BTx623 in green.

4.4.3 SC-lines with ASV+ mostly Share SNPs in Branching Enzymes

Whole genome re-sequencing of SC489, SC491, SC587 and SC589 revealed more than 200 homozygous SNPs in candidate genes for starch biosynthesis (**Table 4.1**). However, only around 107 SNPs were shared in all four lines and approximately 59 of these SNPs were in starch branching genes, like SBEIIb *Sobic.004G163700* (**Table 4.1**).

Table 4.1. SC-lines SC487, SC491, SC587, SC589 share common SNPs in starch biosynthesis genes from whole genome re-sequencing analyses.

Gene	Gene description	SNP, INDELS position (REF --> ALT)
Sobic.001G083900	similar to Alpha 1,4-glucan phosphorylase L isozyme	6491557 (G-->C), 6493277 (A-->C), 6496774 (T-->C)
Sobic.001G239500	similar to putative uncharacterized protein SSI	24609069 (A-->G), 24609559 (C-->T), 24610385 (C-->T), 24610644 (G-->A), 24610664 (G-->A), 24612553 (A-->G), 24612704 (G-->A), 24612729 (T-->A), 24613491 (T-->C), 24614286 (TA-->T)
Sobic.002G116000	similar to GBSS IIa	14336638 (T-->A), 14336679 (C-->T), 14338839 (CT-->C), 14340488 (A-->G), 14340846 (G-->A)
Sobic.002G233600	similar to Isoamylase-type starch debranching enzyme ISO3	62457714 (C-->T), 62458735 (T-->G), 62461676 (C-->A), 62461956 (C-->T), 62462103 (G-->A), 62462895 (T-->C), 62462994 (T-->C), 62463176 (G-->A), 62466487 (A-->T), 62467260 (G-->A), 62467479 (T-->C), 62468169 (T-->A)
Sobic.003G213800	similar to Putative 1,4-alpha-glucan branching enzyme	54790394 (A-->G), 54790433 (C-->T), 54790535 (T-->C), 54790620 (G-->A), 54790751 (G-->T), 54791826 (T-->G), 54791826 (T-->G), 54792264 (G-->A), 54792265 (A-->G), 54793301 (A-->AAGTGAAATTATG), 54793305 (T-->A)
Sobic.004G163700	ae1, SbeIIb, similar to 1,4-alpha-glucan-branching enzyme 2	51292259 (A-->AAC), 51292274 (C-->CT), 51292275 (C-->A), 51292342 (A-->G), 51292345 (G-->C), 51292352 (G-->A), 51292365 (T-->C), 51292416 (G-->C), 51292465 (T-->C), 51292472 (A-->T), 51292476 (T-->C), 51292514 (G-->A), 51292538 (TTTTC-->T), 51292549 (A-->G), 51292554 (CA-->C), 51292556 (GCCCC-->G), 51292563 (A-->G), 51292716 (A-->C), 51293242 (T-->A), 51294803 (C-->T), 51296235 (A-->G), 51297245 (A-->G), 51299170 (A-->G), 51299810 (T-->A), 51300412 (A-->C), 51300803 (T-->A), 51302117 (A-->AT), 51302133 (G-->A), 51303947 (A-->ATT), 51303977 (A-->G), 51304083 (G-->C), 51304124 (A-->AGCG), 51304166 (AG-->A)
Sobic.006G066800	BEIIa, 1,4-Alpha-Glucan-Branching Enzyme 2-1	42708514 (T-->TC), 42710309 (A-->C), 42710479 (G-->A), 42710744 (T-->C), 42711487 (AT-->A), 42712138 (C-->T), 42712157 (T-->C), 42712302 (T-->C), 42712575 (A-->C), 42713817 (G-->A), 42713964 (G-->T),42714307 (C-->A), 42715316 (A-->AAAC), 42716348 (G-->A), 42717317 (AGG-->A)
Sobic.006G221000	SS3, similar to Starch synthase IIIb-1	56789566 (C-->CCTCT), 56789696 (C-->T), 56789794 (G-->A), 56790054 (G-->T), 56790087 (G-->A), 56790140 (T-->A), 56790142 (T-->A), 56790147 (A-->G), 56790164 (G-->A)
Sobic.007G068200	similar to Starch synthase DULL1	7589674 (GGA-->G), 7596076 (G-->T)
Sobic.010G022600	similar to GBSS 1	1861246 (CT-->C), 1861269 (G-->GCTA)
Sobic.010G047700	Starch Synthase 2	3699364 (T-->C), 3700536 (G-->T)
Sobic.010G093400	similar to SSIIa, su2	8300714 (CTA-->C), 8303592 (C-->T), 8304081 (T-->C)

Gene description from Campbell et al. (2016) and *Phytozome* version 12.1 *(2019).*

4.4.4 First Genome-Wide Association Study of ASV in *Sorghum bicolor*

We report the first GWAS of the ASV phenotype in sorghum. The genome wide linkage disequilibrium (LD) decay in the SDP was estimated using GAPIT and reported to be approximatly 12kb (**Figure 4.7**). A mixed linear model (MLM) (Lipka et al. 2012) and a logistic mixed model LMM (Shenstone et al., 2018), both controlling for population structure, with 5 PCAs as covariates, and kinship were used to evaluate the genetic variation of ASV in sorghum. We did not run a combined analysis across environments due to the low trait correlation of 0.23 between environments. Therefore, we report GWAS with location specific loci associated with ASV.

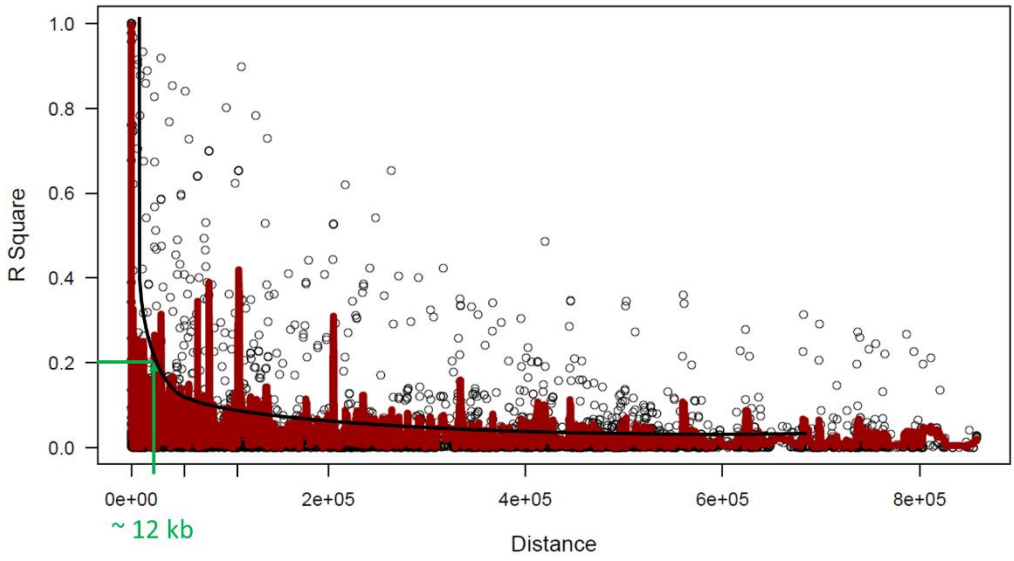


Figure 4.7. Linkage Disequilibrium (LD) decay in the sorghum diversity panel.

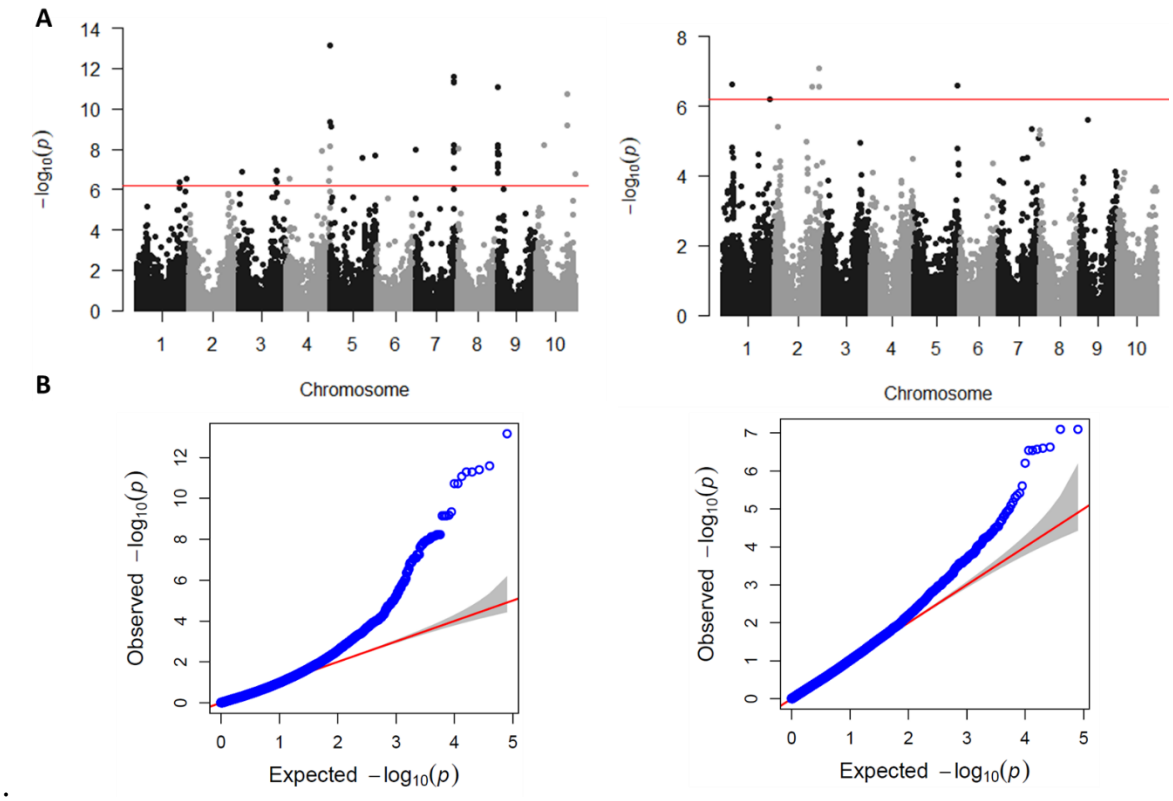


Figure 4.8. GWAS results for ASV for grain samples produced in WL13 (left) and WL17 (right) with (**A**) Manhattan plots and (**B**) QQ plots from MLM.

The number of significant SNPs differed between the MLM and LMM models used in the GWAS (**Figure 4.8 and 4.11**). Using MLM model and the bonferroni threshold of 6.242E-07 (Bonferroni 1936), we identified 53 SNPs for WL2013 and 8 SNPs for WL2017. Using the FDR (Benjamini and Hochberg, 1995) adjusted p values at 0.05, we identified 205 SNPs for WL2013 and 14 SNPs for WL2017. None of the approximately 220 significant SNPs were close (less than approximately 150kb) to a list of known genes reported by Campbell et al. (2016 Table 1) involved in starch biosynthesis in the amyloplast. Only one SNP S10_60808604 was approximately 113kb away from a gene involved in starch biosynthesis, *Sobic.010G273800.*

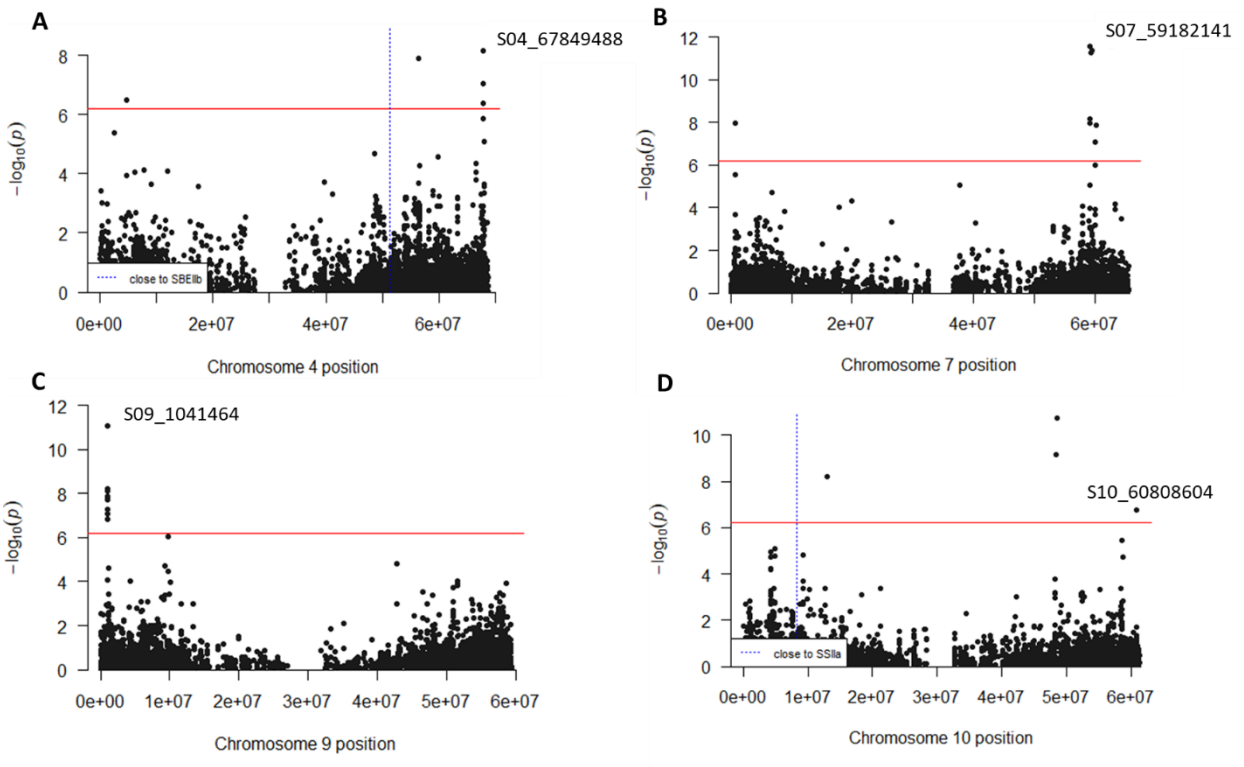


Figure 4.9. **GWAS results** from ASV for grain samples produced in WL13 using the MLM. (**A)** chromosome 4 with locator of *SSIIa* (dotted blue line), (**B**) chromosome 7, (**C**) chromosome 9 and (**D**) chromosome10 with locator of SBEIIb (dotted blue line).

Figure 4.9 shows information on Chromosomes 4, 7, 9 and 10 with the most significant SNP labeled except for Chromosome 10 labeled with the significant SNP S10_60808604. The Chromosomes 4 and 10 are marked with a dotted blue vertical line indicating the location of a SNP close to *Sobic.004G0163700*, a *SBEIIb*, and *Sobic.010G93400*, a *SSIIa*. Both genes were identified as causal genes for ASV in an EMS population (Griebel et al. 2019b). No significant GWAS hit was reported close to these two genes. The Chromosomes 4, 7 and 9 exhibit numerous significant SNPs but with the top SNP not being in close proximity of a gene known in starch biosynthesis. The genes around those SNPs are mostly uncharacterized proteins (**Table 4.2**).

Table 4.2. GWAS results from MLM with significant SNPs and candidate genes for West Lafayette (WL) 2013 (10 most significant SNPs and S10_60808604, S04_67849488) and 2017 at FDR 0.05.

Env.	SNP	P value	MAF	FDR adj. P value	Candidate Gene Annotation
WL13	S10_60808604	1.75E-07	0.03	0.0003	Sobic.010G273800.2.p – similar to starch branching enzyme I precursor (Campbell et al. 2016) Sobic.010G274700.1.p - (1 of 55) PF00069//PF00560//PF08263//PF13855 - Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine rich repeat N-terminal domain (LRRNT_2) // Leucine rich repeat LRR_8) Sobic.010G274800.1.p - similar to Glycosyltransferase QUASIMODO1, putative, expressed Sobic.010G274900.1.p - similar to Putative uncharacterized protein Sobic.010G275001.1.p - (1 of 2) PF01697 - Glycosyltransferase family 92 (Glyco_transf_92)
WL13	S05_875962	7.03E-14	0.03	5.63E-09	Sobic.005G010000.1.p - Predicted protein Sobic.005G010100.1.p - similar to BTB/POZ domain containing protein, expressed Sobic.005G010200.1.p - similar to Hypoxia induced protein conserved region containing protein, expressed
WL13	S05_875955	4.60E-10	0.08	4.10E-06	
WL13	S07_59182141	2.61E-12	0.05	8.18E-08	Sobic.007G157400.2.p - (1 of 8) 4.2.1.105 - 2-hydroxyisoflavanone dehydratase Sobic.007G157500.1.p - similar to Putative uncharacterized protein Sobic.007G157550.1.p - (1 of 5) PTHR23024//PTHR23024:SF137 - Member 'GDXG' Family of Lipolytic enzymes Sobic.007G157600.1.p - Predicted protein Sobic.007G157700.1.p - similar to NADPH HC toxin reductase
WL13	S09_1041464	8.27E-12	0.03	1.10E-07	Sobic.009G011600 - no annotation Sobic.009G011700.1.p – similar to Putative uncharacterized protein Sobic.009G011400.1.p – similar to Putative uncharacterized protein Sobic.009G011500.1.p – Predicted protein Sobic.009G011600 no annotation Sobic.009G011700.1.p – similar to Putative uncharacterized protein
WL13	S07_59500280	4.08E-12	0.05	8.18E-08	Sobic.007G160300.1.p - similar to D-type cyclin Sobic.007G160400.1.p - similar to Putative uncharacterized protein
WL13	S07_59224740	5.10E-12	0.05	8.18E-08	Sobic.007G158000.1.p - weakly similar to Putative uncharacterized protein
WL13	S07_59224782				
WL13	S10_48542785	1.87E-11	0.03	1.88E-07	Sobic.010G164200.1.p - similar to Putative uncharacterized protein Sobic.010G164300.1.p - (1 of 2) PTHR23041:SF65 - E3 UBIQUITIN-PROTEIN LIGASE CCNB1IP1 HOMOLOG
WL13	S10_48542812	1.87E-11	0.03	1.88E-07	
WL13	S10_48403358	6.62E-10	0.04	4.36E-06	Sobic.010G164100.1.p - (1 of 1) PTHR19139//PTHR19139:SF183 - AQUAPORIN TRANSPORTER
WL13	S04_67849488	7.15E-09	0.07	3.10E-05	Sobic.004G349500.1.p - Similar to Plant integral membrance protein TIGR01569 containing protein expressed Sobic.004G349600.1.p – similar to MATE efflux protein like Sobic.004G349650.1 – no annotation ; Sobic.004G349800.2- no annotation
WL17	S02_70502500	8.07E-08	0.07	0.003	Sobic.002G337700.1.p - similar to Putative uncharacterized protein Sobic.002G337800.1.p - similar to Myb protein
WL17	S02_70502518	8.07E-08	0.07		
WL17	S02_70502913	2.84E-07	0.06		
WL17	S02_70502916	2.84E-07	0.06		

Table 4.2 continued

WL17	S01_15639633	2.31E-07	0.03	0.003	Sobic.001G183550.1 – no annotation Sobic.001G183600.1.p - similar to Putative uncharacterized protein Sobic.001G183633.1 – no annotation Sobic.001G183666.1 – no annotation
WL17	S02_59605124	2.66E-07	0.03	0.003	Sobic.002G204400.1.p - similar to Putative uncharacterized protein Sobic.002G204500.1.p - similar to Phospholipase D lambda Sobic.002G204600.1.p - weakly similar to Putative uncharacterized protein Sobic.002G204700.1.p - (1 of 5) PTHR23067//PTHR23067:SF47 - Double-stranded RNA-Binding ZINC FINGER
WL17	S05_68193148	2.54E-07	0.04	0.003	Sobic.005G196600.1.p - Predicted protein Sobic.005G196700.1.p - similar to Sugar transporter family protein, expressed Sobic.005G196800.1.p - similar to Os12g0514100 protein Sobic.005G196900.3.p - similar to Potyvirus VPg interacting protein, putative, expressed Sobic.005G196950.1.p - similar to Os12g0514600 protein
WL17	S01_73843118	6.23E-07	0.05	0.006	Sobic.001G464200.1.p - similar to XPG I-region family protein, expressed Sobic.001G464300.1.p - similar to Helix-loop-helix DNA-binding domain containing protein Sobic.001G464400.3.p - similar to Putative uncharacterized protein Sobic.001G464500.1.p - similar to CUE domain containing protein, expressed Sobic.001G464600.1.p - similar to Transposon protein, putative, unclassified, expressed
WL17	S09_11993119	2.52E-06	0.35	0.022	Sobic.009G082400.1 – no annotation Sobic.009G082500.1 – no annotation Sobic.009G082550.1 – no annotation
WL17	S02_5699759	3.79E-06	0.07	0.030	Sobic.002G058900.1.p - Predicted protein Sobic.002G059000.1.p - (1 of 303) PF00646 - F-box domain (F-box) Sobic.002G059100.2 – no annotation Sobic.002G059200.1.p - similar to Histone H2A
WL17	S07_51741290	4.39E-06	0.04	0.032	Sobic.007G120000.1.p - similar to Os08g0398300 protein
WL17	S08_253410	5.01E-06	0.20	0.033	Sobic.008G002600.1.p - (1 of 1) PTHR10209//PTHR10209:SF143 -Oxidoreductase, 2OG-FE II Oxygenase family Sobic.008G002700.1.p - Predicted protein Sobic.008G002800.1.p - similar to Fatty acid desaturase DES1 Sobic.008G002950.1.p - (1 of 34) PF13365 - Trypsin-like peptidase domain (Trypsin_2)
WL17	S08_274611	6.67E-06	0.18	0.041	Sobic.008G003100.2.p - (1 of 34) PF13365 - Trypsin-like peptidase domain (Trypsin_2) Sobic.008G003200.1.p - similar to Putative fatty acid desaturase
WL17	S07_62928259	8.11E-06	0.08	0.046	Sobic.007G197700.1.p - similar to BHLH transcription factor(GBOF-1)-like Sobic.007G197800.1.p - similar to Os09g0502000 protein Sobic.007G197900.1 – no annotation Sobic.007G198000.1.p - similar to Os08g0524400 protein

Gene annotatons from Phytozome version 12.1 (2019).

All significant SNPs from WL2017 and the top ten significant SNPs from WL2013 including S10_60808604 and S04_67849488 were further evaluated using the integrated genome viewer (IGV; Robinson et al. 2011; Thorvaldsdóttir et al. 2013) The IGV was used to identify flanking genes that were approximately 15kb up and downstream of a significant SNP, related to the estimated genomewide LD decay of approximately 12kb in this study. Those genes and their annotations from phytozome (Phytozome 12, 2019, version 12.1; Goodstein et al. 2012, version 12.1) are listed in **Table 4.2**. The results from WL2013 show one significant SNP S10_60808604 on chromosome 10 that is approximately 113kb away from *Sobic.010G273800*, a 1,4-alpha-glucan branching enzyme I precursor (**Table 4.2**; Campbell et al. 2016). The SNP S10_60808604 is flanked within approximately 15kb of two genes not reported by Campbell et al. (2016), which could be involved in starch biosynthesis. Those two genes are *Sobic.010G274800* (approximately 9kb away) and *Sobic.010G275001* (approximately 15kb away), both are predicted in Phytozome (2019, Goodstein et al. 2012, version 12.1) as glycosyltransferases. As we estimated the genomewide LD decay of approximately 12kb but assumed to vary across chromosomes we created an LD plot for the specific chromosomal region. The LD plot created in TASSEL (Tassel 2018; Bradbury et al. 2007) shows that the SNP S10_60808604 (10s60808604) is in LD with the SNP 10s60831007 (**Figure 4.10**). The SNP 10s60831007 is a flanking marker approximately 8kb upstream of the gene *Sobic.010G275001* and might be in LD with the gene. However, *Sobic.010G275001* is flanked approximately 8kb downstream by another marker 10s60814316, which is not in LD with our significant SNP S10_60808604 (10s60808604) but in LD with the SNP 10s60831007. The significant SNP S10_60808604 is also not in LD with *Sobic.010G273800* (marker 10s60694303 in gene) and not in LD with both flanking markers of *Sobic.010G274800*. However, the SNP 10s60831007, which is in LD with our significant SNP, is also in LD with a

flanking marker (10s60803603) of *Sobic.010G274800* and the gene *Sobic.010G273800*. This indicates that the three genes might be explaining some of the ASV phenotypic variation and need further evaluation.

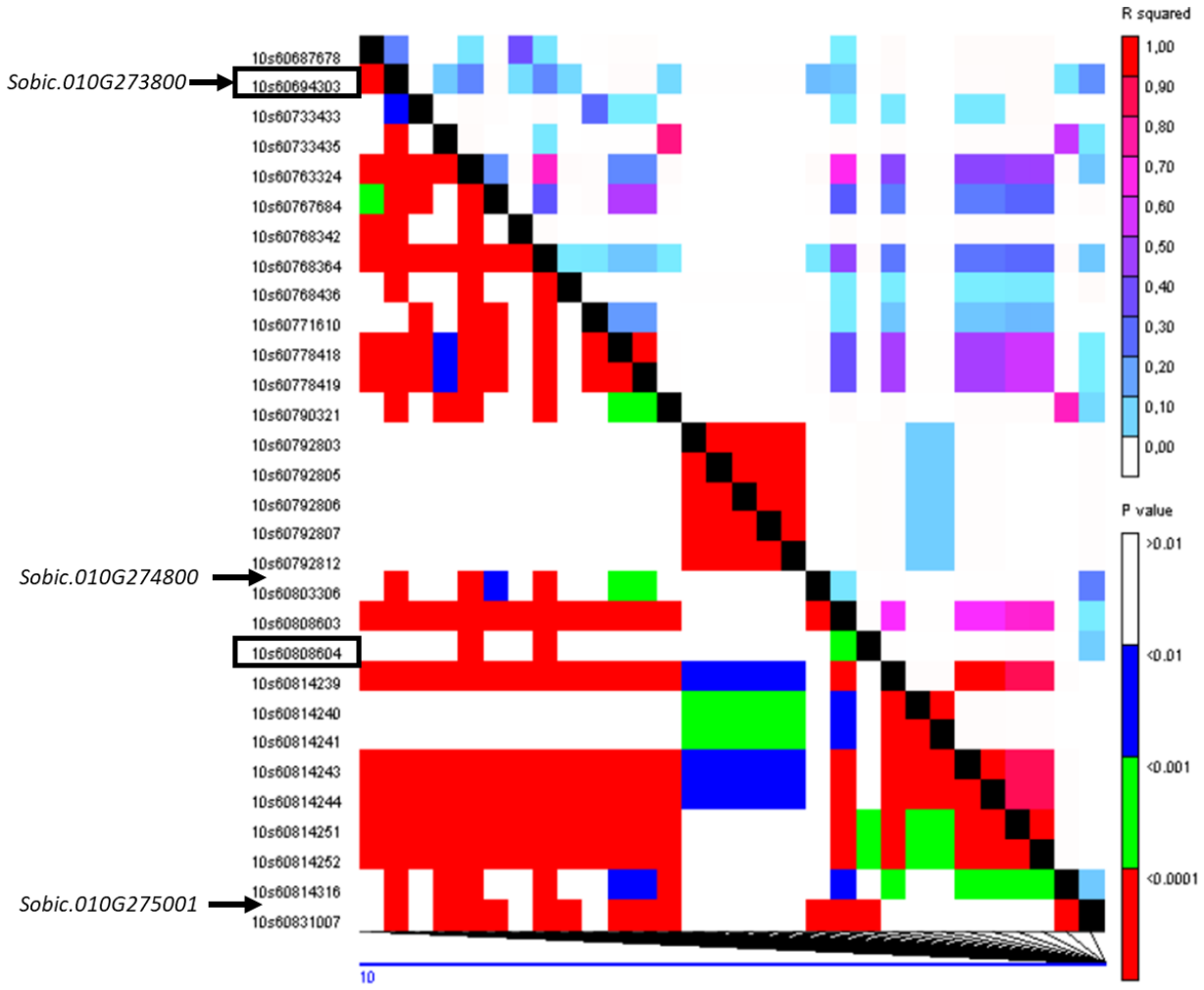


Figure 4.10. LD Plot for the significant SNP S10_60808604 and other GBS markers in close proximity to the genes *Sobic.010G273800, Sobic.010G274800 and Sobic.010G275001*.

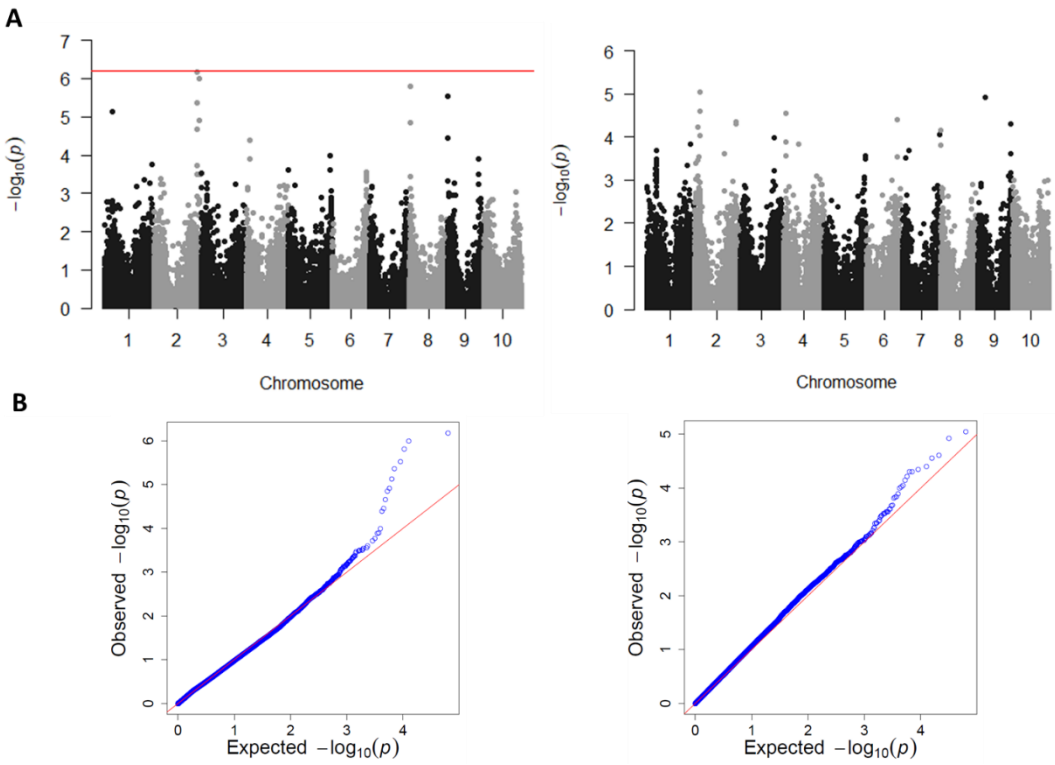


Figure 4.11. GWAS results for ASV for grain samples produced in WL13 (left) and WL17 (right) with (**A**) Manhattan plots and (**B**) QQ plots using a LMM.

A recent study reported that quantitative traits can be converted to binary traits for GWAS using LMM (Shenstone et al. 2018). The LMM was used in this study and the data fitted the model better than the MLM based on QQ plot results (**Figure 4.8, 4.11, B**). However, the LMM model for 2013 identified 10 SNPs using the FDR adjusted p-values at a threshold of 0.05 (**Table 4.3**). In 2017, no significant SNP was identified using the LMM. None of the SNPs from LMM were close (less than approximately 150kb) to a starch biosynthesis candidate gene (Campbell et al., 2016 table 1). The SNP results under LMM are different than from MLM. However, some SNPs significant in MLM WL2017, were not significant in the LMM model but appeared under the top 20 lowest p-values in the LMM (**Table 4.2, 4.3**). The integrated genome viewer (IGV; Robinson et al. 2011; Thorvaldsdóttir et al. 2013) was used to identify the genes approximately 15kb up and

downstream of a significant SNP. Those genes and their annotations from *Phytozome* (Phytozome version 12.1, 2019) are listed in **Table 4.3**.

Table 4.3. GWAS results from LMM with significant SNPs and candidate genes from WL13.

SNP	P value	FDR adj. P value	Candidate Gene Annotation
S02_71447794	6.68E-07	0.013	Sobic.002G349900.1.p - similar to Protein TOC75, chloroplast precursor Sobic.002G350000.1.p - similar to Putative uncharacterized protein Sobic.002G350100.2.p - (1 of 3) PTHR10562//PTHR10562:SF23 - SMALL UBIQUITIN-RELATED MODIFIER // SUBFAMILY NOT NAMED Sobic.002G350200.1.p - similar to Putative uncharacterized protein Sobic.002G350300.1.p - similar to Splicing factor 4-like protein Sobic.002G350400.1.p - similar to Tubulin alpha-2/alpha-4 chain
S02_71447799	4.27E-06	0.030	
S02_71447812	4.27E-06	0.030	
S02_73959322	1.02E-06	0.013	Sobic.002G384000.1.p - similar to Putative uncharacterized protein OJ1339_F05.143 Sobic.002G384150.1 - no annotation Sobic.002G384200.1 - no annotation Sobic.002G384300.1.p - similar to Myb-related protein-like Sobic.002G384400.1.p - similar to Thiazole biosynthetic enzyme 1-1, chloroplast precursor Sobic.002G384500.1.p - weakly similar to Putative uncharacterized protein OJ1448_G06.12 Sobic.002G384600.1 – no annotation
S02_73959309			
S02_73959315			
S02_73959316			
S08_2378161	1.54E-06	0.016	Sobic.008G026200.1.p - similar to Putative Xa1-like protein Sobic.008G026300.1 – no annotation Sobic.008G263400 – no annotation
S09_903513	2.97E-06	0.027	Sobic.009G009500.1.p - similar to MPI Sobic.009G009600.1.p - (1 of 39) PF00280 - Potato inhibitor I family (potato_inhibit)
S01_13279217	7.38E-06	0.047	Sobic.001G161200.1.p - similar to Putative serine/threonine protein kinase Sobic.001G161300.1.p - similar to THA4 Sobic.001G161400.1.p - similar to Putative uncharacterized protein Sobic.001G161500.1.p - similar to Auxin-responsive protein IAA12

Gene annotations from Phytozome version 12.1 (2019).

4.4.5 ASV is Expressed as a Recessive Trait

The four SC lines (SC489, SC491, SC587 and SC589) were crossed to ATx623 (cytoplasm sterile) and seeds from the crosses were evaluated to determine if the ASV is inherited as a dominant or recessive trait (**Table 4.4**). The crosses were screened for ASV with 32 seeds and 0 up to 2 seeds expressed the ASV phenotype. Thus, we assume that the ASV is recessive in the four lines studied.

Table 4.4. ASV is a recessive trait - The lines SC489, SC491, SC587 and SC589 were crossed to ATx623 and crosses were evaluated for ASV.

Environment	Pedigree	ASV+ counts of 32 seeds
WL2018	ATx623/BTx623	0*
WL2018	ATx623/BTx623	0*
WL2018	ATx623/BTx623	0*
WL2018	ATx623/SC489-1-1	1
WL2018	ATx623/SC489-2-1	1
WL2018	ATx623/SC489-2-2	0
WL2018	ATx623/SC489-4-1	1
WL2018	ATx623/SC489-4-2	0
WL2018	ATx623/SC489-5-1	0
WL2018	ATx623/SC489-5-1	0
WL2018	ATx623/SC489-5-2	0
WL2018	ATx623/SC489-6-1	1
WL2018	ATx623/SC491-2-1	2
WL2018	ATx623/SC491-3-1	0
WL2018	ATx623/SC587-1-1	0
WL2018	ATx623/SC587-2-1	0
WL2018	ATx623/SC587-2-1	0
WL2018	ATx623/SC587-3-1	0
WL2018	ATx623/SC587-4-1	0
WL2018	ATx623/SC587-4-1	1
WL2018	ATx623/SC587-4-2	0
WL2018	ATx623/SC589-1-1	0
WL2018	ATx623/SC589-1-1	0
WL2018	ATx623/SC589-2-1	0
WL2018	ATx623/SC589-4-1	1
WL2018	ATx623/SC589-5-1	0

ASV 1.8% KOH, 24h. *Griebel et al. 2019b

4.4.6 ASV Occurs Across Environments Exhibits Penetrance Issues

The four selected SC lines were evaluated for ASV in 2017 and 2018 (**Table 4.5**). These lines exhibited the ASV+ phenotype in both years; however, the homozygous lines show variable expression of the ASV+ phenotype in all 32 seeds tested. With less than 22 seeds of 32 seeds showing an ASV+ depending on the SC line, its panicle and the season. This provides evidence of variable penetrance for the ASV phenotype in the four lines making it difficult to score for ASV appearance if small sample sizes are used.

Table 4.5. Evaluation of the ASV phenotype of the genotypes SC489, SC491, SC587 and SC589 and controls in West Lafayette 2017 and 2018.

Environment	Pedigree	ASV+ count of 32 seeds	Environment	Pedigree	ASV+ count of 32 seeds
WL2017	TX623	0	WL2018	Macia	0
WL2017	TX623	0	WL2018	Macia	0
WL2017	Macia	0	WL2018	Sepon82	0
WL2017	Macia	0	WL2018	Sepon82	0
WL2017	Sepon82	0	WL2018	SC 489-1-B1	7
WL2017	Sepon82	0	WL2018	SC 489-2-B1	9
WL2017	SC 489-1	18	WL2018	SC 489-5-B1	18
WL2017	SC 489-2	22	WL2018	SC 489-6-B1	17
WL2017	SC 489-4	16	WL2018	SC 491-2-B1	10
WL2017	SC 491-1	12	WL2018	SC 491-3-B1	9
WL2017	SC 491-2	19	WL2018	SC 491-3-B1	9
WL2017	SC 491-3	20	WL2018	SC 491-5-B1	13
WL2017	SC 587-1	9	WL2018	SC 587-1-B1	22
WL2017	SC 587-2	8	WL2018	SC 587-2-B1	20
WL2017	SC 587-4	9	WL2018	SC 587-2-B1	12
WL2017	SC 589-1	11	WL2018	SC 587-3-B1	17
WL2017	SC 589-2	11	WL2018	SC 589-2-B1	14
WL2017	SC 589-5	10	WL2018	SC 589-5-B1	20
WL2018	TX623	0	WL2018	SC 589-3-B1	17
WL2018	TX623	0	WL2018	SC 589-3-B2	12

Samples used in starch analysis; ASV 1.8% KOH, 24h

4.4.7 Physiochemical Analysis

The lines SC489, SC491, SC587 and SC589 and controls exhibit total starch (db) values ranging between 70% and 80% and protein (db) between 9% and 14% in WL2013 and 2017, with values being slightly higher in 2017 (**Table 4.6**).

Table 4.6. NIR predicted moisture, starch, and protein contents of grain samples from sorghum landraces from two seasons WL2013 and WL2017

	WL2013					WL2017				
Pedigree	**Moisture**	**Starch (db)**	**Starch as is**	**Protein (db)**	**Protein as is**	**Moisture**	**Starch (db)**	**Starch as is**	**Protein (db)**	**Protein as is**
SC489	11.16 B	77.42 AB	68.78 A	10.46 AB	9.29 ABC	11.41 ABC	80.09 A	70.96 A	11.00 BC	9.74 BC
SC491	10.49 C	76.12 AB	68.14 A	9.31 B	8.34 BC	11.44 AB	79.40 AB	70.31 AB	12.13 B	10.74 B
SC587	14.29 A	79.77 A	68.37 A	9.30 B	7.97 C	11.48 AB	80.72 A	71.45 A	10.87 BC	9.62 BC
SC589	10.79 BC	77.76 AB	69.37 A	10.12 AB	9.03 ABC	11.64 A	77.73 BC	68.69 C	13.54 A	11.96 A
Macia	10.86 BC	72.15 B	64.32 A	11.02 A	9.82 A	11.33 ABC	76.61C	67.93 C	11.02 BC	9.77 BC
Sepon82	10.90 BC	72.70 B	64.78 A	10.90 AB	9.71 AB	11.22 BC	77.44 BC	68.76 BC	10.77 BC	9.56 BC
Tx623	10.77 BC	75.94 AB	67.77 A	10.87 AB	9.70 AB	11.08 C	79.25 AB	70.47 AB	9.95 C	8.85 C

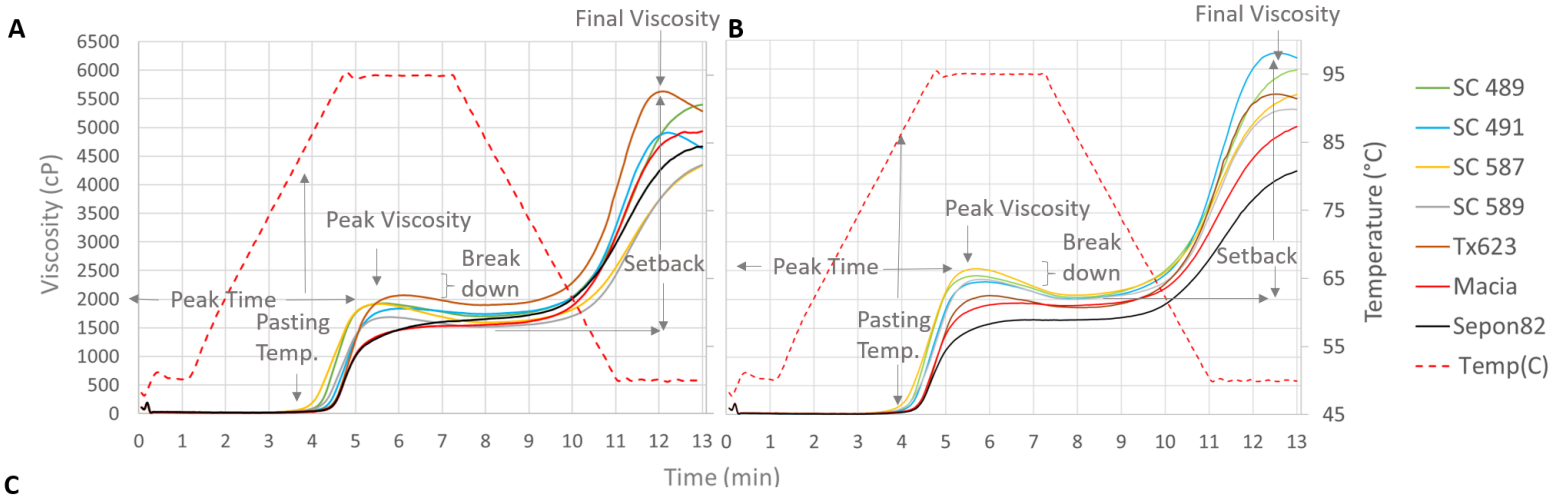
WL2013 only one biol. rep in duplicate

The physiochemical analysis of amylose, starch thermal properties and paste viscosity profiles demonstrated G×E interaction **(Figure 4.12**). The amylose values of the four SC-lines varied between approximately 20% and 24% being slightly lower than controls in WL2013, 2017 and 2018 (**Table 4.7, Figure 4.12**). Significant differences for amylose were observed only for SC489, SC587 and SC589 in the season WL2017 in comparison to Macia and for SC491 and SC589 in WL2018 in comparison to BTx623.

Table 4.7. Evaluation of lines from WL2013 for amylose content and starch thermal properties.

Pedigree	**Amylose [%]**		**To [°C]**		**Tp [°C]**		**Tc [°C]**		**TcminusTo [°C]**		**Enthalpy [J/g]**		**Melting_P [°C]**	
SC489	23.7	A	61.7	B	68.7	C	80.7	B	19.0	AB	3.7	BC	96.1	B
SC491	23.8	A	62.0	B	69.1	B	79.7	B	17.7	CB	2.7	C	96.7	AB
SC587	23.1	A	61.0	B	69.0	CB	81.3	B	20.4	A	3.9	BC	96.7	AB
SC589	21.8	A	59.2	C	68.2	D	79.7	B	20.5	A	4.2	B	95.9	B
BTx623	25.3	A	67.9	A	72.6	A	84.5	A	16.6	C	6.0	A	98.0	A

Mean of laboratory reps, one biol rep per genotype



C

Environment	Pedigree	Peak 1 [cP]	Trough 1 [cP]	Breakdown [cP]	Final Viscosity [cP]	Setback [cP]	Peak Time [min]	Pasting Temp [°C]	Amylose [%]	To [°C]	Tp [°C]	Tc [°C]	Tc-To [°C]	Enthalpy [J/g]	Melting Point [°C]
WL17	Macia	1535.50 A	1472.25 A	63.25 A	4938.25 AB	3466.00 AB	6.90 D	90.14 C	27.15 C	63.96 A	70.88 A	83.77 A	19.82 BC	6.53 AB	101.11 AB
WL17	SC489	1947.78 BC	1704.67 ABC	239.77 AB	5408.97 B	3698.24 B	5.70 AB	86.34 B	23.40 AB	65.02 ABC	70.72 A	83.59 A	18.55 AB	5.95 A	96.47 A
WL17	SC491	1867.00 ABC	1755.23 BC	111.17 A	4950.62 AB	3196.79 AB	5.98 BC	88.08 BC	22.28 A	65.89 BCD	71.68 A	86.87 B	20.99 C	6.85 AB	99.25 AB
WL17	SC587	1968.36 BC	1618.43 AB	340.67 B	4365.68 A	2739.89 A	5.55 A	83.54 A	22.85 AB	66.04 CD	71.05 A	84.79 A	18.71 AB	7.04 AB	100.60 AB
WL17	SC589	1693.02 AB	1505.07 A	182.65 AB	4386.31 A	2825.14 A	5.76 AB	86.98 B	23.32 AB	66.43 D	71.14 A	83.82 A	17.38 A	5.96 A	101.24 B
WL17	Sepon82	1604.00 A	1478.75 A	125.25 AB	4667.00 AB	3188.25 AB	6.98 D	90.35 C	25.45 BC	64.66 AB	70.51 A	82.90 A	18.25 AB	6.87 AB	100.45 AB
WL17	Tx623	2075.00 C	1898.25 C	176.75 AB	5288.25 AB	3390.00 AB	6.17 C	89.36 C	23.70 AB	65.39 BCD	71.11 A	83.59 A	18.20 AB	7.99 B	98.86 AB
WL18	Macia	1943.00 AB	1878.00 AB	65.00 A	5002.75 AB	3124.75 AB	6.70 B	88.43 CD	23.28 AB	67.26 B	72.11 B	86.04 C	18.78 A	8.66 AB	99.18 A
WL18	SC489	2415.00 D	2036.38 B	378.63 BC	5989.62 CD	3953.25 CD	5.67 A	85.75 B	23.01 AB	64.28 A	69.12 A	83.84 AB	19.56 AB	8.73 AB	98.51 A
WL18	SC491	2308.88 CD	2006.88 B	302.00 BC	6196.50 D	4189.63 D	5.87 A	86.36 BC	22.09 A	64.70 A	69.33 A	85.68 C	20.97 C	9.20 AB	99.34 A
WL18	SC587	2537.50 D	2068.38 B	469.13 C	5555.12 BCD	3486.75 BC	5.70 A	83.73 A	22.70 AB	64.19 A	68.98 A	84.92 ABC	20.73 BC	9.50 B	99.24 A
WL18	SC589	2355.38 CD	1998.75 B	356.63 BC	5294.50 BC	3295.75 ABC	5.83 A	85.21 AB	21.48 A	64.75 A	69.41 A	83.55 A	18.80 A	8.00 A	100.09 A
WL18	Sepon82	1651.25 A	1587.00 A	64.25 A	4230.50 A	2643.50 A	6.92 B	88.81 D	22.71 AB	66.99 B	71.75 B	85.55 BC	18.56 A	8.67 AB	98.44 A
WL18	Tx623	2067.75 BC	1856.25 AB	211.50 AB	5483.00 BCD	3626.75 BCD	5.98 A	89.03 D	24.79 B	68.61 C	72.35 B	86.57 C	17.96 A	8.75 AB	98.58 A

S-means; Values followed by the same letters (A-D) in the same column are not significantly different.

Figure 4.12. Starch quality characteristics of natural variation in WL2017 and WL2018 for amylose content, starch thermal properties, paste viscosity profiles.

The starch thermal properties showed significantly lower starch GT in comparison to controls for onset (T_o), peak (T_p) and completion (T_c) GT in West Lafayette 2013 and for onset (T_o), peak (T_p) in West Lafayette 2018 (**Table 4.7, Figure 4.12**).

The evaluation of paste viscosity behavior in West Lafayette 2017 and 2018 showed similar profiles of the four SC-lines in comparison to the controls (**Figure 4.12**). However, the controls, Macia and Sepon82, behave different than Tx623 and the SC lines and show no existing breakdown in both seasons and lower final viscosity in 2018.

4.5 Discussion

The sorghum diversity panel (SDP) used in this study with more than 700 sorghum conversion lines is an excellent source of genetic diversity (Stephens et al. 1967; Rosenow et al. 1997a, 1997b). The SDP is larger than the commonly used Sorghum Association Panel (SAP) (Casa et al., 2008, Morris et al. 2013; Cuevas et al. 2017; Boyles et al. 2017; Adeyanju et al. 2015; Mace et al. 2013; Sukumaran et al. 2012; Shenstone et al. 2018). The race and regional classification from our principal component analysis confirmed previously reported results on the different sorghum races and geographic origins. The SDP represented durras mainly from India and durra-bicolors mostly from Eastern Africa, guineas mainly from Western Africa, the kafirs from Eastern and Southern Africa, and caudatums from a diverse set or origins on the African continent not clearly from Central or Eastern Africa as reported previously (Morris et al. 2013; Billot et al. 2013; Sukumaran et al. 2012; Harlan and de Wet 1972, Dahlberg 2000). The bicolors occurred as reported previously with no clear geographical classification (Morris et al. 2013; Deu et al. 2006; Brenton et al. 2016; Billot et al. 2013).

We identified 13 SC lines exhibiting an ASV+ phenotype across two environments (**Figure 4.6**). The PCA helped to confirm that 11 of those SC lines cluster together as durra types from the

Nandyal working group originating from India (**Figure 4.3, 4.4**). The other lines are one guinea from Western Africa and second a line of no clear race nor geographic classification.

The race durra is reported as having smaller, ovate panicles, exhibit little hairs and are very stiff and densely packed in comparison to the bicolors (**Figure 4.6;** Dahlberg 2000). The durra race is sub divided into the working groups 50, 51 and 52 (Dahlberg 2000). The working group 51, known as 46(1) from Murty and Govil's classification, describes durra types named Nandyal (Dahlberg 2000). We observed for the Nandyal types similar characteristics as reported previously, such as dense packed panicles, creamy to yellow in seed color, a strong glutinous and pearly endosperm and medium to tall plant size (Dahlberg 2000). The Nandyal types are reported as short season durras (Harlan and de Wet 1972) producing bold, glutinous grains (Golomb et al. 1976).

Glutinous seeds are also called waxy (Martin and Jenkins 1950). An earlier study by Waniska (1976) was able to distinguish between waxy and non-waxy sorghum genotypes by treating the whole, pearled sorghum kernels with NaOH. Glutinous rice is desired in many Asian cultures and countries (Roder et al. 1996) as it becomes sticky during cooking (Golomb et al. 1976). Glutinous rice is known for a unique amylopectin structure and considered as a recessive variant of the non-glutinous varieties (Golomb et al. 1976). Thus, the amylopectin structure should be further evaluated in the 11 sorghum SC lines in this study.

Glutinous rice becomes soft and sticky after cooking (Wang et al. 2007; Golomb 1976; Martin and Jenkins 1950), little sweet (Golomb et al. 1976) and tender, glossy when eating, which is a result of low amylose content, low starch GT (Wang et al. 2007) and high gel consistency (Wang et al. 2007; Golomb et al. 1976). The four selected sorghum genotypes SC489, SC491, SC587 and SC589 exhibit similar physiochemical behaviours with low amylose content, high gel consistency and in some seasons a slightly lower starch GT (**Table 4.7, Figure 4.12**). Maybe the

identified sorghums are suitable, such as glutinous rice, for sweet dishes and alcoholic beverages and less suitable for non-sticky and dishes where less viscosity is desired (Golomb et al. 1976). In the 1950s, the sorghum glutinous/waxy product Minute Dessert was known in the USA (Martin and Jenkins 1950). Glutinous/waxy starches are suitable for soft puddings and industrial adhesive uses (Martin and Jenkins 1950).

The natural variation in this study varies in several SNPs compared to Tx623, why several starch biosynthesis genes could be involved in determining ASV and the genotypes physiochemical profile (**Table 4.1**). The 107 SNPs in common in the four lines are not necessarily due to the ASV+ phenotype but due to the close relatedness of the four lines from the same working group and their clear non-relatedness to Tx623 (**Figure 4.6, C**). However, from the 107 SNPs 59 are in starch branching genes, one being *SBEIIb* with 33 SNPs reported previously as a key gene causing ASV in sorghum (Griebel et al. 2019). Several genes control amylose content and gel consistency (GC) in rice (Tian et al. 2009; Gao et al. 2011). Genes controlling amylose content in rice include *Wx, ALK, SSIII-2, SSI, PUL* and for gel consistency *ISA, SBE3* (Tian et al. 2009). In sorghum, we observed that ASV, amylose and starch GT are influenced by a *SSIIa* and *SBEIIb* (Griebel et al. 2019b). The physiochemical behaviour of the four SC lines in the current study is similar to our previous study in an EMS population, where low amylose content, high gel consistency and lower starch GT was observed in sorghum *ssIIa* mutants (Griebel et al. 2019b). The previously reported information on genetics and physiochemical behaviour could help to narrow down the putative candidate genes in the current study and may point to a starch synthase II that causes the ASV variation in our sorghum diversity panel. The ASV has been shown to be controlled by the *ALK* gene, a starch synthase (SS) II, in rice (Gao et al. 2011; Wang et al. 2007; Tian et al. 2009) and by a SSIIa and SBEIIb in sorghum (Griebel et al. 2019b).

For further evaluation, we conducted a GWAS on the entire sorghum diversity panel in 2013 and 2017. The current study describes the first GWAS in sorghum using the ASV phenotype. Results of GWAS for ASV in sorghum were similar to studies in rice with poor trait correlations and inconsistent performance across years (Zhao et al. 2011). The LD decay in rice is slower (Huang et al. 2010) than in sorghum with 1kb up to 20kb (**Figure 4.7**; Morris et al. 2013; Mace et al. 2013). However, some candidate genes in rice were close to the *ALKK (SSII-3)* gene (Huang et al. 2010; Zhao et al. 2011) and the *SSII-2* (Zhao et al. 2011). In this study the GWAS results do not show a significant SNP in close proximity to *Sobic.004G163700*, a *SBEIIb* and *Sobic.010G093400*, a *SSIIa* reported as one of the causal genes for ASV in sorghum (Griebel et al. 2019b).

GWAS was conducted using MLM and LMM. Since the data were collected using binary scores, the LMM may provide a better control for type 1 error rate (Shenstone et al. 2018; Chen et al. 2016). The MLM model identified three candidate genes for ASV. The SNP S10_60687678 was 9kb away from *Sobic.010G274800*, 15kb away from *Sobic.010G275001*, and 113kb away from *Sobic.010G273800* (**Table 4.3**). *Sobic.010G273800 is* annotated as a 1,4-alpha glucan branching enzyme I precursor (**Table 4.3)** and a known gene involved in starch biosynthesis in the amyloplast (Campbell et al. 2016). The other two genes not reported by Campbell et al. (2016) are annotated as glycosyltransferases (Phytozome 12, 2019, version 12.1). These could be putative new sorghum genes involved in starch biosynthesis and are worth further evaluation. The common enzymes involved in starch biosynthesis are starch synthases, which are ADP-glucose-dependent (alpha 1,4-glucan-4-glucosyl transferases; Boyer 1985) transferases, and starch branching enzymes (SBEs) being 1,4-alpha-glucan 6-glucosyl transferases (Tetlow and Emes, 2014; Boyer 1985).

The LD decay in this study was estimated as approximately 12kb but is assumed to vary across the genome. Therefore, an LD plot of a significnat SNP and GBS marker in close proximity to candidate genes was created. The LD plot showed that the SNP 10s60831007, which is in LD with our significant SNP S10_60808604, is also in LD with a flanking marker 10s60803603 of *Sobic.010G274800* and the gene *Sobic.010G273800*. Therefore, these two genes predicted as glycosyltransferases could be in LD with a flanking marker and thus, be considered valuable candidate genes that might explain some of the ASV phenotypes. Furthermore, these two candidate genes if in LD with the flanking marker would be also in LD with *Sobic.010G273800* a known starch branching enzyme I precursor involved in starch biosynthesis. This gene, *Sobic.010G273800*, showed sequence similarity with the gene *Sobic.004G163700* (Phytozome 12 2019, version 12.1), which is reported as a candidate gene for ASV in an sorghum EMS Population (Griebel et al. 2019b). It was also previously reported that the gene *Sobic.010G273800* was the best hit on sequence similarity with maize amylose extender (Griebel et al. 2019b).

4.6 Conclusion

The ASV phenotype used in Griebel et al. (2019) is expressed in the standing variation of sorghum. The samples that exhibit stable ASV across seasons are mainly *durra* types from the working group Nandyal described in previous studies as glutinous varieties. The four selected SC lines exhibit normal amylose content, high GC and environmentally dependent lower starch GT. This makes the lines valuable for the food industry for product development with different levels of stickiness and viscosity. GWAS showed significant associations for one SNP in proximity to a starch biosynthesis gene *Sobic.010G273800* and two candidate genes *Sobic.010G274800* and *Sobic.010G275001* described as glucosyltransferases. Whole genome resequencing (WGS) revealed several SNPs in genes in starch biosynthesis which need to be further evaluated.

Therefore, we developed bi-parental populations for the four SC lines with BTx623, which are ready for evaluation and to be further developed to RIL populations. Additionally, the GWAS could be repeated in more seasons with increased repeatability to screen more than 20 seeds per genotype to account for penetrance and ensure to catch the presence and absence of the ASV phenotype truly.

4.7 Acknowledgments/Funding Sources

This research was 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). The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. Program activities are funded by the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-A-13-00047.

4.8 Material and Methods

4.8.1 Plant Material

A sorghum diversity panel of approximately 840 lines (SC lines and breeding lines) was grown in West Lafayette in 2013 and 2017. The panicles were self-pollinated by covering the panicles with paper bags before anthesis, bulked and screened for the ASV phenotype (Griebel et al. 2019) in a binary manner. Four candidate lines (SC489, SC491, SC587 and SC589) were selected based on ASV+ phenotypes and increased in West Lafayette 2017 and 2018. Three biological replicates per line used for this study in WL17, while four biological replicates were tested in WL2018. Tx623, Sepon82 and Macia were used as controls in both seasons with two biological replicates each.

4.8.2 Phenotypic Data - Alkali Spreading Test

The alkali spreading test was conducted as described in Griebel et al. (2019) with 1.8% KOH for 24h. Each line was evaluated with 2 seeds in West Lafayette 2013 (preliminary study) and with 8 seeds in West Lafayette 2017. A positive ASV phenotype was scored as 1, no visibility of ASV similar to Tx623, Sepon82 and Macia, as 0. The selected four SC lines and their crosses from genetic studies were screened for ASV with 32 seeds each.

4.8.3 ASV - Mode of Inheritance for four selected SC lines

To study whether the ASV is a recessive trait, each of the four lines was crossed to ATx623 in West Lafayette 2018 to produce F1 hybrid seeds. The F1 seeds were scored for the ASV phenotype on 32 seeds.

4.8.4 CTAB DNA Extraction

DNA was extracted from individual plants using a CTAB extraction protocol as described in Griebel et al. (2019b).

4.8.5 Next Generation Sequencing (NGS) and analyses

Next generation sequencing analysis including reads filtering contaminants cleaning, the reference genome read mapping approach and SNP calling using GATK was conducted as described in Griebel et al. (2019b).

4.8.6 Physiochemical Analysis

The flour preparation, starch extraction and starch thermal properties using differential scanning calorimetry (DSC) were conducted as described in Griebel et al. (2019).

The amylose analysis, moisture determination and paste viscosity analysis followed the methodology described in Griebel et al. (2019b).

4.8.7 Proximate Analysis NIR

Proximate Analysis was conducted on whole sorghum seeds from PR16/17 and WL17 samples using the NIR of Perten (Model DA 7250, Serial Number 1211581) running the program calibration “PURDUE Whole Sorghum (Mirror Cup)”. The parameters reported are protein db, moisture, starch (db), starch as is. For each seed sample, two replications were conducted.

4.8.8 Statistical Analysis for Starch Quality Traits

Linear mixed model analysis of variance (ANOVA) with the Tukey-Kramer multiple comparison adjustment was performed in SAS 9.4 using the GLIMMIX procedure to compare GT properties across genotypes and environments as described in Griebel et al. (2019b).

4.8.9 Statistical Analyses for GWAS

Phenotypic Data

The phenotypic data were prepared as binary data with 0 = ASV- and 1=ASV+. Trait correlation was calculated using SAS 9.4 and the proc corr procedure.

Principal Component Analysis

The race information and country of origin for the Sorghum Conversion Program were kindly provided by Robert R. Klein, USDA ARS (personal communication, 2019) and missing information were filled using the database USDA ARS GRIN (2019). The countries of origin were grouped by continent for the minority of lines from Asia and the Americas and kept a single group for India. The majority of the lines are from the African continent why those countries were

grouped into regions based on UN classifications (UN, 2019). The non-SC lines were treated as breeding lines and were not given a race or country name. The population structure and kinship analysis was conducted using GAPIT (based on prcomp() from R, personal communication from Alex Lipka) with 5 PCs. The PCs were plotted using R v. 3.5.1 and the package "ggfortify" (R core Team, 2013; Tang et al. 2016; Horikoshi and Tang, 2016).

Genotypic Data

The imputed genotype-by-sequencing data were kindly provided by Dr. Patrick Brown and raw data are available at University of Illinois database, https://doi.org/10.13012/B2IDB-7570206_V1 (Brown, personal communication, 2019; Brown, 2017). The GBS data were aligned to the sorghum reference genome version 3 from Phytozome11 (Goodstein et al. 2012, version 11) and imputed using the Beagle 4 software (Brown, personal communication, 2019). From the total set of markers, 80103 SNP markers at a minor allele frequency (MAF) 0.025 were selected and used in the genome wide association study.

Genome Wide Association Analysis

The binary data from WL2013 and WL2017 were used as input phenotypes for the GWAS using a mixed linear model (MLM) (Lipka et al. 2012) and a logistics mixed model (LMM) (Shenstone et al. 2018) with 80103 SNP markers.

Mixed Linear Model

The Genome Association and Prediction Integrated Tool (GAPIT) (Zhang et al. 2010; Lipka et al. 2012) was used for the mixed linear model (MLM). The PCs were calculated using GAPIT and set to 5, due to the five race classification in Sorghum. The MLM in GAPIT run under the default settings and the Van Raden kinship algorithm.

Logistic Mixed Model

The logistic mixed model (LMM) was conducted as described in Shenstone et al. (2018) but with 5 principal components. The significant SNPs were determined using bonferroni (Bonferroni 1936) threshold of 6.242E-07 and at an FDR (Benjamini and Hochberg 1995) of adjusted p-values at 0.05.

Manhattan Plot

The manhattan plots were created using the GWAS results and the R package "qqman" (Turner 2014; R Core Team, 2013).

Linkage Disequilibrium (LD) Plot

The LD plot for the significant SNPs and with adjacent SNPs to candidate genes was created in TASSEL 5.2.51 (Tassel 2019; Bradbury et al. 2007).

4.8.10 Neighbour-Joining Tree

The neighbour joining tree was created in TASSEL 5.2.51 (Tassel 2019; Bradbury et al. 2007).

4.9 References

Adeyanju A, Little C, Yu J, Tesso T (2015) Genome-wide association study on resistance to stalk rot disease in grain sorghum. *G3* 5:1165-1175.

Benjamini Y., and Y. Hochberg. 1995. Controlling the false discovery rate:a practical and powerful approach to multiple testing. J. R. Stat. Soc., B 57:289–300.

Billot C, Ramu P, Bouchet S, Chantereau J, Deu M, Gardes L, Noyer J-L, Rami J-F, Rivallan R, Li Y, Lu P, Wang T, Folkertsma RT, Arnaud E, Updadhyaya HD, Glaszman J-C, Hash CT (2013) Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLOS ONE* 8:1-16.

Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilit `a. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3–62

Boyer CD (1985) Soluble starch synthases and starch branching enzymes from developing seeds of sorghum. *Phytochemistry* 24:15-18.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23:2633–2635.

Brenton ZW, Cooper EA, Myers MT, Boyles RE, Shakoor N, Zielinski KJ, Rauh BL, Bridges WC, Morris GP, Kresovich S (2016) A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics* 204:21-33.

Brown PJ (2019) University of California Davis, personal communication. Provided imputed GBS data file on sorghum reference genome Tx623 version 3.

Brown PJ (2017) Sorghum bicolor GBS: diversity study II. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-7570206_V1

Boyles RE, Pfeiffer BK, Cooper EA, Rauh BL, Zielinski KJ, Myers MT, Brenton Z, Rooney WL, Kresovich S (2017) Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theor Appl Genet* 130:697-716.

Campbell BC, Gilding EK, Mace ES, Tai S, Tao Y, Prentis PJ, Thomelin P, Jordan DR, Godwin ID (2016) Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol J* 14:2240-2253.

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. *Crop Science* 48:30-40.

Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celedon JC, Redline S, Papanicolaou GJ, Thornton TA, Laurie CC, Rice Km Lin X (2016) Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet* 98, 653-666.

Cuevas HE, Rosa-Valentin G, Hayes CM, Rooney WL, Hoffmann L (2017) Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics* 18:108 1-17.

Dahlberg JA (2000) Classification and characterization of sorghum. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 99-130.

Deu M, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorhgums using a core collection. *Genome* 49:168-180.

FAO, 2019, http://www.fao.org/land-water/databases-and-software/crop-information/sorghum/en/). FAOSTAT, 2019, http://www.fao.org/faostat/en/) (Jan. 2019).

FAOSTAT, 2016, FAOSTAT Database. http://www.fao.org/faostat/en/ (Oct. 2016).

Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y,Yan M, Jiang H, Dong G, Huang Y, Han B, Li J, Qian Q (2011) *ALK*, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. *J Integr Plant Biol* 53: 756-765.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putman N, Rokhsar DS (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res 40* (Database issue), D1178–86. doi:10.1093/nar/gkr944.

Golomb L (1976) The origin, spread and persistence of glutinous rice as a staple crop in mainland Southeast Asia. *Journal of Southeast Asian Studies* 7:1-15.

Griebel S, Webb MM, Campanella OH, Craig BA, Weil CF, Tuinstra MR (2019) The alkali spreading phenotype in *Sorghum bicolor* and its relationship to starch gelatinization. *J Cereal Sci* 86:41-47.

Griebel S, Westerman RP, Adeyanju A, Addo-Quaye C, Craig BA, Weil CF, Cunningham M S, Bhavesh P, Campanella OH, Tuinstra MR (2019b) Mutations in sorghum *SBEIIb* and *SSIIa* affect alkali spreading value, starch composition, thermal properties and flour viscosity. *In Review. Theor Appl Genet*.

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. *Crop Science* 12:172-176.

Hayes CM, Burow GB, Brown PJ, Thurber C, Xin Z, Burke JJ (2015) Natural variation in synthesis and catabolism genes influences dhurrin content in sorghum. *The Plant Genome* 8:1-9.

Horikoshi M. and Y. Tang. 2016. ggfortify: Data Visualization Tools for Statistical Analysis Results. https://CRAN.R-project.org/package=ggfortify

House LR, Gomez M, Sun Y, Murty DS, Verma BN (2000) Development of some agricultural industries in several African and Asian countries. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 131-190.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Wenig Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42:961-969.

Huang X, ZhaoY, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44:32-41.

Kimber CT (2000) Origins of domesticated *Sorghum* and its early diffusion to India and China. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 3-98.

Kimber CT, Dahlberg JA, Kresovich S (2013) The gene pool of *Sorghum bicolor* and its improvement. *Genomics of the Saccharinae, Plant Genetics and Genomics: Crops and Models 11, ed Paterson AH* (Springer New York), pp 23 – 41.

Klein RR (2019) USDA ARS personal communication. Provided data of race, country of origin and PI numbers for sorghum conversion lines.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and predition integrated tool. *Bioinformatics* 28:2397-2399.

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa´s indigenous cereal crop sorghum. *Nat. commun.* 1-9 DOI: 10.1038/ncomms3320.

Mapchart.net 2019. https://mapchart.net/. Accessed 2019.

Martin JH, Jenkins MT (1950) New uses for waxy-cereal starches. *Yearbook of Agriculture 1950-1951* 159-162.

Mogga M, Sibiya J, Shimelis H, Lamo J, Yao N (2018) Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (Oryza sativa L.) using DArT markers. *PLOS ONE* 1-19 doi.org/10.1371/journal.pone.0198012.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum, *PNAS* 110:453-458.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J and Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551-556.

Phytozome 11. 2016. http://www.phytozome.net, https://phytozome.jgi.doe.gov/ pz/portal.html to download version 11. Accessed 2016.

Phytozome 12. 2019. Version 12.1. http://www.phytozome.net, accessed 2019.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.Rproject.org/.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative Genomics Viewer. Nature Biotechnology 29:24-26.

Roder W, Keoboulapha B, Vannalath K, Phouaravanh B (1996) Glutinous rice and its importance for hill farmers in Laos. *Economic Botany* 50:401-408.

Rosenow DT, Dahlberg JA, Stephens JC, Miller FR, Barnes DK, Peterson GC, Johnson JW, Schertz KF (1997a) Registration of 63 converted sorghum germplasm lines from the sorghum conversion program. *Crop Science* 37:1399-1400.

Rosenow DT, Dahlberg JA, Peterson GC, Clark LE, Miller FR, Sotomayor-Rios A, Hamburger AJ, Madera-Torres P, Quiles-Belen A, Woodfin CA (1997b). Registration of fifty converted sorghums from the sorghum conversion program. *Crop Science* 37:1397-1398.

Shenstone E, Cooper J, Rice B, Bohn M, Jamann TM, Lipka AE (2018) An assessment of the performance of the logistic mixed model for analyzing binary traits in maize and sorghum diversity panels. *PLOS ONE* 1-20 doi.org/10.1371/journal.pone.0207752.

Song J-M, Arif M, Zhang M, Sze S-H, Zhang H-B (2019) Phenotypic and molecular dissection of grain quality using a USDA rice mini-core collection. *Food Chemistry* 284:312-322.

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghum to early combine genotypes. *Crop Science* 7:396.

Sukumaran S, Xiang W, Bean SR, Pedersen JF, Kresovich S, Tuinstra MR, Tesso TT, Hamblin MT, Yu J (2012) Association mapping for grain quality in a diverse sorghum collection. *The Plant Genome* 5:126-135.

Tang Y., M. Horikoshi, W. Li. 2016. "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." The R Journal 8.2 2016: 478-489.

TASSEL (2018) Trait Analysis by aSSociation Evolution and Linkage (TASSEL) version 5.2.51 (https://www.maizegenetics.net/tassel).

Taylor J, Dewar J (2000) Fermented products: Beverages and porridges. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA (John Wiley & Sons, Inc., New York), pp. 751-796.

Tetlow IJ, Emes MJ (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life* 66:546-558.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14:178-192.

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M, Li J (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *PNAS* 106:21760-21765.

Turner, S.D. (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhatta n plots. biorXiv DOI: 10.1101/005165

UN. https:// unstats.un.org/unsd/methodology /m49/. Accessed 2019.

USDA ARS (2019). United States Department of Agriculture (USDA), Agriculture Research Service (ARS), Germplasm Resources Information Network (GRIN). Database https://www.ars-grin.gov/npgs/aboutgrin.html (Accessed 2019).

Wang LQ, Liu WJ, Xu Y, He YQ, Luo LJ, Xing YZ, Xu CG, Zhang Q (2007) Genetic basis of 17 traits and viscosity parameters characterizing the eating and cooking quality of rice grain. *Theor Appl Genet* 115:463-476.

Wang Y-H, Upadhyaya HD, Burrell AM, Sahraeian SME, Klein RR, Klein PE (2013) Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, *sorghum bicolor*. G3 3:783-793

Waniska RD (1976) Methods to assess quality of boiled sorghum, gruel and chapaties from sorghum with different kernel characteristics. *Master thesis*. Texas A&M University, College Station, Texas, USA.

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42:355-360.

Zhao K, Tung C-W , Eizenga GC , Wright MH , Liakat Ali M , Price AH , Norton GJ, Rafiqul Islam M , Reynolds A , Mezey J , McClung AM , Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat. Commun,* 1- 10 DOI: 10.1038/ncomms1467.

4.10 Supplemental Information

Table S4.1: Genotypes used in PCA by race, country and ASV (Race, PI name, country: Klein, personal communication, 2019).

Taxa and SC code	PI name	Race	WL13	WL17	Country	Regions	PC1	PC2	PC3	PC4	PC5
Ajabsido	PI656015	Breeding-line	0	0	Breeding-line	Breeding-line	-21.875	35.080	-18.876	3.632	-16.180
BOK11	PI656002	Breeding-line	0	0	Breeding-line	Breeding-line	-1.593	-10.010	135.223	-12.457	-38.618
BQL41	PI656063	Breeding-line	0	0	Breeding-line	Breeding-line	-4.560	7.637	54.734	-5.500	-24.400
BTx2752	PI656018	Breeding-line	0	0	Breeding-line	Breeding-line	-5.893	-0.766	57.806	-8.203	-17.913
BTx3042	PI656019	Breeding-line	0	0	Breeding-line	Breeding-line	18.511	-3.942	90.199	-11.390	-66.096
BTx3197	PI655992	Breeding-line	0	0	Breeding-line	Breeding-line	-1.500	-8.957	140.438	-14.179	-43.557
BTx378	PI655991	Breeding-line	0	0	Breeding-line	Breeding-line	6.088	-8.106	131.600	-14.530	-47.273
BTx399	PI655993	Breeding-line	0	0	Breeding-line	Breeding-line	-24.450	27.740	42.397	-4.621	-42.785
BTx615	PI656022	Breeding-line	0	0	Breeding-line	Breeding-line	-12.770	4.841	15.165	-5.301	-10.685
BTx623	PI564163	Breeding-line	0	0	Breeding-line	Breeding-line	-34.361	28.206	60.357	-11.062	-35.668
BTx641	PI642793	Breeding-line	0	0	Breeding-line	Breeding-line	-13.801	15.760	50.777	-6.564	-47.616
BTx642	PI656029	Breeding-line	0	0	Breeding-line	Breeding-line	47.733	15.392	12.306	7.792	-57.738
BTx643	NA	Breeding-line	1	0	Breeding-line	Breeding-line	-27.834	20.802	47.308	-8.987	-35.476
BTx645	NA	Breeding-line	0	0	Breeding-line	Breeding-line	-20.711	23.095	53.893	-7.579	-47.704
BTxARG-1	PI561072	Breeding-line	0	0	Breeding-line	Breeding-line	-0.815	-9.756	132.639	-11.346	-39.101
Caprock	PI655986	Breeding-line	0	NA	Breeding-line	Breeding-line	24.824	2.449	70.938	-9.858	-66.461
Comb7078	PI655990	Breeding-line	0	0	Breeding-line	Breeding-line	44.027	12.897	23.145	-3.674	-68.557
Day	PI641874	Breeding-line	0	0	Breeding-line	Breeding-line	34.708	12.355	29.174	-1.129	-58.813
Dorado	PI656034	Breeding-line	0	0	Breeding-line	Breeding-line	-72.129	66.900	-14.701	-3.445	-9.518
DwfYellMilo	NA	Breeding-line	0	0	Breeding-line	Breeding-line	54.617	21.040	-17.112	3.388	-70.371
ElMota	PI656035	Breeding-line	0	0	Breeding-line	Breeding-line	-53.617	43.044	-19.484	-1.055	-3.988
FeteritaGishesh	PI152651	Breeding-line	0	0	Breeding-line	Breeding-line	-62.425	47.495	-12.976	-0.577	-14.020
HEGARI	PI34911	Breeding-line	0	0	Breeding-line	Breeding-line	-75.775	72.084	-24.603	-1.994	-10.693
KS19	PI655998	Breeding-line	0	0	Breeding-line	Breeding-line	-22.487	-51.836	30.371	-10.227	-54.239
MR732	PI656051	Breeding-line	0	0	Breeding-line	Breeding-line	-10.450	7.357	16.037	-3.669	-19.645
Macia	PI565121	Breeding-line	0	0	Breeding-line	Breeding-line	-68.689	62.414	-17.258	-2.605	-11.346
Malisor84_7	PI656048	Breeding-line	0	1	Breeding-line	Breeding-line	-60.301	48.161	24.464	-8.666	-24.263

Table S4.1. continued

Martin	PI655987	Breeding-line	0	0	Breeding-line	Breeding-line	15.076	3.054	119.591	-10.798	-74.018
P898012	PI656057	Breeding-line	0	0	Breeding-line	Breeding-line	-66.453	64.036	-22.252	-1.127	-10.547
P9517	PI656058	Breeding-line	0	0	Breeding-line	Breeding-line	-27.006	12.658	66.223	-7.877	-18.353
P_721	PI656055	Breeding-line	0	0	Breeding-line	Breeding-line	-56.163	53.130	9.406	-4.807	-35.892
RTAM2566	PI655977	Breeding-line	0	0	Breeding-line	Breeding-line	-59.853	67.275	-20.517	0.571	-19.361
RTAM428	PI656009	Breeding-line	0	0	Breeding-line	Breeding-line	-65.194	67.423	1.447	-0.483	-32.820
RTX2536	PI656010	Breeding-line	0	0	Breeding-line	Breeding-line	-23.155	-41.395	-9.023	-5.710	-40.448
RTX2737	PI655978	Breeding-line	0	0	Breeding-line	Breeding-line	-11.199	-39.514	5.946	-6.069	-39.197
RTx2783	PI656001	Breeding-line	0	0	Breeding-line	Breeding-line	-42.269	50.014	6.871	0.249	-26.471
RTx2917	PI629040	Breeding-line	0	0	Breeding-line	Breeding-line	-10.459	16.511	19.960	-7.198	-38.448
RTx430	PI655996	Breeding-line	0	0	Breeding-line	Breeding-line	-33.151	13.121	35.916	-6.793	-34.402
RTx434	PI564165	Breeding-line	0	0	Breeding-line	Breeding-line	-17.871	43.912	11.575	-1.411	-48.999
RTx436	PI561071	Breeding-line	0	0	Breeding-line	Breeding-line	-20.894	11.638	16.947	-4.897	-46.331
RTx437	PI629034	Breeding-line	0	1	Breeding-line	Breeding-line	-46.875	17.553	-6.859	-6.914	-27.553
Redbine	PI655989	Breeding-line	0	0	Breeding-line	Breeding-line	16.154	2.615	102.867	-11.185	-67.717
SC0002	PI534118	Durra	0	1	Ethiopia	Eastern-Africa	43.003	-3.698	-8.501	132.728	19.793
SC0003	NA	Durra	0	0	Ethiopia	Eastern-Africa	34.101	-3.170	42.149	-1.126	-40.614
SC0004	PI534119	Durra	0	0	Ethiopia	Eastern-Africa	65.970	5.571	-10.631	-4.087	17.177
SC0006	PI533902	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	41.219	-0.535	-7.381	142.795	10.708
SC0007	PI534120	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	39.544	-1.573	-4.188	155.127	8.661
SC0010	PI534121	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	45.880	7.260	-13.736	9.575	-27.390
SC0012	PI534122	Bicolor	0	0	Ethiopia	Eastern-Africa	33.003	-15.308	-7.582	145.899	38.153
SC0013	PI534123	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	29.172	-2.078	4.936	122.018	0.179
SC0015	PI534124	Bicolor	0	0	Ethiopia	Eastern-Africa	16.746	1.118	-4.967	113.761	3.514
SC0016	PI534125	Bicolor	0	0	Ethiopia	Eastern-Africa	16.903	-5.912	-0.453	59.756	18.771
SC0017	PI533903	Bicolor	0	0	Ethiopia	Eastern-Africa	26.694	-14.469	-2.977	100.114	31.394
SC0019	PI533904	Bicolor	0	0	Ethiopia	Eastern-Africa	30.116	-12.250	-4.936	119.303	30.606
SC0020	PI534126	Bicolor	0	0	Ethiopia	Eastern-Africa	37.312	-9.008	-4.541	153.059	16.287
SC0021	PI534127	Kafir-Bicolor	0	0	Ethiopia	Eastern-Africa	40.872	3.792	-1.562	159.052	-5.635
SC0022	PI656091	Durra	0	1	Ethiopia	Eastern-Africa	52.254	6.249	-13.908	3.782	-28.432
SC0023	PI534128	Durra	0	0	Ethiopia	Eastern-Africa	10.017	8.790	-11.495	0.489	-3.913
SC0024	PI534129	Durra	1	0	Ethiopia	Eastern-Africa	56.801	13.574	-2.219	8.267	-49.081

Table S4.1. continued

SC0025	PI656092	Durra	0	0	Ethiopia	Eastern-Africa	53.909	4.687	-9.683	-0.610	-23.727
SC0027	NA	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	57.295	6.447	-17.214	63.552	-27.231
SC0028	PI534130	Durra	0	0	Ethiopia	Eastern-Africa	57.238	8.009	-19.042	5.150	-27.284
SC0029	PI533905	Durra	0	0	Ethiopia	Eastern-Africa	57.014	13.193	-8.187	10.071	-58.525
SC0030	PI534131	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	46.640	2.866	1.487	104.866	-16.426
SC0031	PI533906	Durra	0	0	Ethiopia	Eastern-Africa	37.478	6.922	-3.577	1.042	-15.520
SC0033	PI534132	Durra	0	0	Ethiopia	Eastern-Africa	53.594	4.371	-4.307	0.819	-22.966
SC0035	PI534133	Durra	0	0	Ethiopia	Eastern-Africa	60.288	5.729	-20.621	4.293	-26.657
SC0036	PI533907	Durra	0	0	Ethiopia	Eastern-Africa	62.399	9.022	-18.221	1.801	-27.467
SC0037	PI534134	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	41.370	-3.244	-6.661	156.277	13.527
SC0038	PI534135	Durra	1	0	Ethiopia	Eastern-Africa	60.275	4.660	-26.763	3.363	-17.622
SC0041	NA	Durra	0	0	Ethiopia	Eastern-Africa	15.543	4.547	1.772	-2.004	-13.478
SC0042	PI576393	Caudatum	0	0	Ethiopia	Eastern-Africa	-26.085	-3.919	16.466	20.569	13.025
SC0043	PI534136	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	42.440	3.004	-11.930	154.497	3.324
SC0044	PI533908	Bicolor	0	0	Ethiopia	Eastern-Africa	34.895	-14.684	-7.456	148.827	42.494
SC0048	PI533909	Caudatum	0	0	Sudan	Northern-Africa	-39.469	34.906	-6.151	-3.876	-2.609
SC0049	PI656098	Guinea	0	0	Sudan	Northern-Africa	-31.757	25.552	-6.149	6.216	-8.330
SC0050	PI533787	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-23.533	6.977	-8.086	10.902	-6.856
SC0051	PI534137	Caudatum	0	0	Sudan	Northern-Africa	-32.456	24.592	-16.450	-4.183	1.968
SC0052	PI533830	Caudatum	0	0	Sudan	Northern-Africa	-29.888	27.480	-19.388	1.809	7.812
SC0053	PI533788	Caudatum	0	0	Sudan	Northern-Africa	-2.303	23.538	-28.112	-11.722	35.106
SC0054	PI533960	Caudatum	0	0	Sudan	Northern-Africa	-22.938	16.819	-15.357	-2.612	10.319
SC0055	PI533755	Caudatum	0	0	Sudan	Northern-Africa	-24.757	18.797	-24.018	-0.387	15.039
SC0056	PI533910	Caudatum	0	0	Sudan	Northern-Africa	-36.338	33.752	-6.141	-3.711	2.920
SC0057	PI533789	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-33.490	20.711	-22.469	-0.718	18.995
SC0058	PI533911	Caudatum	0	0	Sudan	Northern-Africa	-39.969	23.469	-13.646	-1.150	20.140
SC0059	PI656102	Caudatum-Bicolor	0	0	Sudan	Northern-Africa	-10.814	-3.523	22.632	6.749	4.151
SC0060	PI533962	Caudatum	0	0	Sudan	Northern-Africa	-71.925	58.042	-26.359	-5.214	7.267
SC0061	PI576429	Guinea	0	0	Sudan	Northern-Africa	-31.640	-11.467	-8.177	0.208	23.775
SC0062	PI534138	Caudatum	0	0	Sudan	Northern-Africa	-21.556	-5.482	22.765	-4.222	3.883
SC0063	PI533912	Caudatum	0	0	Sudan	Northern-Africa	-32.143	12.067	15.048	-5.634	6.157
SC0064	PI533757	Caudatum	0	0	Sudan	Northern-Africa	-65.785	58.400	-12.879	-1.597	4.910

Table S4.1. continued

SC0066	PI533913	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-31.877	0.217	18.077	4.714	5.180
SC0067	PI534139	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-18.676	-4.852	5.401	3.376	4.621
SC0068	PI534140	Caudatum	0	0	Kenya	Eastern-Africa	-33.375	26.189	-7.254	-3.675	9.030
SC0069	PI595725	Caudatum	0	0	Kenya	Eastern-Africa	-33.654	6.984	22.774	-2.207	14.943
SC0072	PI 153844	Durra-Caudatum	0	0	Kenya	Eastern-Africa	-18.119	3.493	14.906	8.345	-5.946
SC0073	PI576430	Caudatum	0	0	Kenya	Eastern-Africa	-17.166	-9.222	26.848	6.770	-0.700
SC0074	PI 153846	Unknown/Other	0	0	Kenya	Eastern-Africa	5.621	2.638	60.793	1.006	-37.144
SC0075	PI 153852	Caudatum	0	0	Kenya	Eastern-Africa	-12.136	-3.676	37.325	-4.501	14.868
SC0077	PI534141	Caudatum	0	1	Kenya	Eastern-Africa	-16.128	-2.893	34.172	-4.794	20.428
SC0078	PI533914	Caudatum	0	0	Kenya	Eastern-Africa	-10.335	-8.413	48.014	1.629	19.755
SC0079	PI533915	Unknown/Other	0	0	Kenya	Eastern-Africa	-8.361	-4.294	38.248	0.293	23.686
SC0080	PI534142	Caudatum	0	0	Kenya	Eastern-Africa	-23.070	-4.383	14.168	8.978	-3.663
SC0083	PI534143	Caudatum	0	0	Uganda	Eastern-Africa	-26.299	-17.758	24.545	4.353	36.875
SC0084	PI534144	Caudatum	0	0	Uganda	Eastern-Africa	-21.348	-19.304	16.675	7.397	38.302
SC0085	PI656109	Caudatum-Bicolor	0	0	Uganda	Eastern-Africa	-16.528	8.899	11.440	9.172	-4.455
SC0086	PI595726	Durra-Caudatum	0	0	Kenya	Eastern-Africa	-20.129	-20.757	19.102	-0.774	31.362
SC0087	PI595727	Guinea	0	0	Kenya	Eastern-Africa	-18.741	-30.546	68.954	-3.547	46.159
SC0090	PI533916	Guinea	0	0	Zaire	Middle-Africa	-25.110	-12.336	-7.322	2.042	30.088
SC0091	PI534145	Unknown/Other	0	0	Zimbabwe	Eastern-Africa	-28.512	-92.700	-37.129	-5.399	2.393
SC0092	NA	Durra	0	0	Ethiopia	Eastern-Africa	39.863	-5.153	54.771	-8.628	-16.171
SC0093	PI533751	Caudatum	0	0	Sudan	Northern-Africa	-29.741	-23.806	-7.346	-6.460	31.330
SC0094	PI533950	Guinea	0	0	Sudan	Northern-Africa	-35.414	-112.916	-34.702	-15.331	0.251
SC0096	PI533951	Caudatum	0	0	Nigeria	Western-Africa	-62.138	46.142	-26.155	-3.332	11.514
SC0097	PI533917	Guinea	0	0	Nigeria	Western-Africa	-40.050	-113.154	-40.396	-11.920	-3.546
SC0098	PI533790	Guinea	0	0	Nigeria	Western-Africa	-35.619	-112.233	-43.527	-9.678	2.448
SC0099	PI576383	Guinea	0	0	Nigeria	Western-Africa	-40.464	-93.750	-44.524	-10.894	2.742
SC0100	PI533791	Guinea	0	1	Nigeria	Western-Africa	-36.252	-115.038	-44.839	-9.995	2.157
SC0101	PI533954	Guinea-Caudatum	0	0	S.Africa	Southern-Africa	-18.690	-12.248	77.019	-1.696	10.514
SC0103	PI533752	Caudatum	0	0	S.Africa	Southern-Africa	-60.510	35.506	-21.867	-0.612	29.758
SC0105	PI533745	Durra	0	1	India	India	81.716	15.231	-22.949	-24.987	24.464
SC0106	PI 248238	Bicolor	0	NA	Ethiopia	Eastern-Africa	38.179	-3.571	-5.858	159.457	11.445
SC0108	PI533792	Caudatum	0	0	Ethiopia	Eastern-Africa	-80.001	69.036	-23.884	-5.871	-1.171

Table S4.1. continued

SC0109	PI533793	Caudatum	0	0	Ethiopia	Eastern-Africa	-66.795	69.983	-30.769	0.493	-11.493
SC0110	PI533794	Caudatum	0	0	Ethiopia	Eastern-Africa	-63.105	63.976	-29.933	-2.556	2.371
SC0111	PI534146	Caudatum	0	0	Ethiopia	Eastern-Africa	-74.245	64.784	-29.527	-3.257	4.827
SC0112	PI533918	Caudatum	0	0	Ethiopia	Eastern-Africa	-68.894	62.928	-35.381	-3.294	4.625
SC0113	PI576395	Caudatum	0	0	Uganda	Eastern-Africa	-58.464	41.419	0.525	-3.517	5.388
SC0114	PI533832	Caudatum	0	0	Uganda	Eastern-Africa	-73.184	58.248	-24.642	-3.865	12.165
SC0115	PI533965	Caudatum-Bicolor	0	0	Uganda	Eastern-Africa	-61.766	38.423	-20.141	-1.669	29.935
SC0118	PI533759	Caudatum	0	0	Sudan	Northern-Africa	-57.582	37.459	-23.316	-6.917	24.524
SC0119	PI533973	Caudatum	0	0	Zimbabwe	Eastern-Africa	-59.740	44.246	-16.235	-5.680	15.833
SC0120	PI533760	Caudatum	0	0	Sudan	Northern-Africa	-66.834	57.479	-27.046	-0.535	0.110
SC0121	PI533961	Caudatum	0	0	S.Africa	Southern-Africa	-33.843	10.493	41.583	-9.426	26.894
SC0123	PI534147	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	50.903	7.824	2.754	22.708	-15.477
SC0124	PI533919	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	50.037	5.792	-17.543	93.340	-14.098
SC0126	PI533795	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	-49.719	41.562	-19.305	-4.490	21.798
SC0127	PI533920	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	43.668	-0.653	-4.178	146.112	1.995
SC0131	PI533796	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	28.816	13.886	-6.662	60.504	-17.355
SC0132	PI659695	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	61.901	15.361	-29.831	9.167	-32.471
SC0134	PI656114	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	61.890	14.319	-19.098	33.557	-45.098
SC0135	PI534148	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	52.991	14.095	-15.096	39.754	-30.746
SC0136	PI534149	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	35.944	1.216	-0.431	125.428	0.409
SC0137	PI534150	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	23.629	0.265	2.067	100.455	2.825
SC0138	NA	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	72.751	19.340	-23.682	19.278	-22.883
SC0139	PI534151	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	48.405	10.731	-3.392	54.797	-30.158
SC0140	PI576431	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	51.137	15.592	-14.213	71.633	-30.047
SC0141	PI534152	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	49.026	9.950	-17.419	72.857	-22.501
SC0142	PI534153	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	30.943	11.058	10.257	15.200	-31.829
SC0144	PI533921	Durra-Caudatum	0	0	Ethiopia	Eastern-Africa	0.887	5.543	1.664	-7.370	4.919
SC0145	PI656082	Bicolor	0	0	Ethiopia	Eastern-Africa	31.224	-15.345	0.534	127.335	33.554
SC0146	PI534154	Caudatum	0	0	Ethiopia	Eastern-Africa	-69.023	54.973	-16.608	0.176	5.266
SC0147	PI533922	Bicolor	0	0	Ethiopia	Eastern-Africa	29.049	-17.529	-1.575	120.819	36.362
SC0150	NA	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	51.811	6.529	-7.871	78.417	-25.644
SC0154	PI533797	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	53.427	7.598	-0.307	105.300	-13.238

Table S4.1. continued

SC0155	PI534155	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	51.121	1.598	-16.229	73.685	-2.119
SC0156	NA	Bicolor	0	0	Ethiopia	Eastern-Africa	43.457	-9.727	-6.628	90.135	32.564
SC0157	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	44.217	9.048	19.828	2.617	-52.043
SC0158	PI534156	Durra	NA	1	Ethiopia	Eastern-Africa	73.670	3.202	-20.759	47.800	30.151
SC0159	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	47.469	-0.042	-2.521	93.826	15.652
SC0161	PI595728	Durra-Caudatum	0	0	Ethiopia	Eastern-Africa	52.913	-3.105	-8.520	51.287	17.750
SC0165	PI533923	Caudatum	0	0	Ethiopia	Eastern-Africa	-43.531	31.409	-11.235	-3.008	-0.181
SC0166	PI533924	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	41.553	0.538	-9.570	155.584	8.732
SC0167	PI533925	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	42.423	-0.572	-5.584	157.988	8.283
SC0170	PI534157	Caudatum	0	0	Ethiopia	Eastern-Africa	-79.138	53.244	-23.422	-7.130	9.181
SC0171	PI533798	Caudatum	0	0	Ethiopia	Eastern-Africa	-41.049	33.447	26.311	-10.486	-22.103
SC0172	PI656117	Caudatum	0	0	Ethiopia	Eastern-Africa	-74.611	66.904	-27.096	-5.721	0.773
SC0173	PI533799	Caudatum	0	0	Ethiopia	Eastern-Africa	-77.823	67.787	-29.905	-3.545	5.160
SC0175	PI533800	Caudatum	0	0	Ethiopia	Eastern-Africa	-66.321	59.760	-27.357	-5.932	-2.957
SC0176	PI534158	Caudatum	0	1	Ethiopia	Eastern-Africa	-68.513	58.187	1.171	-6.210	-8.049
SC0178	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	55.943	8.519	-3.707	81.895	-5.444
SC0179	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	71.457	7.909	14.614	-23.366	-20.766
SC0180	NA	Durra	0	0	Ethiopia	Eastern-Africa	33.541	2.520	-7.996	-8.746	11.165
SC0181	PI534159	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	37.731	-0.741	-0.460	132.095	7.689
SC0182	PI533801	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	53.529	-0.864	-22.946	81.468	37.927
SC0183	PI534160	Kafir	0	0	Mexico	Americas	-24.989	0.142	37.253	-0.384	4.403
SC0184	PI597958	Caudatum-Kafir	0	0	S.Africa	Southern-Africa	-18.895	-2.019	35.098	-0.473	26.585
SC0185	PI534161	Caudatum	0	1	Nigeria	Western-Africa	-45.158	18.148	-20.812	-4.979	-0.339
SC0186	PI533802	Caudatum-Durra	0	0	Nigeria	Western-Africa	-24.218	-55.658	-27.310	-2.447	-31.921
SC0187	PI533803	Caudatum-Durra	0	0	Nigeria	Western-Africa	-17.052	-45.770	-21.034	-2.281	-32.097
SC0188	PI533804	Caudatum-Durra	0	0	Nigeria	Western-Africa	-21.833	-42.929	-17.436	-1.723	-30.809
SC0190	NA	Durra	0	0	India	India	66.074	2.171	14.138	-29.639	1.207
SC0191	NA	Durra-Bicolor	0	0	India	India	67.874	11.273	-16.885	-17.866	3.323
SC0192	PI576390	Durra	0	0	India	India	75.064	17.415	-14.481	-26.816	6.267
SC0193	PI533747	Durra-Bicolor	0	0	India	India	97.203	23.633	-29.077	-26.601	-3.758
SC0196	PI597943	Durra	0	0	India	India	74.922	10.027	-24.301	-21.631	20.263
SC0199	PI533810	Durra	1	1	India	India	108.122	25.354	-25.019	-31.300	-2.289

Table S4.1. continued

SC0200	PI533932	Durra	0	0	India	India	106.236	29.234	-22.477	-30.684	-0.404
SC0201	NA	Durra	1	0	India	India	88.936	11.901	-30.803	-33.998	29.544
SC0202	PI533811	Durra	0	0	India	India	90.449	12.510	-20.531	-37.608	28.857
SC0203	PI533812	Durra	0	0	India	India	99.074	14.277	-33.848	-38.188	31.877
SC0205	PI533813	Durra	0	0	India	India	96.175	18.912	-38.064	-36.795	32.723
SC0206	PI533814	Durra	0	0	India	India	97.783	12.029	-38.070	-39.807	37.038
SC0207	PI533815	Durra	0	0	India	India	97.106	14.589	-32.515	-38.678	33.237
SC0208	PI533816	Durra	0	0	India	India	96.717	14.940	-42.224	-40.701	36.174
SC0209	PI533817	Durra	0	0	India	India	90.719	22.189	-18.975	-26.219	1.149
SC0210	PI533942	Durra	0	0	India	India	91.024	11.704	-30.221	-34.777	31.125
SC0211	PI533945	Durra	0	0	India	India	103.811	23.233	-37.660	-33.882	8.473
SC0212	PI533823	Durra	0	1	India	India	95.697	13.060	-32.553	-35.650	31.899
SC0213	PI576391	Bicolor	0	0	India	India	85.400	7.666	-20.284	-30.944	38.117
SC0214	PI533750	Bicolor	0	0	India	India	65.019	2.991	-16.881	-12.338	9.898
SC0215	PI533825	Durra	0	1	India	India	94.430	12.521	-29.750	-37.691	32.503
SC0216	PI533827	Durra	0	0	India	India	86.399	28.891	-15.632	-20.568	-13.249
SC0217	PI533926	Durra	0	1	India	India	104.468	23.108	-33.889	-29.934	-3.727
SC0218	PI534162	Durra	0	0	China	Asia	76.462	13.962	-31.215	-27.976	16.442
SC0220	PI533805	Durra	0	0	India	India	104.000	23.169	-42.329	-32.487	11.978
SC0221	PI533806	Durra	0	1	India	India	102.212	12.050	-39.423	-38.262	34.122
SC0222	PI533931	Durra	0	0	India	India	78.008	12.734	-29.159	-22.866	16.073
SC0223	PI533807	Caudatum	0	0	Nigeria	Western-Africa	-58.225	39.553	-20.659	1.173	9.663
SC0224	PI533927	Bicolor	0	0	Ethiopia	Eastern-Africa	22.836	-12.579	-5.270	103.190	28.796
SC0226	PI533828	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	55.789	11.954	-18.989	4.932	-31.506
SC0227	PI533829	Durra	0	0	Ethiopia	Eastern-Africa	64.216	11.699	-24.296	-4.109	-18.610
SC0228	PI533756	Caudatum	0	0	Ethiopia	Eastern-Africa	-59.600	57.472	-14.887	-6.639	-5.003
SC0229	PI533808	Caudatum	0	0	Nigeria	Western-Africa	-49.157	-3.519	-20.743	-10.731	-15.790
SC0230	PI533809	Durra	0	0	India	India	84.864	7.684	-33.261	-35.861	31.093
SC0231	PI533935	Durra	0	0	India	India	90.017	27.987	-33.935	-22.098	-11.364
SC0233	PI533746	Durra	0	0	India	India	97.714	17.289	-30.966	-33.828	27.137
SC0235	NA	Guinea-Bicolor	0	0	Ethiopia	Eastern-Africa	33.663	1.003	-2.520	118.450	3.938
SC0236	NA	Durra	0	0	Unknown	Unknown	50.709	-3.194	27.867	-9.875	-7.546

Table S4.1. continued

SC0237	PI533834	Caudatum	0	NA	Sudan	Northern-Africa	-73.789	61.580	-25.757	-6.005	12.979
SC0239	PI533837	Caudatum	0	NA	Sudan	Northern-Africa	-61.684	45.621	-22.291	-2.059	-7.553
SC0240	PI533842	Durra	NA	NA	India	India	102.230	15.937	-33.404	-24.165	14.095
SC0241	PI533843	Guinea	0	NA	India	India	28.080	-12.886	39.344	-10.074	40.666
SC0242	PI533844	Guinea	1	1	Nepal	Asia	-3.462	-31.246	65.239	-9.831	86.529
SC0243	PI533845	Guinea	0	1	Nepal	Asia	20.564	-20.181	41.039	-17.018	56.841
SC0244	PI533847	Guinea	0	0	India	India	-2.133	-23.354	60.331	-9.113	69.647
SC0245	PI534011	Durra	0	0	India	India	74.479	5.474	-16.428	-30.937	42.337
SC0247	NA	Guinea	0	0	India	India	2.668	-29.720	51.855	-10.779	74.428
SC0248	PI533853	Guinea	0	0	India	India	9.031	-27.908	46.068	-11.895	82.183
SC0249	NA	Guinea	0	0	India	India	-16.226	15.073	31.940	-1.697	27.033
SC0250	PI533764	Guinea	0	0	India	India	1.592	-28.430	57.340	-11.257	82.478
SC0252	PI533848	Guinea	0	1	India	India	1.229	-28.863	61.662	-11.835	76.620
SC0253	PI533765	Guinea	0	0	India	India	6.878	-27.157	52.482	-10.440	83.994
SC0254	PI533854	Guinea	0	1	India	India	11.326	-24.618	45.275	-9.003	80.467
SC0256	PI533818	Guinea	0	1	Nigeria	Western-Africa	-39.643	-115.711	-43.720	-12.963	4.040
SC0257	PI533819	Guinea	0	1	Zimbabwe	Eastern-Africa	-23.761	-64.104	4.585	-2.019	8.889
SC0258	PI533749	Guinea	0	1	Africa	Eastern-Africa	-21.361	-28.327	76.150	-3.292	64.592
SC0259	PI576334	Guinea	0	0	Sudan	Northern-Africa	-28.100	13.602	59.703	-11.940	-8.805
SC0260	NA	Guinea	NA	NA	Mali	Western-Africa	-18.931	-59.921	6.638	-9.956	3.263
SC0261	PI533841	Guinea	0	0	Nigeria	Western-Africa	-33.030	-62.575	-20.282	-7.901	11.960
SC0262	PI534002	Guinea	0	0	Mali	Western-Africa	-31.336	-35.373	-18.357	-8.862	13.798
SC0263	NA	Guinea	0	0	Mali	Western-Africa	-23.852	-74.553	-17.640	-10.372	7.854
SC0265	PI533766	Guinea	0	0	BurkinaFaso	Western-Africa	-33.049	-81.220	-28.039	-9.718	12.065
SC0266	NA	Guinea	0	0	BurkinaFaso	Western-Africa	-16.445	-68.575	-5.615	-9.340	-2.728
SC0267	PI533836	Guinea	0	0	Sudan	Northern-Africa	-29.444	-48.595	-8.847	-6.264	13.487
SC0268	PI533768	Guinea	0	0	Sudan	Northern-Africa	-40.695	15.947	-14.381	-6.371	23.721
SC0269	PI534058	Guinea	0	0	Nigeria	Western-Africa	-31.401	-106.774	-37.087	-12.583	-4.311
SC0270	PI534059	Guinea	0	0	Nigeria	Western-Africa	-30.981	-101.895	-39.488	-12.411	0.407
SC0271	PI533772	Guinea	0	0	Nigeria	Western-Africa	-19.728	-57.104	-23.992	-7.916	-12.349
SC0272	PI533773	Guinea	0	0	Nigeria	Western-Africa	-29.033	-100.840	-33.758	-7.319	-18.568
SC0273	PI534062	Guinea	0	0	Nigeria	Western-Africa	33.076	-2.189	-1.069	106.204	1.686

Table S4.1. continued

SC0275	PI534065	Guinea	0	0	Nigeria	Western-Africa	-34.496	-104.178	-43.364	-12.801	0.620
SC0276	NA	Guinea	0	0	Nigeria	Western-Africa	-24.426	-12.955	-5.538	-2.166	-23.551
SC0277	PI534067	Guinea	0	0	Nigeria	Western-Africa	-36.921	-107.135	-35.138	-13.285	1.208
SC0278	PI533763	Guinea	0	0	Nigeria	Western-Africa	-28.055	-71.628	-23.349	-5.625	10.982
SC0279	PI534070	Guinea	0	0	Nigeria	Western-Africa	-39.065	-113.730	-42.245	-12.009	0.180
SC0280	PI534085	Guinea	0	0	Nigeria	Western-Africa	-35.131	-103.116	-26.968	-11.777	-3.596
SC0281	PI533786	Guinea	0	0	Nigeria	Western-Africa	-12.653	-45.478	17.487	6.123	20.596
SC0282	PI533784	Guinea	0	0	Nigeria	Western-Africa	-34.839	-98.082	-43.122	-11.461	1.042
SC0283	PI533869	Guinea	0	0	Tanzania	Eastern-Africa	-16.013	-22.988	72.473	-5.618	43.412
SC0284	PI533767	Guinea	0	0	BurkinaFaso	Western-Africa	-38.015	-105.629	-35.376	-10.252	1.910
SC0285	PI533862	Guinea	0	0	Chad	Middle-Africa	-38.590	-111.778	-42.777	-13.931	-0.460
SC0287	NA	Guinea	0	0	Nigeria	Western-Africa	-38.320	-110.563	-40.850	-13.201	0.693
SC0289	PI533780	Guinea	0	0	Nigeria	Western-Africa	-32.358	-101.725	-36.821	-11.041	-1.804
SC0290	PI533883	Guinea	0	0	Nigeria	Western-Africa	-31.226	-110.170	-48.605	-6.026	-16.675
SC0291	PI533888	Guinea	0	0	Nigeria	Western-Africa	-24.023	-59.683	33.719	-8.955	36.379
SC0292	PI533782	Guinea	0	0	Nigeria	Western-Africa	-35.899	-103.604	-29.451	-11.816	-3.992
SC0293	PI533783	Guinea	0	0	Nigeria	Western-Africa	-35.955	-100.593	-44.804	-10.119	-2.011
SC0295	PI656093	Guinea	0	1	Nigeria	Western-Africa	-35.099	-109.903	-32.839	-11.990	-0.911
SC0296	PI533885	Guinea	0	1	Nigeria	Western-Africa	-36.086	-90.974	-32.721	-12.004	1.544
SC0297	NA	Guinea	0	1	Nigeria	Western-Africa	-34.893	-110.195	-39.061	-14.263	1.351
SC0298	PI533887	Guinea	0	1	Nigeria	Western-Africa	-23.121	-70.061	-33.089	-9.709	0.466
SC0299	PI533785	Guinea	0	0	Nigeria	Western-Africa	-26.920	-89.382	-31.248	-10.602	0.191
SC0300	PI534095	Guinea	0	0	Nigeria	Western-Africa	-36.880	-90.417	-34.160	-10.375	4.123
SC0301	PI656094	Guinea	0	0	Mali	Western-Africa	-28.128	-84.831	-29.663	-11.856	7.242
SC0303	PI533839	Guinea	0	1	Nigeria	Western-Africa	-3.872	-54.520	2.459	20.688	19.052
SC0305	PI534037	Durra	0	0	Chad	Middle-Africa	-14.266	-43.688	15.896	18.400	-0.135
SC0306	PI533958	Bicolor	0	0	USA	Americas	33.452	-16.303	-8.176	146.541	38.836
SC0308	PI533820	Bicolor	0	0	Malawi	Eastern-Africa	-18.572	-33.815	68.072	-1.758	58.628
SC0309	PI533754	Bicolor	0	0	Sudan	Northern-Africa	27.417	-4.524	-16.824	4.820	8.914
SC0311	PI533753	Bicolor	0	1	Sudan	Northern-Africa	25.421	-4.351	-15.363	4.137	9.624
SC0314	NA	Bicolor-Guinea	0	0	India	India	2.451	-12.288	85.568	-1.251	-5.036
SC0315	PI533851	Guinea-Bicolor	0	1	India	India	15.066	-10.884	49.907	-5.285	52.563

Table S4.1. continued

SC0317	PI533855	Guinea-Bicolor	0	1	India	India	14.343	-13.009	61.705	-3.181	44.279
SC0319	PI533833	Caudatum	0	0	Uganda	Eastern-Africa	-58.479	65.064	-24.321	4.711	-16.924
SC0320	PI533863	Kafir	0	0	Chad	Middle-Africa	-37.024	17.674	-15.419	-1.537	11.011
SC0322	PI533821	Caudatum	0	0	Tanzania	Eastern-Africa	-30.018	-8.516	28.289	3.875	18.895
SC0323	PI576399	Caudatum	0	0	Sudan	Northern-Africa	-38.712	-4.987	24.859	-0.362	11.736
SC0324	PI576396	Caudatum	0	0	Uganda	Eastern-Africa	-67.872	50.352	-11.560	-6.321	9.348
SC0325	PI533957	Caudatum	0	0	USA	Americas	-26.446	-5.363	8.350	15.036	19.228
SC0328	PI534112	Caudatum	0	0	Uganda	Eastern-Africa	-72.213	62.059	-15.733	-5.451	9.784
SC0329	PI533838	Caudatum	0	0	Nigeria	Western-Africa	-43.150	9.553	-9.857	-4.852	-10.842
SC0330	PI534106	Caudatum	0	0	Uganda	Eastern-Africa	-64.690	49.605	-12.179	-3.699	11.925
SC0331	PI533824	Durra-Caudatum	0	0	Nigeria	Western-Africa	22.836	-9.579	-2.388	112.312	28.265
SC0333	PI533761	Caudatum	0	0	Ethiopia	Eastern-Africa	-39.411	20.523	-19.800	-5.871	19.840
SC0334	PI533986	Caudatum	0	0	Sudan	Northern-Africa	-46.404	24.370	-19.391	-2.598	15.999
SC0335	PI533865	Caudatum	0	0	Sudan	Northern-Africa	-21.619	3.165	-13.468	-5.248	16.773
SC0336	PI533995	Caudatum	0	0	Sudan	Northern-Africa	-12.257	7.731	12.806	-4.652	-8.379
SC0337	PI533868	Caudatum	0	0	Sudan	Northern-Africa	-36.153	11.856	-9.601	-5.487	23.305
SC0338	PI534044	Caudatum	0	0	Sudan	Northern-Africa	-59.501	36.810	-17.727	-4.098	24.493
SC0339	PI576412	Caudatum	0	0	Sudan	Northern-Africa	-22.966	4.258	-12.797	2.569	5.836
SC0340	PI533867	Caudatum	0	0	Sudan	Northern-Africa	-21.994	10.233	-21.673	0.515	9.961
SC0342	PI534064	Caudatum	0	0	Nigeria	Western-Africa	-34.800	-65.234	-27.989	-7.778	8.983
SC0343	NA	Caudatum	0	0	Nigeria	Western-Africa	-37.066	21.011	5.924	-6.543	13.555
SC0344	PI533774	Caudatum	0	0	Nigeria	Western-Africa	-22.584	-17.432	-17.647	-6.037	9.297
SC0345	PI534073	Caudatum	0	0	Nigeria	Western-Africa	-13.746	-11.084	-15.738	-5.801	6.909
SC0347	PI534074	Caudatum	0	0	Nigeria	Western-Africa	-26.117	-77.871	-31.001	-12.660	-18.145
SC0348	PI534075	Caudatum	0	0	Nigeria	Western-Africa	-22.119	-77.241	-29.020	-10.211	-15.046
SC0349	PI595712	Caudatum	0	0	Nigeria	Western-Africa	-19.277	-7.790	-13.298	-3.651	7.089
SC0350	PI576357	Caudatum	0	0	Nigeria	Western-Africa	-37.347	-16.157	-22.796	-4.618	16.404
SC0351	PI534060	Caudatum	0	0	Nigeria	Western-Africa	-32.548	-96.471	-32.725	-8.541	-1.570
SC0352	PI534042	Caudatum	0	0	Sudan	Northern-Africa	-19.564	9.140	-1.768	-5.788	12.496
SC0353	PI533777	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-19.877	-48.890	-28.963	0.624	-28.255
SC0354	PI533890	Caudatum	0	0	Nigeria	Western-Africa	-34.061	-94.113	-28.832	-12.949	-1.964
SC0356	PI533895	Caudatum	0	0	Nigeria	Western-Africa	-37.264	-61.616	-35.620	-11.697	7.700

Table S4.1. continued

SC0358	PI533879	Caudatum	0	0	Nigeria	Western-Africa	-19.116	-72.581	-33.310	-9.070	-30.111
SC0362	PI533775	Caudatum	0	0	Nigeria	Western-Africa	-28.832	-63.822	-21.053	-11.003	-31.237
SC0366	PI533771	Caudatum	0	0	Nigeria	Western-Africa	-21.229	-69.973	-36.294	-6.051	-33.468
SC0367	PI533872	Caudatum	0	0	Nigeria	Western-Africa	-21.796	-67.102	-28.525	-6.694	-45.128
SC0368	PI533873	Caudatum	0	0	Nigeria	Western-Africa	-29.042	-68.497	-29.396	-8.462	-44.179
SC0369	PI533874	Caudatum	0	0	Nigeria	Western-Africa	-19.105	-71.688	-30.856	-6.891	-32.441
SC0370	PI533776	Caudatum	0	0	Nigeria	Western-Africa	-29.150	-66.082	-35.906	-13.246	-28.983
SC0371	NA	Caudatum	0	0	Nigeria	Western-Africa	-30.361	-61.703	-33.444	-12.916	-32.999
SC0372	PI533878	Caudatum	0	0	Nigeria	Western-Africa	-17.814	-65.275	-25.664	-2.989	-68.504
SC0373	PI656095	Caudatum	0	0	Nigeria	Western-Africa	-39.304	-54.987	-33.116	-12.844	-32.369
SC0374	PI533778	Caudatum	0	0	Nigeria	Western-Africa	-38.380	-82.298	-40.703	-10.749	-20.634
SC0377	PI534082	Caudatum	0	0	Nigeria	Western-Africa	-38.630	-90.203	-29.754	-10.864	-18.676
SC0380	PI533781	Caudatum	0	0	Nigeria	Western-Africa	-21.716	-39.424	-33.500	0.143	-22.383
SC0382	PI534088	Caudatum	0	0	Nigeria	Western-Africa	32.272	-21.896	-30.948	-14.727	-25.724
SC0384	PI534089	Caudatum	0	0	Nigeria	Western-Africa	-24.949	-81.301	-37.374	-8.786	-32.847
SC0386	PI656119	Caudatum	0	0	Nigeria	Western-Africa	-22.209	-76.574	-13.478	-7.671	-37.116
SC0387	PI533891	Caudatum	0	0	Nigeria	Western-Africa	-32.138	-76.050	-22.654	-12.421	-23.840
SC0388	PI533893	Caudatum	0	0	Nigeria	Western-Africa	-68.518	62.235	-27.891	-0.482	-2.444
SC0389	PI533894	Caudatum	0	0	Nigeria	Western-Africa	-30.837	-65.033	-28.349	-14.179	-27.775
SC0391	PI656096	Caudatum	0	0	Egypt	Northern-Africa	-33.539	-58.942	-34.043	-13.376	-35.746
SC0392	PI533770	Caudatum	0	0	Nigeria	Western-Africa	-18.545	-63.627	-34.669	-3.040	-48.323
SC0393	PI534066	Caudatum	0	0	Nigeria	Western-Africa	-51.177	15.948	-20.156	-8.435	-14.753
SC0394	PI533880	Caudatum	0	0	Nigeria	Western-Africa	-29.424	-67.582	-28.278	-14.964	-34.469
SC0396	PI533877	Caudatum	0	0	Nigeria	Western-Africa	-29.398	-64.854	-36.381	-13.431	-27.427
SC0397	PI534076	Caudatum	0	0	Nigeria	Western-Africa	-24.114	-43.728	-20.418	-11.331	-31.132
SC0398	PI533881	Caudatum	0	0	Nigeria	Western-Africa	-31.033	-65.303	-28.531	-12.494	-36.246
SC0399	PI533882	Caudatum	0	0	Nigeria	Western-Africa	-16.584	-50.589	-31.533	-3.468	-40.846
SC0400	NA	Caudatum	0	1	Nigeria	Western-Africa	-20.732	-62.441	7.128	-15.435	-17.030
SC0401	PI533896	Caudatum	0	0	Nigeria	Western-Africa	-29.893	-65.558	-30.170	-10.714	-22.398
SC0402	PI533897	Caudatum	0	0	Nigeria	Western-Africa	-2.250	-30.568	-26.159	-19.400	-8.246
SC0403	PI533898	Caudatum	0	0	Nigeria	Western-Africa	-34.877	-55.224	-30.709	-9.690	-32.697
SC0405	PI533835	Caudatum	0	0	Sudan	Northern-Africa	-55.382	52.671	-24.198	3.797	-9.711

Table S4.1. continued

SC0406	PI597953	Guinea	0	0	Nigeria	Western-Africa	-27.621	-48.900	-33.259	-0.437	-30.030
SC0407	PI533779	Guinea-Caudatum	0	0	Nigeria	Western-Africa	-3.631	-43.409	-25.845	5.378	-38.193
SC0408	PI533884	Caudatum	0	0	Nigeria	Western-Africa	-14.202	-63.657	-20.285	-2.989	-26.966
SC0409	NA	Caudatum-Guinea	0	0	Nigeria	Western-Africa	-16.108	-39.625	-18.551	2.718	-25.280
SC0411	PI533866	Caudatum-Bicolor	0	0	Sudan	Northern-Africa	-10.305	-5.305	-4.620	-2.201	-21.964
SC0412	NA	Caudatum-Bicolor	0	0	Sudan	Northern-Africa	17.913	12.365	-12.382	11.530	-15.200
SC0413	PI534079	Caudatum-Bicolor	0	0	Nigeria	Western-Africa	-35.196	-14.343	-24.084	0.832	-19.102
SC0414	PI533831	Caudatum	0	1	Sudan	Northern-Africa	-29.422	32.549	-14.754	4.075	-3.934
SC0417	PI533861	Caudatum	0	1	Senegal	Western-Africa	46.989	15.852	-22.489	4.932	-42.059
SC0418	PI533822	Caudatum	0	0	Tanzania	Eastern-Africa	-41.747	21.109	5.853	3.403	8.529
SC0420	PI533769	Caudatum	0	0	Sudan	Northern-Africa	-34.859	9.927	-7.920	-2.935	-17.779
SC0422	PI576421	Caudatum	0	0	Uganda	Eastern-Africa	-35.069	4.650	11.370	5.677	7.759
SC0423	PI533758	Caudatum	0	0	Sudan	Northern-Africa	-68.727	54.168	-18.503	-4.129	4.167
SC0424	PI533901	Caudatum	0	0	Japan	Asia	-36.231	26.210	-6.740	-4.302	-1.620
SC0425	PI533762	Caudatum	0	0	Sudan	Northern-Africa	-33.388	21.699	-20.983	1.366	3.101
SC0426	PI533900	Caudatum	0	0	Japan	Asia	3.429	-25.314	-23.894	8.211	-46.465
SC0427	PI533899	Caudatum	0	0	Nigeria	Western-Africa	18.202	-5.470	-17.974	10.397	-18.949
SC0428	PI534057	Caudatum	0	0	Nigeria	Western-Africa	5.148	7.555	14.495	5.863	9.198
SC0430	PI576398	Durra	0	1	Ethiopia	Eastern-Africa	43.228	-11.179	-5.405	128.904	34.085
SC0431	PI533846	Durra	0	0	India	India	91.008	13.743	-12.879	-32.291	17.191
SC0432	PI533941	Durra	0	0	Tanzania	Eastern-Africa	-9.037	-25.002	34.674	-9.413	11.285
SC0435	PI576331	Durra	0	1	India	India	21.235	-16.821	47.993	-10.150	71.660
SC0436	NA	Durra	0	1	India	India	84.569	9.887	-19.885	-31.858	27.101
SC0437	PI533947	Durra	0	1	India	India	86.449	12.126	-25.029	-30.914	26.324
SC0438	NA	Durra	0	0	Ethiopia	Eastern-Africa	37.305	-7.669	62.475	-9.357	-14.367
SC0441	PI534009	Durra	0	0	India	India	83.816	8.649	-16.747	-34.223	37.265
SC0442	PI534010	Durra	0	1	India	India	89.760	8.408	-29.101	-34.709	35.539
SC0445	NA	Durra	0	1	India	India	12.375	-18.313	48.265	-14.516	79.939
SC0448	NA	Durra	1	1	India	India	97.671	16.640	-33.299	-36.720	27.761
SC0449	PI597950	Durra	0	1	India	India	21.712	-9.572	48.369	-10.325	61.930
SC0450	PI533852	Durra	0	0	India	India	75.265	8.025	-18.133	-19.608	16.416
SC0451	PI533934	Durra	0	0	India	India	-60.084	48.860	-29.830	0.407	8.739

Table S4.1. continued

SC0452	PI533944	Durra	0	0	India	India	107.257	22.121	-36.769	-29.911	9.947
SC0454	PI576362	Durra	0	0	India	India	89.552	8.491	-35.978	-35.066	41.095
SC0455	PI534013	Durra	0	0	India	India	85.428	14.175	-22.128	-37.281	27.562
SC0456	NA	Durra	0	0	India	India	80.951	11.338	-7.592	-32.221	18.490
SC0457	PI533849	Durra-Bicolor	0	0	India	India	81.213	16.698	-0.653	-21.929	9.486
SC0458	NA	Durra	0	0	India	India	75.985	7.972	0.679	-29.913	25.761
SC0459	PI533850	Durra	0	0	India	India	97.015	19.460	-12.946	-32.354	11.811
SC0460	PI534017	Durra	0	0	India	India	84.248	15.693	-23.095	-25.160	20.860
SC0462	PI534018	Durra	0	0	India	India	87.877	12.263	-24.285	-32.156	21.117
SC0463	NA	Durra	0	0	India	India	88.231	13.549	0.918	-34.370	10.995
SC0464	PI534031	Durra	0	0	India	India	83.710	9.368	-17.168	-36.778	24.587
SC0465	PI533997	Guinea-Durra	0	0	Arabia	Arabia/MiddleEast	-15.341	-25.221	37.379	2.151	16.328
SC0466	PI533981	Durra-Bicolor	0	0	India	India	9.054	-5.481	68.079	-8.168	-24.077
SC0467	PI533943	Durra-Bicolor	0	0	India	India	83.899	15.326	7.777	-27.581	0.724
SC0468	PI533946	Durra-Bicolor	0	0	India	India	64.506	13.343	33.582	-13.701	4.506
SC0470	NA	Durra	0	0	India	India	83.597	14.103	17.566	-23.361	-13.907
SC0471	PI534014	Durra	0	0	India	India	67.295	14.411	-25.453	-28.273	17.716
SC0472	PI534016	Durra	0	0	India	India	105.476	24.748	-31.084	-28.655	12.942
SC0473	PI534028	Durra	0	1	India	India	96.431	26.436	-18.619	-30.481	-4.933
SC0475	PI534022	Durra	0	1	India	India	105.711	23.811	-27.829	-32.596	9.225
SC0477	PI534026	Durra	1	1	India	India	107.361	20.568	-24.043	-33.570	11.193
SC0479	NA	Durra	1	1	India	India	100.331	29.733	-24.401	-32.889	2.917
SC0480	PI656097	Durra	0	1	India	India	111.528	29.432	-33.952	-34.221	7.542
SC0482	PI533859	Durra	0	1	India	India	113.682	29.067	-30.322	-34.480	3.234
SC0483	PI534034	Durra	1	1	India	India	110.708	32.518	-24.460	-32.170	-5.809
SC0484	PI534036	Durra	0	1	India	India	94.976	18.996	-30.476	-1.450	7.848
SC0485	NA	Durra	0	NA	India	India	62.900	5.953	36.873	-19.759	-4.783
SC0489	PI533856	Durra	1	1	India	India	110.026	30.565	-23.895	-32.893	-8.663
SC0490	PI534024	Durra	0	0	India	India	96.242	18.550	-27.491	2.356	2.812
SC0491	PI534025	Durra	1	1	India	India	111.713	25.529	-28.460	-33.762	-2.463
SC0492	PI534027	Durra	0	1	India	India	108.891	30.440	-25.428	-31.434	-1.789
SC0493	PI533857	Durra	0	1	India	India	113.427	31.293	-30.622	-31.329	-2.285

Table S4.1. continued

SC0494	PI534030	Durra	0	0	India	India	109.850	31.801	-24.551	-32.621	-5.503
SC0496	PI534032	Durra	0	1	India	India	107.338	29.891	-19.823	-33.732	-0.890
SC0497	PI534033	Durra	0	1	India	India	106.381	28.507	-22.051	-29.063	-11.624
SC0498	PI656099	Durra	0	1	India	India	117.228	35.018	-29.654	-33.294	-5.418
SC0499	PI533858	Durra	1	1	India	India	110.242	24.052	-29.650	-35.210	6.850
SC0500	PI656100	Durra	0	1	India	India	107.276	28.454	-23.043	-29.026	-10.566
SC0501	PI533860	Caudatum	0	0	India	India	-5.904	34.948	-26.757	-3.764	-9.437
SC0502	PI533996	Caudatum	0	0	Sudan	Northern-Africa	-23.753	19.528	-17.129	-8.109	16.706
SC0504	PI533864	Caudatum	0	0	Sudan	Northern-Africa	-45.094	29.300	-16.632	-0.193	13.719
SC0505	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	40.326	-1.008	-1.266	133.360	-2.833
SC0508	PI533999	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	41.286	-6.792	-6.808	141.787	1.265
SC0512	PI534008	Guinea	0	0	India	India	32.026	-13.545	32.969	-17.141	61.706
SC0514	PI534019	Guinea	0	0	India	India	18.716	-17.805	51.959	-9.091	62.686
SC0515	PI534012	Guinea	0	NA	India	India	-0.153	-26.026	59.564	-12.768	78.453
SC0516	PI576408	Guinea	0	0	India	India	16.139	-19.874	43.888	-10.022	65.084
SC0517	NA	Guinea	0	0	Ghana	Western-Africa	-9.148	-50.458	18.686	-9.933	35.791
SC0519	PI534078	Guinea	0	1	Nigeria	Western-Africa	-32.523	-107.541	-39.237	-6.944	-18.293
SC0520	PI534080	Guinea	0	1	Nigeria	Western-Africa	-34.975	-114.502	-32.816	-10.194	-2.379
SC0521	NA	Guinea	0	0	Nigeria	Western-Africa	-25.865	-98.973	-10.608	-11.836	-6.395
SC0522	PI534081	Guinea	0	0	Nigeria	Western-Africa	-38.167	-113.798	-43.832	-12.300	-0.315
SC0523	PI534083	Guinea	0	1	Nigeria	Western-Africa	-39.234	-115.649	-41.853	-12.944	2.668
SC0525	PI656101	Guinea	0	0	Nigeria	Western-Africa	-32.396	-103.638	-29.686	-14.914	-6.500
SC0526	PI534086	Guinea	0	0	Nigeria	Western-Africa	-32.425	-105.270	-32.166	-15.536	-2.228
SC0527	PI534087	Guinea	0	0	Nigeria	Western-Africa	-31.059	-97.422	-32.896	-11.974	-7.816
SC0528	PI533889	Guinea	0	0	Nigeria	Western-Africa	-31.527	-95.177	-42.596	-8.390	-21.755
SC0529	PI534090	Guinea	0	0	Nigeria	Western-Africa	-35.084	-105.482	-31.409	-11.573	-2.765
SC0530	PI533892	Guinea	0	NA	Nigeria	Western-Africa	-33.420	-103.078	-35.932	-14.642	-2.323
SC0532	PI597951	Guinea	0	0	BurkinaFaso	Western-Africa	-23.614	-96.992	-29.153	-6.138	-30.683
SC0534	PI533886	Guinea	0	0	Nigeria	Western-Africa	-29.239	-112.802	-37.529	-11.117	-1.391
SC0536	PI534084	Guinea	0	0	Nigeria	Western-Africa	-34.168	-112.690	-37.859	-11.400	-0.158
SC0537	PI597954	Guinea	0	0	Nigeria	Western-Africa	-30.203	-108.255	-36.943	-6.934	-2.811
SC0538	PI534091	Guinea	0	0	Nigeria	Western-Africa	-21.561	-71.021	-16.785	-13.982	3.114

Table S4.1. continued

SC0540	NA	Guinea	0	0	Nigeria	Western-Africa	-32.104	-96.975	-9.858	-12.516	-4.059
SC0541	PI534094	Guinea	0	0	Nigeria	Western-Africa	-31.298	-113.396	-40.594	-12.923	-1.007
SC0542	NA	Guinea	0	0	Nigeria	Western-Africa	-14.670	-20.335	28.721	-11.706	-6.634
SC0543	PI534071	Guinea	0	0	Nigeria	Western-Africa	-26.495	-107.453	-32.340	-13.646	-6.244
SC0544	PI534069	Guinea	0	0	Nigeria	Western-Africa	-33.826	-114.587	-35.245	-10.542	-1.594
SC0545	PI534072	Guinea	0	0	Nigeria	Western-Africa	-25.497	-105.701	-32.426	-11.984	-4.831
SC0546	PI534077	Guinea	0	0	Nigeria	Western-Africa	-34.623	-105.681	-33.414	-12.345	0.704
SC0547	NA	Guinea	0	0	Nigeria	Western-Africa	-27.793	-95.264	-26.976	-11.961	-0.538
SC0549	PI533840	Guinea	0	0	Nigeria	Western-Africa	-34.609	-110.181	-41.253	-13.320	-0.395
SC0550	NA	Guinea	0	0	BurkinaFaso	Western-Africa	-4.547	-56.914	-0.488	23.460	24.364
SC0551	NA	Guinea	0	0	Nigeria	Western-Africa	-34.797	-112.677	-46.434	-14.457	-1.199
SC0553	PI534063	Durra	0	0	Nigeria	Western-Africa	-33.123	-97.606	-40.607	-5.788	-20.813
SC0556	NA	Kafir	0	0	Bechuanaland	Unknown	7.504	-5.210	107.045	-12.588	-38.675
SC0557	PI533939	Caudatum	1	0	Mozambique	Eastern-Africa	-25.936	-9.642	26.809	0.200	20.649
SC0558	PI533938	Caudatum	0	0	Zaire	Middle-Africa	-16.210	-10.963	37.317	2.343	7.663
SC0559	PI534001	Caudatum	0	NA	Ethiopia	Eastern-Africa	-51.583	45.508	-9.649	0.594	-2.818
SC0562	PI533987	Caudatum	0	NA	Sudan	Northern-Africa	-37.987	15.944	-24.187	-5.858	18.486
SC0563	PI533876	Caudatum	0	0	Nigeria	Western-Africa	-27.898	-1.420	-12.491	-6.433	14.377
SC0564	PI534053	Caudatum	0	0	Uganda	Eastern-Africa	-37.041	17.673	-19.105	-3.789	16.223
SC0565	PI534055	Caudatum	0	0	Niger	Western-Africa	-24.617	-51.494	3.853	-9.168	31.454
SC0566	PI533871	Caudatum	0	0	Nigeria	Western-Africa	-36.632	-82.283	-32.984	-10.106	5.808
SC0567	PI534093	Caudatum	0	0	Nigeria	Western-Africa	-22.719	-79.791	-26.804	-14.061	2.144
SC0568	PI576343	Caudatum	0	0	Nigeria	Western-Africa	-22.204	-76.939	-20.420	-11.705	-8.201
SC0569	PI534092	Caudatum	0	0	Nigeria	Western-Africa	-31.252	-70.207	-32.810	-15.120	-34.761
SC0572	PI533980	Caudatum	0	0	China	Asia	-33.202	32.114	-18.589	3.050	-3.450
SC0574	PI534114	Caudatum	0	0	Pakistan	Asia	-61.860	38.595	-14.852	-4.717	23.643
SC0575	PI533992	Caudatum	0	0	Sudan	Northern-Africa	-27.319	27.665	-20.409	3.134	-2.678
SC0577	PI533978	Caudatum	0	0	Sudan	Northern-Africa	-25.884	27.081	-21.150	1.181	-6.980
SC0578	PI533870	Caudatum	0	0	Nigeria	Western-Africa	-23.619	-61.062	-22.370	0.322	-18.161
SC0580	PI534006	Durra	1	0	India	India	107.140	20.994	-32.247	-40.270	32.022
SC0581	NA	Durra	0	0	India	India	81.942	11.296	-13.486	-29.852	21.397
SC0582	NA	Durra-Bicolor	0	0	India	India	42.094	0.163	31.717	-6.156	23.206

Table S4.1. continued

SC0584	PI534015	Durra	0	1	India	India	106.548	30.149	-20.159	-33.341	-2.094
SC0586	PI534035	Durra	1	1	India	India	111.723	32.598	-27.838	-33.321	-0.538
SC0587	PI534021	Durra	1	1	India	India	112.641	21.969	-36.766	-33.589	12.859
SC0589	PI534023	Durra	1	1	India	India	115.445	24.332	-31.424	-34.557	4.107
SC0590	PI534102	Caudatum	0	0	Uganda	Eastern-Africa	-15.107	-2.748	20.511	6.056	11.230
SC0593	PI576344	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	16.593	-4.716	18.913	89.624	-2.463
SC0598	PI576337	Unknown/Other	0	0	Uganda	Eastern-Africa	-28.868	13.357	47.445	-17.128	-3.309
SC0599	PI534163	Caudatum	0	0	USA	Americas	-58.530	37.530	5.712	-8.562	18.258
SC0600	PI576336	Guinea	0	0	Sudan	Northern-Africa	-36.386	-43.805	-21.282	-9.084	15.290
SC0601	PI534103	Guinea	0	0	Uganda	Eastern-Africa	-33.437	-3.333	34.331	-7.805	36.178
SC0602	NA	Guinea	0	0	Uganda	Eastern-Africa	-2.152	8.327	79.783	-7.859	-24.834
SC0603	PI533936	Guinea	0	0	Tanzania	Eastern-Africa	-0.755	-28.496	90.175	-11.654	8.921
SC0604	NA	Guinea	0	0	Africa	Eastern-Africa	-24.429	-45.069	10.278	-9.902	17.284
SC0605	PI534096	Guinea	0	1	Kenya	Eastern-Africa	-7.148	-49.950	2.406	23.279	23.822
SC0606	PI597946	Guinea-Bicolor	0	0	China	Asia	-13.867	23.957	26.814	4.564	-17.617
SC0609	PI576332	Bicolor	0	0	China	Asia	4.701	-20.104	99.281	-7.473	18.365
SC0610	PI656103	Bicolor	0	0	China	Asia	57.444	-1.341	-17.601	-6.842	26.267
SC0614	PI533940	Bicolor	0	0	Tanzania	Eastern-Africa	12.109	-21.362	90.479	-15.348	6.279
SC0615	NA	Bicolor-Kafir	0	0	India	India	41.559	1.614	49.300	-9.558	-8.348
SC0616	NA	Bicolor	0	0	India	India	65.563	-5.277	-5.458	-26.662	24.481
SC0618	NA	Bicolor	0	1	Burma	Asia	71.262	-1.924	1.968	-27.001	36.155
SC0620	PI576341	Bicolor	0	1	S.Africa	Southern-Africa	8.245	-32.122	35.841	-1.576	26.499
SC0621	PI656104	Bicolor	NA	1	India	India	63.529	-1.813	2.109	-16.568	33.906
SC0623	PI533956	Durra	0	0	Congo	Middle-Africa	-6.252	-5.278	45.445	1.222	-0.870
SC0624	PI576366	Durra	0	0	India	India	52.560	4.971	1.385	-10.359	20.755
SC0625	PI534097	Kafir	0	1	Japan	Asia	16.349	-11.260	100.444	-8.747	-11.440
SC0626	PI534098	Kafir	0	0	Japan	Asia	7.743	-20.783	104.419	-9.017	2.468
SC0627	PI576345	Kafir	0	0	S.Africa	Southern-Africa	-4.416	-6.457	115.072	-10.654	-16.960
SC0628	PI533979	Kafir	0	0	S.Africa	Southern-Africa	8.070	-12.367	123.771	-10.002	-27.493
SC0629	PI597947	Kafir	0	0	Botswana	Southern-Africa	-8.881	-11.678	110.921	-12.863	-3.426
SC0630	PI533937	Kafir	0	0	Zambia	Eastern-Africa	-9.035	-10.377	122.502	-14.039	-7.833
SC0631	NA	Kafir	0	0	Bechuanaland	Unknown	0.478	-16.257	115.654	-15.839	-5.097

Table S4.1. continued

SC0632	PI576400	Kafir	0	0	Unknown	Unknown	1.039	-20.199	137.562	-19.539	-19.157
SC0634	PI533929	Kafir	0	0	Mexico	Americas	8.708	-8.352	101.480	-13.390	8.405
SC0635	PI576329	Kafir	0	0	USA	Americas	1.334	-17.707	134.345	-18.410	-13.815
SC0636	NA	Kafir-Caudatum	0	0	India	India	61.557	5.686	40.467	-24.695	-24.319
SC0637	PI534105	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-52.656	44.836	-2.512	-2.232	-2.828
SC0639	PI656105	Kafir-Caudatum	0	0	India	India	-40.451	35.843	24.765	-4.211	-9.975
SC0640	NA	Kafir-Caudatum	NA	0	India	India	-44.184	40.704	21.196	-6.974	-8.505
SC0641	PI534104	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-64.451	57.843	-14.196	-3.782	5.837
SC0642	PI534109	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-65.186	59.524	-12.666	-3.077	2.340
SC0643	PI534110	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-57.631	52.731	-1.427	-4.538	4.064
SC0644	PI534111	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-64.383	56.369	-19.623	-4.980	7.304
SC0645	PI534108	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-57.706	53.596	-1.329	-3.278	3.950
SC0646	PI534107	Kafir-Caudatum	0	0	Uganda	Eastern-Africa	-2.259	-9.020	108.425	-11.243	-23.741
SC0647	PI533952	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	-4.032	-11.317	114.347	-8.979	-16.283
SC0648	PI533955	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	1.095	-17.526	114.092	-9.699	-4.720
SC0649	PI576338	Kafir-Caudatum	0	0	Zimbabwe	Eastern-Africa	-1.380	-13.809	125.563	-11.053	-11.770
SC0650	PI576340	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	53.032	1.780	34.452	-21.247	18.865
SC0652	NA	Kafir-Caudatum	0	0	India	India	3.107	-11.718	124.463	-12.195	-18.389
SC0653	PI533953	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	-17.632	6.351	11.712	-7.626	-5.273
SC0654	PI533975	Kafir-Caudatum	0	0	Mozambique	Eastern-Africa	6.999	-13.107	123.498	-10.132	-23.106
SC0655	PI533976	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	-52.606	41.218	-12.381	1.655	5.214
SC0657	PI533982	Kafir-Caudatum	0	0	S.Africa	Southern-Africa	1.949	-13.306	122.270	-10.897	-5.842
SC0659	PI576333	Guinea-Kafir	0	0	USA	Americas	-1.789	-20.166	119.124	-11.063	0.729
SC0663	PI533948	Guinea-Kafir	0	0	USA	Americas	-3.568	-18.477	121.233	-11.586	-8.762
SC0671	PI534054	Kafir	0	0	Kenya	Eastern-Africa	-12.058	-24.451	80.073	1.578	31.006
SC0672	PI595702	Kafir-Caudatum	0	0	Zimbabwe	Eastern-Africa	-2.757	-12.431	110.411	-7.821	-1.375
SC0673	PI576339	Kafir	0	0	Zimbabwe	Eastern-Africa	-1.798	-16.848	117.962	-9.549	-0.119
SC0679	PI534046	Guinea-Caudatum	0	0	Sudan	Northern-Africa	-24.611	9.928	9.562	20.777	12.160
SC0680	PI576374	Caudatum	0	0	Tanzania	Eastern-Africa	-73.248	58.948	-23.807	-7.407	23.789
SC0681	PI533994	Caudatum	0	0	Sudan	Northern-Africa	-73.495	55.047	-13.024	-6.359	7.821
SC0682	PI576392	Caudatum	0	0	S.Africa	Southern-Africa	-1.743	-14.330	116.395	-14.827	-9.376
SC0683	PI533963	Caudatum	0	0	Sudan	Northern-Africa	-63.025	50.972	-10.238	-7.207	4.077

Table S4.1. continued

SC0685	NA	Caudatum	0	0	FrenchEquatorial	Middle-Africa	-57.886	49.359	-17.514	-0.653	-10.658
SC0686	PI534113	Caudatum	0	0	Uganda	Eastern-Africa	-74.137	61.616	-21.771	-7.690	11.664
SC0687	PI534038	Caudatum	0	0	Sudan	Northern-Africa	-77.065	57.825	-23.612	-5.364	4.090
SC0689	PI533969	Caudatum	0	0	Uganda	Eastern-Africa	-58.530	50.110	-14.776	-4.951	0.883
SC0690	PI533971	Caudatum	0	0	Uganda	Eastern-Africa	-73.414	61.113	-17.333	-7.383	10.286
SC0691	PI534050	Caudatum	0	0	Cent.Afr.Rep.	Middle-Africa	-59.585	43.438	-13.463	-4.925	3.818
SC0692	PI533966	Caudatum	0	0	Uganda	Eastern-Africa	-69.923	61.874	-14.391	-2.880	2.014
SC0693	PI534052	Caudatum	0	0	Uganda	Eastern-Africa	-78.134	61.575	-16.597	-9.786	12.856
SC0694	PI534056	Caudatum	0	0	Nigeria	Western-Africa	-68.830	52.364	-12.160	-5.514	5.668
SC0695	PI656106	Caudatum	0	0	Tanzania	Eastern-Africa	-65.337	44.054	-9.212	-8.910	25.387
SC0700	PI576346	Caudatum	0	0	S.Africa	Southern-Africa	-19.102	-21.118	73.402	-14.381	19.345
SC0701	PI533985	Caudatum	0	0	Sudan	Northern-Africa	-60.475	41.503	-26.987	-4.282	20.948
SC0702	PI656107	Caudatum	0	0	Sudan	Northern-Africa	-38.163	29.912	-15.758	-1.752	19.497
SC0704	PI534099	Caudatum	0	0	Japan	Asia	-67.760	48.138	-24.118	-6.205	22.626
SC0705	PI576420	Caudatum	0	0	Japan	Asia	-59.327	36.815	-14.531	-4.206	22.388
SC0706	PI659693	Caudatum	0	0	Sudan	Northern-Africa	-66.525	56.020	-17.520	-3.656	0.891
SC0707	PI533959	Caudatum	0	0	Unknown	Unknown	-58.781	49.744	-13.738	-5.984	6.829
SC0708	PI533970	Caudatum	0	0	Uganda	Eastern-Africa	-63.105	55.566	-23.672	-3.185	13.408
SC0709	PI533990	Caudatum	0	0	Sudan	Northern-Africa	-20.421	19.221	37.587	-4.599	-21.117
SC0712	PI576372	Caudatum	0	0	Japan	Asia	-68.851	45.867	-23.530	-4.708	14.881
SC0713	NA	Caudatum	0	0	Sudan	Northern-Africa	-61.546	49.584	-22.443	-3.936	-2.682
SC0715	PI534039	Caudatum	0	0	Sudan	Northern-Africa	-68.239	57.876	-23.320	-5.944	1.564
SC0716	PI576355	Caudatum	0	0	Sudan	Northern-Africa	-55.828	42.943	-24.229	-6.849	11.684
SC0719	PI534047	Caudatum	0	0	Sudan	Northern-Africa	-45.569	38.958	-16.618	-0.369	-8.964
SC0720	PI659696	Caudatum	0	0	Kenya	Eastern-Africa	-47.314	28.189	35.846	-9.822	3.208
SC0721	PI534100	Caudatum	0	0	Japan	Asia	-20.365	7.392	66.601	-13.083	-22.947
SC0723	PI533988	Caudatum	0	0	Sudan	Northern-Africa	-76.805	61.118	-20.773	-4.485	1.686
SC0724	PI533993	Caudatum	0	0	Sudan	Northern-Africa	-61.027	50.213	-32.467	-1.928	5.947
SC0725	PI534101	Caudatum	0	0	Japan	Asia	-52.456	38.311	-16.555	-6.552	12.587
SC0726	PI534051	Caudatum	0	0	Cent.Afr.Rep.	Middle-Africa	-68.554	58.317	-22.843	-5.103	4.728
SC0727	PI534045	Caudatum	0	0	Sudan	Northern-Africa	-65.195	54.606	-15.975	-3.702	3.790
SC0728	PI533968	Caudatum	0	0	Uganda	Eastern-Africa	-45.197	43.361	-14.919	1.271	-7.888

Table S4.1. continued

SC0730	PI576411	Caudatum	0	0	Sudan	Northern-Africa	-60.679	53.784	-31.933	-0.820	3.980
SC0731	PI576356	Caudatum	0	0	Sudan	Northern-Africa	-63.340	53.781	-24.093	-6.096	3.451
SC0732	PI534049	Caudatum	0	0	Sudan	Northern-Africa	-70.738	59.231	-20.552	-5.656	4.756
SC0733	PI534061	Caudatum	0	0	Nigeria	Western-Africa	-56.038	45.472	-30.004	2.782	9.194
SC0734	PI576394	Caudatum	0	0	Sudan	Northern-Africa	-7.544	-4.882	105.241	-14.820	-22.459
SC0736	PI534040	Caudatum	0	0	Sudan	Northern-Africa	-19.822	19.032	52.031	-2.068	-11.472
SC0737	PI533983	Caudatum	0	0	Sudan	Northern-Africa	-71.544	56.401	-26.591	-3.004	9.518
SC0738	PI597952	Caudatum	0	1	Sudan	Northern-Africa	-47.665	38.791	-6.814	-6.478	-0.886
SC0741	NA	Caudatum	0	0	Nigeria	Western-Africa	86.521	20.168	-4.167	-21.864	-17.887
SC0747	NA	Caudatum	0	0	via(Mexico)	Americas	-25.337	29.729	29.222	-6.702	-3.256
SC0748	PI533991	Caudatum	0	0	Sudan	Northern-Africa	-53.013	52.481	-4.289	-4.306	-10.589
SC0749	PI576373	Caudatum-Bicolor	0	0	Japan	Asia	-4.146	19.166	46.009	-0.854	-36.789
SC0751	PI576354	Caudatum	0	0	Sudan	Northern-Africa	-9.501	-19.651	58.227	-4.938	39.400
SC0753	NA	Caudatum-Kafir	0	0	India	India	-18.940	17.409	50.483	-4.340	-23.736
SC0755	PI576350	Caudatum	0	1	USA	Americas	5.895	-11.025	42.870	11.673	12.164
SC0756	PI576368	Caudatum	0	0	Sudan	Northern-Africa	-52.276	52.831	-25.366	2.555	-2.586
SC0757	PI576352	Caudatum	0	0	Botswana	Southern-Africa	-29.295	16.180	65.134	-6.747	-4.610
SC0759	NA	Caudatum-Kafir	0	0	via(Mexico)	Americas	-37.091	25.921	5.609	-4.647	7.587
SC0760	PI533949	Caudatum	0	0	Sudan	Northern-Africa	-53.321	46.112	-15.824	-0.929	-2.428
SC0761	PI533928	Caudatum	0	0	Mexico	Americas	3.885	-15.354	110.494	-10.034	-9.067
SC0762	PI576363	Caudatum	0	0	India	India	-66.979	59.263	-24.867	-1.448	3.788
SC0763	PI534041	Caudatum	0	0	Sudan	Northern-Africa	-49.862	39.010	-1.414	-2.810	-1.778
SC0764	PI533974	Caudatum	0	0	Mozambique	Eastern-Africa	-55.814	47.743	3.253	-6.600	2.069
SC0770	PI576328	Caudatum	0	0	Swaziland	Southern-Africa	-49.163	25.871	11.582	-5.574	-5.842
SC0773	PI576351	Caudatum	0	0	China	Asia	-62.825	58.155	-24.796	-2.344	2.407
SC0774	PI576353	Caudatum	0	0	Sudan	Northern-Africa	-58.566	55.527	-15.121	-1.239	0.278
SC0777	NA	Caudatum-Kafir	0	0	Sudan	Northern-Africa	-52.004	44.830	-0.610	-3.939	-3.128
SC0779	PI576413	Caudatum	0	0	Sudan	Northern-Africa	-11.297	4.296	75.868	-7.184	-19.077
SC0780	PI533984	Caudatum	0	0	Sudan	Northern-Africa	-55.339	46.060	-23.617	0.877	7.915
SC0781	PI576371	Caudatum	0	0	Nigeria	Western-Africa	-60.994	56.700	-1.839	-3.727	1.032
SC0782	PI576364	Caudatum	0	0	India	India	1.417	-10.163	101.282	-11.698	2.095
SC0784	PI576369	Caudatum	0	0	Cent.Afr.Rep.	Middle-Africa	-60.685	40.450	-18.683	-4.405	8.052

Table S4.1. continued

SC0790	PI656120	Caudatum	0	0	India	India	-47.787	46.873	1.687	-9.828	-17.319
SC0797	PI576397	Caudatum	0	1	Zimbabwe	Eastern-Africa	-5.817	-12.976	84.922	-11.750	9.490
SC0798	PI533989	Caudatum	0	0	Sudan	Northern-Africa	-67.998	55.919	-21.336	-6.135	7.647
SC0800	PI595709	Caudatum	0	0	Sudan	Northern-Africa	-75.300	59.004	-24.816	-5.515	3.902
SC0803	PI533964	Caudatum	0	0	Sudan	Northern-Africa	-75.349	61.207	-28.331	-5.201	8.515
SC0804	PI534043	Caudatum	0	0	Sudan	Northern-Africa	-66.977	57.786	-22.917	-6.134	6.998
SC0805	PI533967	Caudatum	1	0	Uganda	Eastern-Africa	-61.562	47.713	-11.589	-6.997	14.398
SC0807	PI576370	Caudatum	0	0	Zimbabwe	Eastern-Africa	4.157	14.869	6.296	3.744	-23.680
SC0808	PI597944	Caudatum-Durra	0	0	Uganda	Eastern-Africa	-30.699	31.147	11.278	-1.145	-17.188
SC0810	PI533930	Caudatum	0	0	Sudan	Northern-Africa	-60.915	52.100	-7.751	-2.135	2.879
SC0814	NA	Durra	0	0	via(Mexico)	Americas	62.087	3.904	21.052	-24.690	-3.238
SC0817	PI576330	Durra	0	0	India	India	90.006	15.897	-30.678	-33.328	16.817
SC0819	PI595700	Durra	1	0	India	India	91.318	13.762	-27.167	-36.409	30.489
SC0821	PI595701	Kafir	0	0	S.Africa	Southern-Africa	1.673	-18.650	110.720	-14.331	10.667
SC0823	NA	Durra	0	0	Mozambique	Eastern-Africa	40.862	-2.019	-4.751	144.934	4.247
SC0826	PI576402	Durra	0	0	India	India	85.529	17.958	-30.432	-33.531	28.551
SC0827	PI576403	Durra	0	0	India	India	81.975	19.762	-24.493	-26.289	17.674
SC0830	PI595704	Durra	0	0	India	India	77.574	13.243	-20.607	-24.077	17.275
SC0831	PI576404	Durra	0	0	India	India	94.036	14.886	-26.905	-32.912	24.764
SC0832	PI534005	Durra	0	0	India	India	102.509	22.575	-33.446	-38.681	19.604
SC0833	PI656108	Durra	0	0	India	India	95.567	14.238	-22.446	-35.517	31.091
SC0834	NA	Durra	0	0	India	India	84.333	16.574	-27.160	-25.311	16.290
SC0835	PI534007	Durra	0	0	India	India	98.255	12.617	-24.070	-36.378	27.398
SC0837	NA	Durra	0	0	India	India	77.698	2.761	8.952	-31.022	3.888
SC0839	PI576406	Durra	0	0	India	India	96.902	13.934	-27.943	-38.165	24.075
SC0841	PI595708	Durra	0	0	India	India	41.906	-0.821	36.913	-18.397	16.414
SC0842	PI576367	Durra	0	0	India	India	2.650	-15.934	112.928	-8.170	-3.756
SC0847	NA	Durra	0	0	India	India	88.885	17.080	-23.772	-33.897	18.506
SC0848	NA	Durra	0	0	India	India	100.366	15.282	-34.368	-37.805	28.763
SC0851	PI576342	Durra	0	0	Egypt	Northern-Africa	22.062	-4.050	34.577	-4.033	-12.553
SC0852	PI534000	Durra	0	0	Ethiopia	Eastern-Africa	47.955	5.595	-19.927	-2.821	-7.863
SC0854	NA	Durra	0	0	India	India	58.464	6.899	23.729	-21.898	-2.032

Table S4.1. continued

SC0855	PI597945	Durra	0	0	Egypt	Northern-Africa	32.852	11.306	-2.216	11.584	-28.950
SC0858	NA	Durra	0	0	India	India	78.508	19.609	12.391	-23.090	-25.303
SC0859	PI534003	Durra	1	0	India	India	96.569	16.813	-28.772	-28.921	10.376
SC0860	NA	Durra	0	0	India	India	91.223	24.043	-23.466	-17.674	-12.793
SC0863	PI576360	Durra	0	0	India	India	91.287	27.073	-20.848	-21.050	-4.959
SC0864	NA	Durra	0	0	India	India	67.891	13.409	12.511	-19.143	-15.223
SC0865	PI534004	Durra	0	0	India	India	71.178	18.797	-11.068	-19.504	-3.864
SC0868	NA	Durra	0	0	India	India	-7.994	1.142	30.834	-8.949	-6.720
SC0870	NA	Durra	0	0	India	India	84.241	19.166	4.420	-25.534	-14.890
SC0871	NA	Durra	0	0	India	India	96.455	26.743	-5.681	-23.288	-22.354
SC0875	PI576358	Durra	0	0	India	India	91.089	26.030	-19.246	-25.645	-2.420
SC0876	PI595706	Durra	0	0	India	India	106.268	29.363	-27.483	-29.377	2.974
SC0877	NA	Durra	0	0	India	India	50.449	8.749	51.152	-14.859	-38.857
SC0878	NA	Durra	0	0	India	India	75.423	18.757	5.335	-16.449	-18.177
SC0888	PI659753	Durra	0	1	India	India	101.350	22.029	-29.375	-28.593	6.131
SC0891	PI595707	Durra	0	NA	India	India	100.055	28.019	-29.087	-27.558	-3.663
SC0893	NA	Durra	0	0	India	India	59.379	11.094	51.368	-18.093	-36.602
SC0895	NA	Durra	0	0	India	India	80.377	21.571	8.301	-23.917	-22.421
SC0899	PI534020	Durra	0	0	India	India	72.623	16.575	-31.284	-21.297	25.744
SC0902	PI576365	Durra	0	0	India	India	87.618	26.457	-20.863	-19.324	-8.278
SC0905	PI534068	Durra	0	0	Nigeria	Western-Africa	95.089	30.034	-18.989	-27.406	-4.506
SC0906	PI534048	Caudatum	0	0	Sudan	Northern-Africa	-46.961	62.176	-22.645	-1.614	-7.378
SC0910	PI576359	Guinea-Durra	0	1	India	India	64.992	6.574	-0.718	-23.326	34.898
SC0913	PI576410	Durra-Bicolor	0	0	Sudan	Northern-Africa	-35.201	18.147	36.665	-1.782	3.758
SC0914	NA	Durra-Bicolor	0	0	India	India	-51.097	53.796	-13.064	3.121	-3.492
SC0919	PI576407	Durra	0	0	India	India	95.988	27.363	-22.739	-22.562	-9.931
SC0921	NA	Durra	0	1	India	India	69.583	2.917	17.386	-30.877	10.643
SC0923	NA	Durra	0	1	India	India	-12.800	1.564	9.970	-6.007	-0.595
SC0924	PI534029	Durra	NA	NA	India	India	98.025	24.869	-14.261	-30.957	-8.393
SC0929	PI595699	Durra	0	0	India	India	96.547	26.506	-22.833	-26.007	-8.363
SC0935	PI576388	Durra	0	NA	USA	Americas	36.746	22.003	-13.562	5.031	-56.942
SC0937	PI576348	Bicolor	0	1	USA	Americas	12.749	-13.496	17.451	10.310	23.693

Table S4.1. continued

SC0941	PI576347	Bicolor	0	1	USA	Americas	1.120	-24.266	50.914	1.949	40.858
SC0942	PI576349	Bicolor	0	1	USA	Americas	11.115	-13.374	13.599	13.698	25.278
SC0947	PI656121	Unknown/Other	0	0	India	India	-45.527	6.086	-11.375	-4.050	11.454
SC0949	PI533998	Bicolor	0	0	USA	Americas	-20.225	15.805	26.483	-1.311	3.228
SC0950	PI576389	Caudatum-Bicolor	0	0	USA	Americas	7.540	-20.786	89.698	-11.953	19.049
SC0951	PI576335	Guinea	0	0	Sudan	Northern-Africa	-8.305	-24.704	104.729	-10.125	18.907
SC0956	PI576416	Guinea	0	0	Nigeria	Western-Africa	-33.682	-69.008	-26.485	-10.509	12.148
SC0958	PI576361	Durra	0	0	India	India	68.997	30.133	-22.577	-9.649	-35.243
SC0963	PI533977	Caudatum	0	0	S.Africa	Southern-Africa	-64.146	43.119	-19.300	0.360	17.654
SC0964	PI533972	Caudatum	0	0	Uganda	Eastern-Africa	-81.822	62.134	-25.533	-6.212	26.188
SC0965	NA	Unknown/Other	0	0	Uganda	Eastern-Africa	28.863	-0.385	-11.640	17.522	12.276
SC0968	PI656110	Unknown/Other	0	0	Zimbabwe	Eastern-Africa	53.344	5.367	-8.625	69.037	-19.085
SC0969	PI534164	Durra	0	0	Uganda	Eastern-Africa	-44.989	34.043	-11.643	-10.261	15.070
SC0970	PI576386	Guinea-Caudatum	0	0	Uganda	Eastern-Africa	-55.071	38.975	-17.230	-4.243	23.411
SC0971	PI656111	Unknown/Other	1	1	PuertoRico	Americas	-3.542	-11.496	10.789	-7.695	17.177
SC0972	PI534165	Caudatum	0	0	Sudan	Northern-Africa	-75.453	64.198	-30.219	-4.052	2.080
SC0975	PI576379	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	40.109	-0.809	-2.966	148.530	1.671
SC0979	PI576428	Caudatum	0	0	Ethiopia	Eastern-Africa	-74.359	62.731	-15.641	-8.883	-6.970
SC0982	PI576380	Caudatum	0	0	Ethiopia	Eastern-Africa	-66.655	60.899	-25.716	0.767	-7.189
SC0984	PI534115	Caudatum	0	0	Ethiopia	Eastern-Africa	-80.869	67.741	-27.974	-8.309	-2.297
SC0987	PI534116	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	32.616	-0.496	6.450	78.985	-15.251
SC0991	PI534117	Bicolor	0	0	Uganda	Eastern-Africa	-19.782	-15.055	38.834	10.361	-3.815
SC0997	PI534166	Unknown/Other	0	0	Israel	Arabia/MiddleEast	-33.831	15.990	48.586	-13.398	-6.934
SC0998	PI534167	Durra-Bicolor	0	0	Unknown	Unknown	27.481	2.135	-8.401	-0.658	-8.665
SC0999	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	44.316	-7.250	-13.017	147.577	17.957
SC1014	PI576375	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	49.784	9.049	-10.389	14.385	-50.720
SC1015	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	52.749	17.596	-9.205	19.166	-50.901
SC1017	PI576376	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	59.498	21.472	-19.282	15.260	-51.504
SC1019	PI656071	Caudatum	0	0	Ethiopia	Eastern-Africa	-45.132	44.126	-18.634	-3.231	1.905
SC1021	NA	Durra	0	0	Ethiopia	Eastern-Africa	52.845	7.298	23.670	2.542	-24.673
SC1022	PI576377	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	53.060	-3.692	-10.585	83.709	12.924
SC1023	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	42.914	-0.139	26.560	50.460	-29.194

Table S4.1. continued

SC1024	NA	Durra	0	0	Ethiopia	Eastern-Africa	69.703	14.653	-7.727	6.567	-43.869
SC1025	PI576378	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	-2.261	23.767	-13.180	43.247	3.707
SC1031	PI597956	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	52.822	5.799	-13.043	2.813	-17.500
SC1033	PI576426	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	60.847	17.975	-12.064	22.198	-50.479
SC1038	PI576381	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	55.267	18.663	-14.866	7.427	-55.017
SC1039	PI576382	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	65.509	20.722	-26.401	13.131	-49.106
SC1040	PI595724	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	87.043	22.918	-19.007	-15.429	-9.374
SC1046	PI576423	Durra	0	0	Ethiopia	Eastern-Africa	69.620	21.785	-17.545	5.849	-58.479
SC1047	PI656072	Durra	0	0	Ethiopia	Eastern-Africa	67.263	16.291	-19.754	11.268	-57.285
SC1049	PI576424	Durra	0	0	Ethiopia	Eastern-Africa	66.525	17.349	-18.079	8.551	-58.052
SC1055	PI595739	Caudatum	0	0	Sudan	Northern-Africa	-58.826	59.479	-21.638	2.008	-11.340
SC1056	PI576387	Caudatum	0	0	Sudan	Northern-Africa	-69.863	57.363	-14.590	-3.031	0.309
SC1057	PI595740	Caudatum	0	0	Uganda	Eastern-Africa	-73.370	53.248	-18.172	-7.174	11.056
SC1058	NA	Unknown/Other	0	0	Cameroon	Middle-Africa	74.821	3.028	-9.081	-24.573	28.182
SC1063	PI595741	Guinea	0	0	Senegal	Western-Africa	-16.577	-53.576	27.706	-9.898	27.301
SC1065	PI597959	Caudatum	0	0	Senegal	Western-Africa	-72.593	54.655	-25.030	-6.730	16.534
SC1067	PI576432	Unknown/Other	0	0	Senegal	Western-Africa	-31.618	3.860	-19.121	0.807	1.658
SC1069	PI576384	Unknown/Other	0	0	Nigeria	Western-Africa	-9.922	-48.461	62.439	-9.831	-22.701
SC1070	PI576385	Unknown/Other	0	NA	Nigeria	Western-Africa	-6.510	-53.911	40.290	-6.598	-51.413
SC1072	PI576433	Kafir	0	0	Nigeria	Western-Africa	13.756	3.019	45.724	-17.182	5.694
SC1074	PI656073	Kafir	0	0	Nigeria	Western-Africa	-19.614	-57.244	17.545	-13.630	-10.650
SC1076	PI597960	Caudatum	0	0	Nigeria	Western-Africa	-23.227	-51.428	-23.221	-6.199	-30.336
SC1077	PI597961	Caudatum	0	0	Nigeria	Western-Africa	-60.945	59.921	-23.040	2.903	-5.661
SC1079	PI595714	Caudatum	0	0	Sudan	Northern-Africa	-66.665	51.437	-23.292	-4.233	7.721
SC1080	PI576422	Kafir	0	0	S.Africa	Southern-Africa	1.628	-15.267	128.113	-10.113	-9.034
SC1082	NA	Unknown/Other	1	0	Nigeria	Western-Africa	-8.639	-39.448	31.823	12.687	-5.105
SC1083	PI595729	Guinea	0	0	Nigeria	Western-Africa	-16.173	-50.166	10.262	18.365	8.500
SC1084	PI595730	Guinea	0	0	Nigeria	Western-Africa	-17.864	-45.881	28.742	6.616	34.693
SC1085	PI576401	Durra	0	0	India	India	87.262	25.513	-21.831	-16.932	-7.277
SC1088	PI597948	Durra	NA	NA	India	India	-45.839	35.320	-10.891	-6.395	21.759
SC1089	PI595705	Durra	0	1	India	India	75.074	28.326	-23.898	-15.585	-6.536
SC1097	NA	Durra	0	1	India	India	58.701	11.816	55.177	-16.125	-51.048

Table S4.1. continued

SC1101	PI595703	Durra-Kafir	0	1	India	India	88.971	24.851	-21.767	-18.193	-11.865
SC1103	PI576434	Bicolor	0	0	Nigeria	Western-Africa	-55.011	29.927	-23.046	-4.555	24.367
SC1104	PI576435	Bicolor	0	0	Uganda	Eastern-Africa	-49.612	51.962	2.692	-1.043	-6.446
SC1107	NA	Bicolor	0	0	Japan	Asia	33.082	-9.537	40.481	-7.290	22.563
SC1108	PI597949	Guinea	0	0	India	India	0.448	-27.985	59.170	-10.491	83.679
SC1109	PI576405	Guinea-Bicolor	0	1	India	India	-13.085	-20.965	73.162	5.689	42.575
SC1111	PI576409	Caudatum	0	0	Sudan	Northern-Africa	-43.268	29.257	-15.802	1.298	-0.363
SC1114	NA	Caudatum	0	0	Sudan	Northern-Africa	-19.206	4.660	16.844	-9.798	-12.546
SC1116	PI576414	Caudatum	0	0	Nigeria	Western-Africa	-31.272	12.812	47.046	-12.978	-13.860
SC1117	PI595710	Caudatum	0	0	Nigeria	Western-Africa	-35.029	-8.281	-10.549	-7.535	-2.859
SC1118	PI595711	Guinea	0	1	Nigeria	Western-Africa	-23.141	-88.952	-32.479	19.464	7.956
SC1119	NA	Guinea	0	0	Nigeria	Western-Africa	-37.228	-77.052	-27.014	-15.055	-1.150
SC1120	NA	Caudatum	0	0	Nigeria	Western-Africa	-19.910	-0.871	-19.848	-2.842	0.558
SC1123	PI576417	Guinea	0	0	Nigeria	Western-Africa	-24.733	-80.579	-7.424	-9.704	28.171
SC1124	PI576418	Guinea	0	0	Nigeria	Western-Africa	-39.314	-113.355	-48.181	-13.018	3.197
SC1125	PI576419	Guinea	0	0	Nigeria	Western-Africa	-38.564	-109.877	-43.297	-12.867	1.631
SC1133	NA	Caudatum	0	0	Kenya	Eastern-Africa	-27.279	0.471	21.017	-7.450	13.001
SC1154	PI595720	Durra-Bicolor	NA	0	Ethiopia	Eastern-Africa	63.500	18.969	-27.727	9.307	-53.157
SC1155	PI576425	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	62.672	21.370	-31.645	13.474	-51.715
SC1156	PI595721	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	61.804	15.199	-22.066	12.093	-45.269
SC1157	PI595722	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	62.981	15.425	-19.920	20.042	-46.087
SC1158	PI597957	Durra	0	0	Ethiopia	Eastern-Africa	58.842	7.334	-15.718	8.913	-50.063
SC1159	PI595723	Durra	0	0	Ethiopia	Eastern-Africa	60.119	18.915	-25.291	6.448	-41.627
SC1160	PI576427	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	22.715	0.979	20.999	99.194	-15.066
SC1166	NA	Unknown/Other	0	0	Ethiopia	Eastern-Africa	52.902	5.245	7.663	13.722	-43.841
SC1170	NA	Durra	0	0	India	India	54.408	5.274	37.568	-2.589	-42.593
SC1172	PI576436	Caudatum	0	0	USA	Americas	-39.283	31.574	32.504	-7.446	0.779
SC1177	PI595742	Bicolor	0	0	Ethiopia	Eastern-Africa	33.347	-7.041	0.366	138.575	16.526
SC1178	NA	Unknown/Other	0	0	Nigeria	Western-Africa	-43.971	21.878	-4.062	0.116	-9.916
SC1179	NA	Unknown/Other	0	NA	Nigeria	Western-Africa	77.628	15.992	-2.234	-28.523	-7.788
SC1184	PI595719	Durra-Bicolor	0	0	Ethiopia	Eastern-Africa	-3.137	-4.756	48.825	-0.116	-39.352
SC1186	PI595731	Kafir	0	0	Sudan	Northern-Africa	47.916	3.717	-14.691	0.008	-12.137

Table S4.1. continued

SC1193	NA	Durra	0	0	Ethiopia	Eastern-Africa	58.736	12.547	-2.813	-14.623	-15.440
SC1201	PI595743	Guinea-Caudatum	0	0	Unknown	Unknown	-2.708	-9.780	54.422	-4.022	-28.635
SC1203	PI576437	Bicolor	0	0	Brazil	Americas	-35.417	19.385	14.692	2.806	-4.001
SC1205	PI597965	Caudatum	0	0	Senegal	Western-Africa	-72.588	58.809	-31.525	-6.619	4.753
SC1211	PI595744	Kafir-Caudatum	0	0	Guatemala	Americas	-20.467	4.618	22.542	0.067	-0.506
SC1212	PI597966	Caudatum	0	0	Venezuela	Americas	-73.517	61.711	-9.200	-8.928	7.516
SC1214	PI595745	Guinea-Caudatum	0	0	BurkinaFaso	Western-Africa	-45.664	40.356	-20.721	-3.588	9.180
SC1215	PI656112	Unknown/Other	0	0	Niger	Western-Africa	11.826	6.649	48.042	-1.225	-40.615
SC1216	NA	Unknown/Other	0	0	Niger	Western-Africa	7.142	-13.296	61.472	-6.580	54.878
SC1218	PI656074	Unknown/Other	0	0	Unknown	Unknown	-18.783	11.585	-8.922	7.523	-13.003
SC1222	PI595713	Caudatum	0	0	Japan	Asia	-35.611	20.548	-18.461	7.372	12.997
SC1229	PI576415	Guinea	0	0	Nigeria	Western-Africa	-35.303	-68.968	-29.678	-11.176	9.151
SC1237	PI597955	Caudatum-Kafir	0	0	Uganda	Eastern-Africa	-24.718	-3.469	19.735	3.611	8.112
SC1246	PI595718	Kafir-Caudatum	0	0	Chad	Middle-Africa	-33.455	32.152	-12.751	-0.658	-3.881
SC1251	PI656075	Caudatum	0	0	Sudan	Northern-Africa	-64.150	46.990	-3.602	-7.649	3.868
SC1261	PI595715	Caudatum	0	0	Sudan	Northern-Africa	-68.154	58.143	-23.825	-6.093	1.526
SC1262	PI595716	Caudatum	0	0	Sudan	Northern-Africa	-72.053	60.327	-18.760	-7.074	4.967
SC1263	PI595717	Caudatum	0	0	Sudan	Northern-Africa	-54.399	54.469	-20.539	-2.978	-4.203
SC1271	PI656076	Caudatum	0	0	Ethiopia	Eastern-Africa	-80.129	65.158	-21.644	-6.858	-6.347
SC1277	PI656113	Durra	0	0	India	India	77.466	25.626	-15.575	-7.884	-32.648
SC1287	PI595746	Caudatum	0	1	India	India	-73.022	62.921	-11.846	-5.118	-5.295
SC1293	PI595747	Caudatum	0	0	Senegal	Western-Africa	-53.291	11.761	-19.695	-8.063	-8.686
SC1300	PI595732	Caudatum	0	0	Ethiopia	Eastern-Africa	-80.286	67.486	-23.191	-6.396	-2.190
SC1302	PI597962	Caudatum	0	0	Ethiopia	Eastern-Africa	-82.916	67.909	-24.492	-6.937	5.525
SC1305	PI595733	Caudatum	0	0	Ethiopia	Eastern-Africa	-66.684	60.403	-8.334	-3.545	-11.553
SC1307	PI595734	Caudatum	0	1	Ethiopia	Eastern-Africa	-64.088	65.436	-18.371	-1.782	-16.480
SC1313	PI595735	Caudatum	0	0	Ethiopia	Eastern-Africa	-69.537	57.734	-22.970	-4.500	9.177
SC1314	PI595736	Caudatum	0	0	Ethiopia	Eastern-Africa	-69.468	53.228	-23.920	-3.694	13.997
SC1316	PI595737	Caudatum	0	0	Ethiopia	Eastern-Africa	-80.780	65.246	-22.495	-6.258	-2.433
SC1317	PI595738	Caudatum	0	0	Ethiopia	Eastern-Africa	-76.472	61.986	-23.955	-4.292	3.951
SC1318	PI597963	Caudatum	0	0	Ethiopia	Eastern-Africa	-63.166	65.084	-29.341	1.163	-6.218
SC1319	PI597964	Caudatum	0	0	Ethiopia	Eastern-Africa	-75.346	66.018	-24.512	-2.933	1.603

Table S4.1. continued

SC1320	PI597967	Caudatum	0	0	Ethiopia	Eastern-Africa	-72.228	56.787	-18.903	-6.525	-0.059
SC1321	PI597968	Caudatum	0	0	Sudan	Northern-Africa	-18.477	29.654	-3.146	-2.413	-8.809
SC1322	PI597969	Durra-Bicolor	0	0	Sudan	Northern-Africa	21.211	5.976	-9.403	9.751	-11.043
SC1325	PI597970	Caudatum	0	0	Sudan	Northern-Africa	-3.182	17.346	-25.085	5.753	0.193
SC1328	PI597971	Caudatum	0	0	Sudan	Northern-Africa	-35.327	30.740	-5.128	-1.892	-6.295
SC1329	PI597972	Durra-Caudatum	0	0	Sudan	Northern-Africa	17.589	9.376	-8.318	12.392	-25.036
SC1330	PI597973	Durra-Caudatum	0	0	Sudan	Northern-Africa	13.897	0.365	-25.287	1.271	7.964
SC1332	PI597974	Guinea	0	0	Mali	Western-Africa	-21.300	-66.657	-21.289	-11.693	13.947
SC1333	PI597975	Guinea	0	0	Mali	Western-Africa	-38.043	-112.422	-42.188	-11.566	0.164
SC1337	PI597976	Guinea	0	0	Mali	Western-Africa	-24.988	-78.518	-27.167	-9.218	11.170
SC1338	PI595748	Guinea	0	0	Mali	Western-Africa	-24.465	-73.982	-19.792	-12.684	5.855
SC1339	PI597977	Guinea	0	0	Mali	Western-Africa	-28.427	-74.316	-21.317	-8.608	5.138
SC1341	PI597978	Guinea	0	0	Mali	Western-Africa	-26.797	-77.978	-23.370	-8.884	9.635
SC1342	PI597979	Guinea	0	1	Mali	Western-Africa	-24.795	-76.807	-21.321	-10.348	11.742
SC1345	PI597980	Guinea	0	0	Mali	Western-Africa	-70.616	55.570	-26.191	-5.842	9.764
SC1351	PI597981	Caudatum	0	0	Sudan	Northern-Africa	-40.129	24.925	-14.487	-0.537	10.497
SC1356	PI597982	Caudatum	0	0	Sudan	Northern-Africa	-42.108	23.853	-7.484	-2.156	14.504
SC1416	PI656077	Unknown/Other	0	0	Niger	Western-Africa	-10.807	0.319	10.854	-7.057	1.279
SC1424	PI656078	Caudatum	0	1	Mali	Western-Africa	-12.687	-39.351	18.614	12.330	24.913
SC1426	PI656079	Bicolor	0	1	Mali	Western-Africa	-29.806	-76.852	-25.685	-9.783	8.153
SC1429	PI656080	Bicolor	0	1	Unknown	Unknown	-20.897	-32.749	79.889	-4.851	62.705
SC1439	PI656081	Guinea	0	1	Gambia	Western-Africa	0.804	-48.222	6.994	32.005	8.598
SC1440	PI656115	Guinea	1	0	Gambia	Western-Africa	-9.052	-50.092	8.061	20.793	19.427
SC1441	NA	Caudatum	0	NA	Zambia	Eastern-Africa	-19.888	-27.791	32.200	-2.373	33.518
SC1442	NA	Guinea	0	0	Malawi	Eastern-Africa	-17.198	-42.834	67.219	-1.514	60.884
SC1446	NA	Kafir	0	0	Zambia	Eastern-Africa	-21.796	-31.763	34.485	-2.976	39.181
SC1451	PI656083	Caudatum	0	0	Malawi	Eastern-Africa	-51.754	32.055	-11.736	-4.647	2.185
SC1463	PI656084	Caudatum	NA	0	No.Kordofan,Sudan	Northern-Africa	-56.206	45.265	-18.136	-3.892	3.367
SC1465	PI656085	Unknown/Other	0	0	No.Kordofan,Sudan	Northern-Africa	-60.132	55.746	-13.431	-0.968	-8.744
SC1471	PI656086	Unknown/Other	0	0	No.Kordofan,Sudan	Northern-Africa	8.150	22.359	-15.944	9.892	-17.575
SC1476	PI656087	Unknown/Other	0	0	No.Kordofan,Sudan	Northern-Africa	-34.985	16.492	43.136	-13.621	-5.892
SC1484	PI656088	Unknown/Other	0	0	Somalia	Eastern-Africa	81.057	18.545	-23.757	-8.714	-23.034

Table S4.1. continued

SC1489	PI656089	Durra	0	0	Somalia	Eastern-Africa	-61.336	64.649	-3.530	1.044	-22.602
SC1494	PI656090	Unknown/Other	0	0	Sudan	Northern-Africa	-30.330	13.140	-10.590	8.113	-0.912
SC1506	PI656116	Unknown/Other	0	0	Mali	Western-Africa	-25.931	1.861	41.959	-7.212	8.062
SRN39	PI656027	Breeding-line	0	1	Breeding-line	Breeding-line	-34.563	47.354	-24.245	-2.744	-8.952
SURENO	PI561472	Breeding-line	0	1	Breeding-line	Breeding-line	-49.604	43.423	29.547	-4.758	-18.222
SanChiSan	PI542718	Breeding-line	0	0	Breeding-line	Breeding-line	63.580	16.970	-11.578	-7.517	-6.745
Segaolane	PI656023	Breeding-line	0	0	Breeding-line	Breeding-line	-4.789	-10.536	122.752	-9.887	-14.194
ShanQuiRed	PI656025	Breeding-line	0	0	Breeding-line	Breeding-line	56.410	15.637	-6.930	-5.594	-3.628
Soberano	PI656026	Breeding-line	0	0	Breeding-line	Breeding-line	-72.110	67.260	-15.753	-1.933	-15.642
SpurFeter	PI655973	Breeding-line	0	0	Breeding-line	Breeding-line	-24.954	38.136	-19.178	3.894	-21.098
Tx2741	PI655979	Breeding-line	0	0	Breeding-line	Breeding-line	-34.445	11.072	6.723	-4.068	-38.855
Tx2911	PI607931	Breeding-line	0	0	Breeding-line	Breeding-line	-29.609	24.618	31.761	-0.806	-23.410

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions and Future Work

The modified alkali assay was used to screen for the Alkali Spreading Value (ASV) in a sorghum EMS population (Griebel et al. 2019). The assay identified nine ems mutants with an ASV+ phenotype having functional mutations in the genes, *Sobic.004G163700* and *Sobic.010G093400* associated with starch biosynthesis (Griebel et al. 2019; Griebel et al. 2019b). A linkage analysis reported that the ASV+ co-segregates with the new alleles (Griebel et al. 2019b). An allele dosage test identified that the ASV+ is inherited recessively but in an allele dependent manner. The genes could be identified as likely a *SBEIIb*, possible sorghum homolog of maize *amylose extender*, and a *SSIIa*. The seven new alleles of *SBEIIb* (*Sobic.004G163700*) and two new alleles of *SSIIa (Sobic.010G093400*) exhibited distinct classes in their functionality (Griebel et al. 2019b). The *sbeIIb*-mutants behaved similar to amylose extender mutants in maize and resulted in higher starch gelatinization temperature (GT), higher amylose content and lower paste viscosity profiles in comparison to the wild type and *ssIIa*-mutants. The *ssIIa*-mutants showed lower starch GT, lower paste viscosity profiles and similar amylose content as the wildtype but higher paste viscosity than the sbeIIb-mutants (Griebel et al. 2019b).

The adjustment and usage of the low-cost ASV assay helps to identify sorghum genotypes that contribute to improved starch quality and structures influencing food product quality (Griebel et al. 2019; Griebel et al. 2019b). The newly identified alleles are of a great value and should be continuously introgressed, not only in the SMIL 2 project phase, into local varieties in Europe, USA and developing countries such as Niger, Senegal, Ethiopia, Uganda, Zambia.

Improved local varieties carrying the *ssIIa-3920-1 or ssIIa-4308-1 allele*, could have lower starch GT and would benefit for example the local beer brewing industries (Griebel et al. 2019, Griebel et al. 2019b). The starch GT impacts malting and beer quality (House et al. 2000) because the high starch GT in sorghum makes it difficult to gelatinize starch and undergo saccharification into fermentable sugars (Taylor and Dewar 2000; Taylor 1992).

Additionally, starch GT influences paste viscosity and functionality of starchy products (Biliaderis 2009). Improved varieties carrying a *ssIIa*-mutant allele could have a lower starch GT, higher peak and final viscosity, and good gel consistency (Griebel et al. 2019b). While improved varieties carrying one of the *sbeIIb*-mutant alleles could have higher starch GT, higher amylose values and lower paste viscosity profiles (Griebel et al. 2019b). Both bear great potential for food scientists. Thus, the improved sorghum varieties could be used for the food processing industry for the development of (1) gluten free, maybe slow digested starch products for humans with diabetes and gluten allergy and (2) food products that require different thickening strength and viscosity parameters. The improved varieties woud provide potential for local smallholder farmers to have a new product for the market

The results of this dissertation provide a foundation to built upon in future work. The mutant's variation in starch GT, amylose content and paste viscosity profiles are suggested to be a result of a lack or reduced enzyme activity of SSIIa (*Sobic.010G093400* product) and SBEIIb (*Sobic.004G163700* product) (Griebel et al. 2019b). Therefore, it would be valuable to further evaluate four things (1) the enzymes activity, (2) protein complex composition (3) branch chain length distribution and frequency and (4) occurance of resistant and slow digested starches.

Why is it important to evaluate the (1) enzymes activity and (3) branch chain length distribution? The mutants' variation in starch GT, amylose content, and other paste viscosity

characteristics are assumed to be the result of the amylose:amylopectin ratio and the branch chain length distribution including the length of single glucan chains produced by starch biosynthesis enzymes (Jane 2009; Tetlow and Emes 2014; Li et al. 2008; Liu et al. 2012a; Nakamura 2015b; Nishi et al. 2001). The *sbeIIb*-mutants could have longer internal branch chains of amylopectin (Jane 2009, Tetlow and Emes 2014; Nishi et al. 2001), are less branched in outer chains (Jane 2009; Tetlow and Emes 2014) and have less short chains (Jane 2009; Tetlow and Emes 2014; Nishi et al. 2001). The SBEIIb is known to transfer short chains with degree of polimeryzation (DP) approximately 6-7 in other cereals (Tetlow and Emes 2014; Nakamura 2015). Thus, the newly identified *sbeIIb* mutants might have amylopectins with shorter glucan chains smaller than DP8, a less branched molecule and maybe enrichment of chains of DP 12-16. Starches with more chains of DP 12-16 are associated with an increase of starch GT (Nakamura 2015b). The enzyme SSIIa elongates short chains into intermediate size chains (DP 12-24) (Tetlow et al. 2004; Nakamura 2015b), which might be inactivated or reduced in activity in the *ssIIa*-mutants, thus likely resulting in an increase in short chains and less intermediate chains as reported for other species (Nakamura 2015). The increase of shorter chains (smaller than DP12) and less intermediate chains is related to lower starch GT (Nakamura, 2015b). Therefore, it would be of great importance to further evaluate the products of the new identified alleles for their enzyme activity and branch chain length distribution and branching frequency of the resulting starches.

Why is is important to further evaluate the (2) protein complex? In other species the different classes of enzymes involved in starch biosynthesis built a protein complex to create the starch granule (Tetlow et al. 2004; Preiss 2009, Ahmed et al. 2015, Liu et al. 2012b). In barly the enzyme protein complex composition differs for the creation of different types of starch granules (Ahmed et al. 2015). The enzymes in the protein complex influence each other if one is mutated

(Tetlow et al. 2004, Ahmed et al. 2015). It was observed, that phosporylation changed the protein charges of starch branching enzyme (SBE) isoforms and with that the protein-protein interactions, like how SBEs co-work with SS isoforms in a protein complex (Tetlow and Emes, 2014). It has been reported, that the SSI, SSIIa, SBEIIa, and SBEIIb form a protein complex (Ahmed et al. 2015, Tetlow and Emes 2014) and are especially there for the amylopectin biosynthesis in other species (Tetlow and Emes 2014). When the SBEIIb was reduced in activity, the protein complex composition changed affecting starch structure in barley (Ahmed et al. 2015). A lack in SSIIa resulted in an entire loss of the other enzymes acting on the starch molecule in barley (Ahmed et al. 2015) since the SSIIa has the ability to bind the protein complex to the starch granule (Tetlow and Emes 2014). The sorghum ems mutants have distinct functionalities and unique alleles, which makes it a great pool to understand the interplay of the enzymes in sorghum starch biosynthesis and if the protein complex is different affected in *ssIIa* mutants than in *sbeIIb* mutants.

Why is it important to analyse for (4) resistant starch and slow digested starches? In other species amylose extender starches are related to high amylose content and correlated with resistant starch (Jane 2009; Tetlow and Emes 2014; Li et al. 2008). Those starches were less enzyme digestible (Tetlow and Emes 2014; Li et al. 2008). The mutants identified in this study showed similar behaviour such as amylose extender starches (Griebel et al. 2019b). However, not the amylose levels alone, but the branch chain length and branching frequency influence starch digestibility (Tetlow and Emes 2014). Therefore, it is of great importance to determine the branch chain length distribution, and the content of resistant starch or slow digested starches. It would be valuable to evaluate whether the *sbeIIb* mutants produce more resistant or slowly digested starches. The resistant starches belong to the less available carbohydrates as they mostly resist digestion in the human small intestine (Englyst et al. 2007). The resistant carbohydrates are hardly absorbed

into the human metabolism and are considered as non-glycaemic and undigestible (Englyst et al. 2007). Food products with a low glycaemix index usually have slow digested starches and are of important impact on human health (Englyst et al., 2007). The *sbeIIb* mutants identified in our study might have diverse effects on humans with diabetes as reported in a study feeding humans with high amylose maize cookies (Gower et al. 2016).

The ASV phenotype could be shown to be expressed in the standing variation of sorghum (Griebel et al. 2019c). The ASV genotypes that were stably expressed across seasons were identified as *durra* (Nandyal) types (Griebel et al. 2019c). Nandyal types were reported to be glutinous varieties (Dahlberg 2000). Four SC lines were further evaluated and observed with normal amylose content, high viscosity, good gel consistency and environmental dependent lower starch GT (Griebel et al. 2019c). The genotypes are valuable for food products that require more stickiness and viscosity. GWAS showed significant associations for one SNP in proximity to a starch biosynthesis gene *Sobic.010G273800* and two candidate genes *Sobic.010G274800* and *Sobic.010G275001* described as glucosyltransferases. These three genes should be further evaluated for their impact on ASV and starch quality. The four SC lines were also subject to whole genome resequencing (WGS) and showed that the genotypes have many SNPs in genes in starch biosynthesis in comparison to Tx623. It needs to be further evaluated if the SNPs in common are due to the close relatedness of the four genotypes in comparison to Tx623 or causal for ASV.

For gene validation from GWAS the already developed bi-parental F_3 populations (SC489/Tx623; SC491/Tx623; SC587/Tx623; SC589/Tx623) could be used to validate the linkage of these three genes with regard to ASV and starch quality phenotypes. The F_3 populations are ready to be evaluated or, ideally, further advanced to recombinant inbred line (RIL) populations (Griebel et al. 2019c). The sorghum diversity panel could be grown for two more seasons to be

evaluated for ASV for another GWAS to obtain more data points. It would be important to increase the repeatability by screening more than 20 seeds per genotype and plot to account for penetrance and truly catch the presence and absence of the ASV phenotype (Griebel et al. 2019c).

5.2 References

Ahmed Z, Tetlow IJ, Ahmed R, Morell MK, Emes MJ (2015) Protein-protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heterometric enzyme complexes in the synthesis of A and B granules. *Plant Science* 233:95-106.

Biliaderis CG (2009) Structural transitions and related physical properties of starch. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam) pp 293 - 372

Englyst KN, Liu S, Englyst HN (2007) Nutritional characterization and measurement of dietary carbohydrates. *Eur J Clin Nutr*, 61:S19-S39.

Dahlberg JA (2000) Classification and characterization of sorghum. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 99-130.

Gower BA, Bergman R, Stefanovski D, Darnell B, Ovalle F, Fisher G, Sweatt SK, Resuehr HS, Pelkman C (2016) Baseline insulin sensitivity affects response to high-amylose maize resistant starch in women: a randomized, controlled trial. Nutr Metab (Lond). 13:2 DOI 10.1186/s12986-016-0062-5.

Griebel S, Webb MM, Campanella OH, Craig BA, Weil CF, Tuinstra MR (2019) The alkali spreading phenotype in *Sorghum bicolor* and its relationship to starch gelatinization. *J Cereal Sci* 86:41-47.

Griebel S, Westerman RP, Adeyanju A, Addo-Quaye C, Craig BA, Weil CF, Cunningham M S, Bvabesh P, Campanella OH, Tuinstra MR (2019b) Mutations in sorghum *SBEIIb* and *SSIIa* affect alkali spreading value, starch composition, thermal properties and flour viscosity. *In Review. Theor Appl Genet.*

Griebel S, Adeyanju A, Tuinstra MR (2019c) Genetic analysis of the alkali spreading phenotype in a sorghum diversity panel. *Unpublished manuscript.*

House LR, Gomez M, Sun Y, Murty DS, Verma BN (2000) Development of some agricultural industries in several African and Asian countries. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 131-190.

Jane J-l (2009). Structural features of starch granules II. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Elsevier, Academic Press, Netherlands), pp 193-236.

Li L, Jiang H, Campbell M, Blanco M, Jane J-l (2008) Characterization of maize amylose-extender (*ae*) mutant starches. Part I: Relationship between resistant starch contents and molecular structures. *Carbohydr Polym* 74:396-404.

Liu F, Ahmed Z, Lee EA, Donner E, Lui Q, Ahmed R, Morell MK, Emes MJ, Tetlow IJ (2012a) Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein-protein interactions. *J Exp Bot* 63:1167-1183.

Liu F, Romanova N, Lee EA, Ahmed R, Evans M, Gilbert EP, Morell MK, Emes MJ, Tetlow I. J (2012b) Glucan affinity of starch synthase IIa determines binding of starch synthase I and starch-branching enzyme IIb to starch granules. *Biochem J* 448:373-387.

Nakamura Y (2015). Biosynthesis of reserve starch. *Starch metabolism and structure*, eds Nakamura Y, (Springer, Japan), pp 161-209.

Nakamura Y (2015b) Manipulation of rice starch properties for application. *Starch metabolism and structure*, eds Nakamura Y, (Springer, Japan), pp 335-369.

Nishi A, Nakamura Y, Tanaka N, Satoh H (2001) Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiol* 127:459-472.

Preiss J (2009) Biochemistry and molecular biology of starch biosynthesis. *Starch Chemistry and Technology*, eds BeMiller J, Whistler R (Academic Press, Amsterdam), pp 83 – 148.

Taylor J, Dewar J (2000) Fermented products: Beverages and porridges. *Sorghum Origin, History, Technology and Production*, eds Smith CW, Frederiksen RA, (John Wiley & Sons, Inc., New York), pp. 751-796.

Taylor JRN (1992) Mashing with malted grain sorghum. *J. Am. Soc. Brew. Chem*. 50:13-18.

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. *J Exp Bot* 55:2131-2145.

Tetlow IJ, Emes MJ (2014) A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life* 66:546-558.