GENOMIC STUDY AND GENETIC IMPROVEMENT OF SORGHUM [Sorghum bicolor (L) MOENCH] FOR HIGH PROTEIN DIGESTIBILITY

By

ELISABETH DIATTA
(10512871)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY DEGREE IN PLANT BREEDING

WEST AFRICA CENTRE FOR CROP IMPROVEMENT
COLLEGE OF BASIC AND APPLIED SCIENCES
UNIVERSITY OF GHANA
LEGON

December 2018
DECLARATION

I hereby declare that except for references to works of other researchers, which have been duly cited, this work is my original research and that neither part nor whole has been presented elsewhere for the award of a degree.

..................................................
Elisabeth DIATTA
Student

..................................................
Prof. Pangirayi B. TONGOONA
Supervisor

..................................................
Prof. Eric Y. DANQUAH
Supervisor

..................................................
Dr Agyemang DANQUAH
Supervisor

..................................................
Dr Ndiaga CISSE
Supervisor
ABSTRACT

Sorghum is a staple crop and a major source of food and energy for the World’s developing countries. However, its utilization as human food is constrained by the low availability of its proteins after wet cooking which could lead to malnutrition, especially in sub-Saharan Africa. West Africa has a great diversity of sorghum yet, its study and use for improved protein digestibility remains unexplored. To provide knowledge on the characterization of West African sorghum germplasm for its exploitation in breeding programmes, a collection of 385 West African sorghum lines was assessed for seed quality, protein digestibility and anti-nutritional factors. Only 21% of the accessions were tannin-free, 56% had coloured seeds and protein content ranged from 9 to 18 g/100 g of samples. The largest variability for protein digestibility was found in accessions from Niger (1-55%). However, none of the lines of the entire collection was found to have highly digestible proteins after wet cooking i.e. at least 60% digestibility hence the need to explore other sources for variability. Chemical mutagenesis, applied to seeds of BTx623, generated three ethyl methanesulfonate (EMS) mutants with 26 to 37% higher digestible proteins as well as 50 to 100% increase in lysine and 30 to 50% more tryptophan content than their wild type parent. Two of those mutants had invaginated protein body (PB) phenotype while the third mutant displayed a phenotype close to the wild type round PB. Further characterization of the two mutants displaying contrasting PB phenotypes revealed that SbEMS3324 had more digestible proteins than P721Q while SbEMS1613 had harder seed kernels than P721Q, the first highly digestible mutant generated in 1975. The increase in digestible proteins in the mutant with invaginated PBs was controlled by a single recessive allele and linked to a point mutation on a kafirin gene (Sobic.005G189000) in SbEMS3324. However, in the mutant with round PBs, SbEMS1613, a missense mutation on a 26S proteasome PMSD10 subunit complex (Sobic.005G083340) located
on chromosome 5 was causing the change in phenotype. Two pairs of SNP markers (ProF and ProR, KafF and KafR) were developed in this study and successfully shown to be linked to the genes causing the phenotype in each mutant. These primers could play a key role in accelerating the introgression of the high protein digestibility gene into locally-adapted sorghum varieties. Furthermore, to contribute to lower the humongous cost of wheat importation in Senegal as well as the use of maize in formulation of infant food and for poultry feed, the Senegalese research institute developed white grained high yielding sorghum varieties well adapted to different environments in Senegal, nevertheless, these varieties lack the essential amino acid lysine and have poorly digested proteins. To increase their nutritional value, these varieties were crossed to a mutant with high digestible proteins, P721Q. One of the crosses, exploited in this study led to the development of 128 BC3F3 progenies generated from a cross between lowly digestible Faourou and P721Q that were evaluated at Bambey in Senegal during the rainy season in 2017 along with the two parents. Amongst the 128 progenies, only 18 had highly digestible proteins after wet cooking. Eight progenies outperformed the donor parent in terms of digestibility which could be explained by the combination of favourable alleles in those progenies known as transgressive segregation. The identified highly digestible progenies, once released, will be potential varieties which would impact the food and feed industry in Senegal. It is recommended that they are further advanced and tested for yield performance across environments for possible release in Senegal.
DEDICATION

I dedicate this thesis to my loving parents, Idrissa DIATTA and Angelique B. MANSALY, and my 6 siblings: Sylvain, Martine, Lydie, Fiacre, Maurice, and Bertrand. Thank you for your love, support, prayers and encouragements throughout my studies. May God bless you with a long prosperous and healthy life.
ACKNOWLEDGEMENTS

There are several organizations and people who have been instrumental in my PhD journey, that made possible for me to even start the journey as well as complete it. This work was completed across three countries on two continents, and I am tremendously grateful to everyone who played a role in helping me from start to finish and along the way.

First and foremost, I am grateful to God almighty, my heavenly father for the gift of life, health, wisdom and knowledge without which this work would not have been possible. To Him be the Glory now and Forever.

My heartfelt gratitude to USAID Feed the Future Sorghum and Millet Innovation Lab (SMIL) for providing the scholarship and research funds that made it possible to travel and complete my experiments.

I thank WACCI for the excellent training that prepared me with the knowledge and skills necessary to plan and execute experiments. Special regards to Prof Eric Danquah, founding director of WACCI and 2018 World Agriculture Prize winner, you will always be a reference for me as a young scientist. Thanks to WACCI staff members and lecturers for technical support. Genuine gratitude to my supervisory committee members: Prof Pagirayi B. Tongoona, Prof Eric Danquah, Dr Agyemang Danquah, and Dr Ndiaga Cisse. Thank you all for your critical reviews, suggestions and corrections brought to this document.

Sincere thanks to my in-country supervisor Dr Ndiaga Cisse, director of CERAAS research institute in Senegal, for guidance, advice, training opportunities and providing a host institute and part of the material used in this study. Thank you for pushing me forward and trusting in my abilities to perform this study. You have been a great mentor.

My gratitude to Dr Khalil Kane for corrections to the thesis proposal and for supervising part of my work at CERAAS and for advice given; to Dr Cyril Diatta for providing technical advice and help with field experiments and data analysis. The two of you have been great “big brothers”. My appreciation goes to Dr Daniel Fonseca, Khady Ndour, Ndeye Coura Fall, Anta Y. Sy, Yvette Rachelle Djiboune, Mbaye Ndoye Sall, and Elisa Diop for providing moral, lab and field technical support. Thanks to CERAAS administration for facilitating paperwork during my stay.
Heartfelt thanks to my mentors Professors Clifford Weil and Mitch Tuinstra for hosting me in your labs at Purdue University, USA where I had the opportunity to conduct excellent research and training under your supervision. I am very grateful for your guidance, advice and believing in my abilities to conduct this research. I will be forever grateful for the opportunity.

During my time at Purdue University, I had the privilege to meet awesome people of God: my Chi Alpha family especially Shalyse Iseminger and Horane A. Holgate. I cannot thank you enough for your love, all the prayers, encouragement and moral support you provided that kept me going. Thanks to each one of my Chi Alpha brothers and sisters for the fun times and the pleasant company that gave me a healthy balance between PhD research and social life.

Very special thanks to:

- The technicians at Bambey: Ousmane Aidara, Mamadou Bounama Sall, Pape Ndiaye, Abdoulaye Ndao, Alamine Badiane, and Alassane Diouf for your tremendous help during field data collection.
- My fellows, WACCI Cohort 8 members, for your encouragement and company in Ghana.
- Jacques M. Faye and Prof. Geoffrey Morris from Kansas State University for providing WASAP GBS data.
- My former lab mates at Purdue University: Jacqueline Anderson, Dr Addie Thompson, Eli Hugges, Patrick Sweet, Dr Roselyn Hatch, David Schlueter, Andy Linvill, Eugene Glover, Matt, Joe, Ryan and Holly for collaboration, assistance in the lab and during field work.
- Moloko G. Mathipa for the coffee breaks at Lavazza, your friendship and lovely company while at Purdue.
- Dr Hudson and Dr Kaity Rainin Martin for providing QPCR and NIRS facilities.
- To my beloved Horane A. Holgate for your love, prayers, moral support, your help in covering and uncovering the plants in Lilly-LSPS greenhouse. Thanks for reading and bringing corrections to this document.
TABLE OF CONTENTS

DECLARATION ................................................................................................................................................. i

ABSTRACT ...................................................................................................................................................... ii

DEDICATION .................................................................................................................................................... iv

ACKNOWLEDGEMENTS ....................................................................................................................................... v

TABLE OF CONTENTS ....................................................................................................................................... vii

LIST OF TABLES ............................................................................................................................................ xiv

LIST OF FIGURES ............................................................................................................................................ xvi

LIST OF ABBREVIATIONS ............................................................................................................................... xviii

CHAPTER ONE ................................................................................................................................................... 1

1.0 GENERAL INTRODUCTION .................................................................................................................... 1

CHAPTER TWO ................................................................................................................................................... 5

2.0 LITERATURE REVIEW ............................................................................................................................... 5

2.1 Sorghum ....................................................................................................................................................... 5

2.1.1 Botany ..................................................................................................................................................... 5

2.1.2 Origin, domestication and spread ......................................................................................................... 5

2.1.3 Economic importance, production, and utilization .............................................................................. 7

2.2 Composition and nutritional value of sorghum grain .............................................................................. 8

2.3 Factors reducing protein digestibility in sorghum grain ........................................................................ 9

2.3.1 Endosperm structure, protein interactions, and protein bodies ......................................................... 9
3.2.2.4 Determination of Starch, protein, total digestible nutrients and acid detergent fiber using near-infrared reflectance spectroscopy ................................................................. 24

3.2.3 Genotyping by sequencing ........................................................................................................... 25

3.2.4 Data analysis ........................................................................................................................................... 26

3.2.4.1 Protein digestibility........................................................................................................................... 26

3.2.4.2 Genetic diversity of WASAP ........................................................................................................... 26

3.3 Results ......................................................................................................................................................... 28

3.3.1 Distribution of seed colour of the collection ...................................................................................... 28

3.3.2 Variation for protein content ................................................................................................................. 28

3.3.3 Variation for protein digestibility ........................................................................................................ 30

3.3.4 Tannin in seeds of WASAP .................................................................................................................. 32

3.3.5 Distribution of acid detergent fiber, total digestible nutrients, and starch in WASAP .................. 33

3.3.6 Correlations between traits ................................................................................................................ 36

3.3.7 Genetic diversity in the West African sorghum collection................................................................. 37

3.4 Discussion .................................................................................................................................................. 39

3.5 Conclusion.................................................................................................................................................. 43

CHAPTER FOUR ........................................................................................................................................ 44

4.0 Characterization of highly digestible sorghum EMS mutants using Bulked Segregant Analysis ........................................................................................................................................ 44

4.1 Introduction ............................................................................................................................................. 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 Material and methods</td>
<td>47</td>
</tr>
<tr>
<td>4.2.1 Plant materials</td>
<td>47</td>
</tr>
<tr>
<td>4.2.2 Phenotypic characterization of the EMS mutants</td>
<td>47</td>
</tr>
<tr>
<td>4.2.2.1 Determination of percent protein digestibility</td>
<td>47</td>
</tr>
<tr>
<td>4.2.2.2 Protein body morphology of sorghum mutants</td>
<td>48</td>
</tr>
<tr>
<td>4.2.2.3 Determination of protein, total lysine and tryptophan contents</td>
<td>48</td>
</tr>
<tr>
<td>4.2.2.4 Measurement of seed diameter</td>
<td>48</td>
</tr>
<tr>
<td>4.2.1.5 Determination of seed hardness</td>
<td>49</td>
</tr>
<tr>
<td>4.2.3 Identification of allele(s) and gene(s) controlling high protein digestibility through bulked segregant analysis (BSA)</td>
<td>49</td>
</tr>
<tr>
<td>4.2.3.1 Mapping population development</td>
<td>49</td>
</tr>
<tr>
<td>4.2.3.2 Development of F3 recombinant populations</td>
<td>50</td>
</tr>
<tr>
<td>4.2.3.3 Measurement of protein digestibility for the mapping populations</td>
<td>50</td>
</tr>
<tr>
<td>4.2.4 Mapping of protein digestibility genes in EMS mutants</td>
<td>51</td>
</tr>
<tr>
<td>4.2.4.1 DNA extraction and sequencing</td>
<td>51</td>
</tr>
<tr>
<td>4.2.4.2 Bulked segregant analysis for protein digestibility</td>
<td>51</td>
</tr>
<tr>
<td>4.2.5 Analysis of conserved mutations between progeny with one shared parent</td>
<td>52</td>
</tr>
<tr>
<td>4.2.6 Data analysis</td>
<td>54</td>
</tr>
<tr>
<td>4.2.6.1 Phenotypic data</td>
<td>54</td>
</tr>
<tr>
<td>4.2.6.2 Sequencing data</td>
<td>54</td>
</tr>
</tbody>
</table>
4.3 Results ........................................................................................................................................ 56

4.3.1 Protein digestibility of sorghum ............................................................................................. 56

4.3.2 Protein body morphology of the EMS mutants ........................................................................ 57

4.3.3 Crude protein, lysine and tryptophan contents of sorghum mutants ..................................... 58

4.3.4 Seed diameter of EMS mutants .............................................................................................. 59

4.3.5 Seed hardness of EMS mutants .............................................................................................. 60

4.3.6 Whole genome sequencing .................................................................................................... 61

4.3.7 Identification of causative alleles in EMS932 x EMS1613 mapping population .......... 61

4.3.7.1 Variation in protein digestibility in SbEMS1613 mapping population ......................... 61

4.3.7.2 Bulked segregant analysis in SbEMS1613 mapping population ..................................... 63

4.3.7.3 SNP validation .................................................................................................................. 66

4.3.8 Identification of causative alleles in SbEMS3324 x SbEMS932 mapping population 68

4.3.8.1 Distribution of protein digestibility in SbEMS3324 mapping population ................. 68

4.3.8.2 Bulked segregant analysis in SbEMS3324 mapping population .................................. 69

4.3.8.3 SNP validation .................................................................................................................. 72

4.4 Discussion .................................................................................................................................. 73

4.5 Conclusion ................................................................................................................................. 78

CHAPTER FIVE ................................................................................................................................ 79

5.0 Development and Evaluation of a Highly Digestible Sorghum Population in Senegal ....... 79

5.1 Introduction ................................................................................................................................ 79
5.2 Material and methods ........................................................................................................................................... 81
  5.2.1 Plant material .................................................................................................................................................... 81
  5.2.1.1 Description of the parents used ................................................................................................................... 81
  5.2.1.2 Population development ............................................................................................................................... 82
  5.2.1.3 Quality control using foreground markers .................................................................................................. 86
  5.2.1.4 DNA extraction and quantification .............................................................................................................. 87
  5.2.1.5 PCR amplification and sequencing ............................................................................................................... 88
  5.2.2 Description of the experimental site .............................................................................................................. 89
  5.2.3 Field experimental design and management .................................................................................................. 90
  5.2.4 Data collection .................................................................................................................................................. 91
  5.2.5 Data analysis .................................................................................................................................................... 92
5.3 Results .................................................................................................................................................................. 94
  5.3.1 Variation for measured traits ......................................................................................................................... 94
  5.3.2 Performance of BC₃F₃ families ..................................................................................................................... 94
  5.3.3 Performance of high and low protein digestible entries ................................................................................ 97
  5.3.4 Variance components and heritability estimates ............................................................................................ 97
  5.3.5 Phenotypic correlation between protein digestibility and agronomic traits ............................................. 98
  5.3.6 Classification of BC₃F₃ families ..................................................................................................................... 99
5.4 Discussion ............................................................................................................................................................ 102
5.5 Conclusion .......................................................................................................................................................... 105
CHAPTER SIX

6.0 GENERAL CONCLUSIONS AND RECOMMENDATIONS

6.1 General conclusions

6.2 Recommendations

REFERENCES

APPENDICES
LIST OF TABLES

Table 3.1 Distribution of seed colour in WASAP .......................................................... 28
Table 3.2 Top 20 accessions of the WASAP for protein content .................................... 29
Table 3.3: Mean square, variance components and broad sense heritability of protein digestibility .................................................................................................................. 30
Table 3.4 Top 20 accessions of the WASAP for protein digestibility .............................. 31
Table 3.5 Presence of tannins in WASAP seeds ............................................................... 32
Table 4.1 Primer information per mutant ................................................................. 53
Table 4.2: PCR reaction volumes per samples according to high fidelity master mix manufacturer’s conditions (items put in order) ................................................................. 53
Table 4.3: PCR cycling steps and conditions according to the master mix manufacturer’s conditions .......................................................................................................................... 53
Table 4.4: Crude protein, total lysine and tryptophan contents of wild type (BTx623) and sorghum mutants .......................................................................................................................... 59
Table 4.5: Mean square for seed diameter in F3 families ............................................. 60
Table 4.6: Seed diameter of sorghum mutants (SbEMS3324, SbEMS1613, P721Q) and their parent lines (BTx623 and P721N). ................................................................. 60
Table 4.7: Mean squares for protein digestibility in F3 families ..................................... 62
Table 4.8 Protein digestibility means per entries of high digestible bulk for BSA. ............. 64
Table 4.9 Protein digestibility means per entries of low digestible bulk for BSA. ............. 65
Table 4.10: Analysis of Variance for protein digestibility in F3 families of SbEMS3324 x SbEMS932 .......................................................................................................................... 68
Table 4.11: Protein digestibility means per entry of high and low bulks chosen for BSA.... 70
Table 5.1: Comparison between the donor (P721Q) and the recurrent (Faourou) parents........ 81
Table 5.2: PCR reaction components and concentration per sample........................................ 88
Table 5.3: PCR cycling steps and conditions............................................................................. 89
Table 5.4 Precipitation, temperature and relative humidity at Bambey research station during field trials........................................................................................................................................ 90
Table 5.5: Mean square values for protein digestibility and agronomic traits for BC₃F₃ families 94
Table 5.6: Analysis of variance between highly digestible and lowly digestible groups ........ 95
Table 5.7: Mean performance of twenty highest and ten lowest digestible entries................. 96
Table 5.8: Comparison between the performance of high and low protein digestible BC₃F₃ families........................................................................................................................................ 97
Table 5.9: Variance components and heritability estimates at Bambey research station ......... 98
Table 5.10: Correlation coefficients between protein digestibility and agronomic traits........ 99
Table 5.11: Classification of the lines by Agglomerative Hierarchical Clustering (AHC) ......... 100
Table 5.12: Lambda Wilk test for the traits measured.............................................................. 100
LIST OF FIGURES

Figure 3.1 Distribution of protein content of all WASAP lines. ............................................................ 29
Figure 3.2 Distribution of protein digestibility of WASAP lines............................................................... 31
Figure 3.3: Bleach test analysis results................................................................................................. 33
Figure 3.4: Boxplots for Protein digestibility, Protein content, Starch content, Total digestible
nutrients, and Acid detergent fiber for accessions from Mali, Niger, Senegal, and Togo.............. 35
Figure 3.5: Correlation matrix of seven measured traits on WASAP accessions............................... 36
Figure 3.6: Heterozygosity level of lines in the West African Sorghum Association Panel.............. 37
Figure 3.7: Genetic distribution of WASAP accessions........................................................................... 38
Figure 4.1: Comparison between protein digestibility of uncooked and cooked maize (B73, M017,
and W22) and sorghum (BTx623, SbEMS932, and P721N).................................................................. 56
Figure 4.2: Comparison between average (5 replicates) protein digestibility of P721N and its
mutant P721Q; between BTx623 and its EMS mutants SbEMS1613, SbEMS1227, SbEMS3324.
......................................................................................................................................................... 57
Figure 4.3: Transmission electron microscopy images showing the protein body structure of the
non-mutagenized parent BTx623 (A) and the 3 highly digestible mutants SbEMS1613 (B),
SbEMS1227 (C), and SbEMS3324 (D)................................................................................................... 58
Figure 4.4: Seed hardness of mutants P721Q, SbEMS1613 and SbEMS3324 compared to their
parent lines P721N and BTx623 respectively...................................................................................... 61
Figure 4.5: Distribution of the BLUEs of protein digestibility............................................................... 63
Figure 4.6: Bulked segregant analysis in F2 segregants of SbEMS932 x SbEMS1613 mapping
population. ........................................................................................................................................... 66
Figure 4.7: Partial alignment of the 26S proteasome complex regulatory subunit PSMD10...... 67
Figure 4.8: Distribution of the corrected protein digestibility values (BLUEs). ......................... 69

Figure 4.9: Bulked segregant analysis in F3 segregants of SbEMS932 x SbEMS3324 mapping population. ........................................................................................................................................... 71

Figure 4.10: Partial alignment of the Kafirin PSKR2 precursor-like protein amplified from the genomic DNA of 15 low and 35 high protein digestible (HD) progenies................................. 72

Figure 5. 1: Comparison of leaf arrangement between Faourou (left) and P721Q (right). ........ 82

Figure 5. 2: Tan plant (A) Faourou, Presence of anthocyanin on (B) P721Q, and true F1 (C) generated from Faourou x P721Q.............................................................................................................................................. 84

Figure 5. 3: Breeding scheme showing the development of the study population. ............... 86

Figure 5.4: Discriminant Analysis of the study population showing entries grouped in 3 main clusters. .......................................................................................................................................................... 101
LIST OF ABBREVIATIONS

ACRE: Agronomy Centre for Research and Education

ADF: Acid Detergent Fiber

AFLP: Amplified Fragment Length Polymorphism

ANOVA: Analysis of Variance

AOAC: Association of Official Agricultural Chemists

BC: Before Christ

BLAST: Basic Local Alignment Search Tool

BLUE: Best Linear Unbiased Estimator

BLUP: Best Linear Unbiased Predictor

bp: base pairs

BSA: Bulked Segregant Analysis

CERAAS: Centre d'Etude Régional pour l’Amélioration de l’Adaptation à la Sècheresse

CIAA: Chloroform Iso-Amyl Alcohol

CTAB: Cetyl Trimethylammonium Bromide

CV: Coefficient of Variation

DNA: Deoxyribonucleotide

DSFLO: Days to 50% flowering

EMMA: Evolutionary Model-based Multi response Approach

EMS: Ethyl Methane Sulfonate

FAO: Food and Agriculture Organisation

GAPIT: Genome Association and Prediction Integrated Tool

GBS: Genotyping by Sequencing
GLM: General Linear Model
GWAS: Genome-Wide Association Studies
HD: High digestible
HPL: Plant Height
IDT: Integrated DNA Technology
ISRA: Institut Sénégalais de Recherche Agricole
ITA: Institut de Technologie Alimentaire
KH₂PO₄: Potassium phosphate
LPAN: Panicle Length
LPED: Peduncle Length
MAF: Minor Allele Frequency
MAS: Marker Assisted Selection
MATAB: Mixed Alkyl Trimethylammonium Bromide
NaOH: Sodium Hydroxide
NGS: Next Generation Sequencing
NIRS: Near Infrared Reflectance Spectrophotometry
PCA: Principal Component Analysis
PCR: Polymerase Chain Reaction
PDigest: Protein Digestibility
QTL: Quantitative Trait Loci
RAPD: Random Amplified Polymorphic DNA
REML: Restricted Maximum Likelihood
RFLP: Restriction Fragment Length Polymorphism
RILs: Recombinant Inbred Lines
rpm: rotation per minute
SMIL: Sorghum and Millet Innovation Lab
SNPs: Single Nucleotide Polymorphisms
SSR: Simple Sequence Repeats
SVS: SNP and Variation Suite
TASSEL: Trait Analysis by association, Evolution, and Linkage
TCA: Trichloroacetic acid
TDN: Total Digestible Nutrients
USA: United States of America
UV: UltraViolet
WAAPP: West Africa Agricultural Productivity Program
WASAP: West African Sorghum Association Panel
WPAN: Panicle Width
WT: Wild type
CHAPTER ONE

1.0 GENERAL INTRODUCTION

Sorghum (Sorghum bicolor (L.) Moench) is an important feed and staple food crop that plays a major role in agricultural development and food security in many countries around the world. Originated from North-Eastern Africa, sorghum is grown today all over the World and is well adapted to many climatic regions especially the warm climate of Sub-Saharan Africa because of its ability to tolerate drought and high temperature which makes it the ideal crop especially under the phenomenon of climate change (Paterson et al., 2009). Sorghum’s potential to contribute to food security can be enhanced through fundamental and applied research especially focusing not only on increased yield but also on nutrient availability.

Globally, sorghum ranks fifth among the most produced cereals after wheat, rice, maize, and barley (FAO, 2010). Despite its great value as a cereal crop, sorghum production is limited by both biotic and abiotic factors that reduce its yield performance (Kenga et al., 2004). In Africa, the average sorghum production from 2006 to 2016 was 25 Million Tons (MT) from an average harvested area of 26 Million hectares (Mha) which is low compared to other cereals like maize that, for 33 Mha, produced 64 MT (FAO, 2018). In Senegal, sorghum is the fourth most important cereal crop after rice, millet, and maize with a mean production of 139,691 Tons over the past five years (2013-2017) for 159,726 ha and an average yield of only 874 Kg/ha (ANSD, 2018).

Sorghum is a worldwide used cereal crop. In the developing countries, sorghum is mainly used for human consumption while in the developed countries like United States of America (USA) it is used as fodder for poultry and cattle, in beer industries, to produce ethanol and sugar (sweet sorghum) (Dicko et al., 2006a). In Africa, Asia, and Latin America, sorghum is produced for its grain that can be used to prepare food and beverages for human consumption such as tortillas,
porridges (e.g. tô, fonde, fura), bread (kisra, injera), granulated foods (e.g. couscous), and local fermented beers from Africa (e.g. dolo, tchapallo) (Dicko et al., 2006a). Sorghum stems can also be used for roofing. In addition, sorghum is referred to as a promising crop because it costs relatively less to grow and requires less water than maize (Gherbin et al., 1996).

Compared to maize, sorghum proteins and starch are less accessible for enzyme degradation in animal, human or industrial processing of the grain (Spicer et al., 1982, 1983; Gualtieri and Rapaccini, 1990; Dowling et al., 2002). In fact, when wet cooked, only about 46% of the total protein found in sorghum is digestible, as compared with 81% in wheat, 73% in maize, and 66% in rice (Maclean Jr et al., 1981). This leads to severe malnutrition in people whose diet is based on sorghum as a primary source of protein (Maclean Jr et al., 1981; Wu et al., 2013). Malnutrition rate in Sub Saharan Africa (SSA) was reported to be the second highest worldwide after south India (Chopra and Darton-Hill, 2006). Malnutrition can lead to stunted growth or disease like kwashiorkor, which results from a protein-poor diet. In addition, in SSA, the rates of child deaths keep increasing and about 48 million children under 5 years old die every year due to malnourishment (Chopra and Darton-Hill, 2006). Furthermore, animals fed with low digestible sorghum lines might show reduced growth rates and lower nutritional value (Lucbert and Castaing, 1986). Therefore, to fight malnutrition and achieve food security, research focus should be placed on proteins and their availability because proteins are the primary expression of genes and play a fundamental role in the metabolism of plants, animals, and humans (Alberts et al., 2002).

There is a lack of information about natural highly digestible sorghum cultivars. Therefore, exploiting natural collections of sorghum accessions will be very beneficial to identify, if there exists, promising lines that can be used in breeding programmes for the improvement of the digestibility of sorghum proteins in areas where sorghum is consumed as a staple crop especially
in West Africa. In that perspective, there was a need for germplasm collection of West African sorghum lines to provide a phenotypic and genotypic characterization of the accessions for their practical use in breeding programmes.

In developing countries, sorghum is the main source of energy for over 300 million people (Godwin and Gray 2000) and represents 60% of the total land area for the production of cereal crops in Africa (FAO, 2018) where it is a staple crop for millions of people. This shows the important role that sorghum could play in achieving food security by promoting its use and improving its nutritional value and market availability. Furthermore, in many parts of Africa, sorghum is the main cereal crop grown and consumed for its grain. Its tolerance to heat and drought and high nitrogen use efficiency make sorghum a key cereal crop that could substitute maize in areas with high temperatures and low water supply (Gardner et al., 1994). However, its low protein digestibility requires the need for supplementation, in diets, by other crops rich in protein but susceptible to drought such as soybean. In addition, its low protein availability makes sorghum a poor-quality staple food unless consumed as part of a varied diet which can be beyond most people’s means especially in the developing world where sorghum is mostly consumed in poor rural areas.

The Senegalese Agricultural Research Institute developed tannin-free high grain quality varieties with a yield between 2000 and 3000 kg ha\(^{-1}\) and are well adapted to many environments in Senegal (ISRA, 2012). However, those varieties have poorly digested proteins after wet cooking which when consumed as staple food could lead to malnourishment. Therefore, to overcome this situation, identifying highly digestible lines or improving the digestibility of locally-adapted sorghum varieties is imperative.
In 1975, a high lysine, high protein digestible sorghum mutant named P721Q (Q for opaque endosperm) was developed at Purdue University by applying chemical mutagenesis on seeds of a lowly digestible sorghum line, P721N (N for normal endosperm) (Mohan, 1975). Unfortunately, this highly digestible mutant had a floury endosperm and small soft seeds which were susceptible to insect damage and not very suitable for some African meals. Despite the low seed quality of P721Q, its existence showed that the high digestibility trait is under genetic control and therefore, it would be possible to develop lines that combined the good nutritional grain quality. To do this, mutagenesis was used again but this time on seeds of BTx623, the sorghum reference genome, (Krothapalli et al., 2013) and led to the development of new highly digestible mutants (unpublished). The genetic mechanisms underlying the increase in digestibility in those EMS mutants remained unknown.

The objectives of this study were to:

1. characterize a West African sorghum association panel for protein digestibility and related nutritional traits,
2. characterize highly digestible EMS mutants to identify the genomic regions controlling high protein digestibility after wet cooking, and
3. improve the digestibility of a Senegalese well-adapted sorghum variety through marker-assisted backcrossing.
CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Sorghum

2.1.1 Botany

Sorghum (Sorghum bicolor (L.) Moench) is an important feed and staple food crop that plays an important role in the development of agriculture and food security in many countries around the world. It is a C4, diploid (2n=2x=20) cereal crop that belongs to the family of Poaceae and the tribe of Andropogonoidea that also comprises maize and sugarcane (Chanterau et al., 2013). Sorghum is predominantly a self-pollinating crop with different proportions of outcrossing and high levels of genetic diversity (Sanchez et al., 2002). Sorghum’s genome has been sequenced from the accession BTx623 and is about 730Mbp, approximately five times the size of Arabidopsis (Paterson et al., 2009). Sorghum’s ability to adapt to different environments is due to its root system and biology which makes it a high water use efficient crop that can be grown in temperate climates as well as the tropics (Gherbin et al., 1996). Sorghum is well adapted to conditions from 300 mm (Mali, Niger, and Senegal) to 1500 mm (Guinea and Sierra Leone) of rain in West Africa (Chanterau et al., 2013).

2.1.2 Origin, domestication and spread

Domesticated in North Eastern Africa more precisely in the North-Eastern part of Africa between Ethiopia and South Sudan about 5,000 years ago, sorghum is well-adapted to heat and drought conditions of the semi-arid tropics (De Wet and Harlan, 1971). The negative impacts of climate change are major causes of food insecurity, especially among the World's poorest people. It is therefore important to consider better practices as well as enhance agricultural production using
crops such as sorghum which is drought tolerant. Initially, a grass growing in the savannas and steppes of Africa, sorghum, through the process of domestication about 3000 years BC (Smith and Frederiksen, 2000), became a staple crop for millions of people in the World, especially in Sub-Saharan Africa. Most of the variability within the sorghum species is reported to be found in north-eastern Africa between Ethiopia, Sudan and Eritrea indicating its area of origin and domestication (Vavilov, 1951; De Wet and Harlan, 1971; Thomas et al., 1996). The diversity can be assessed using molecular, morphological (Motlhaodi et al., 2017), as well as biochemical traits. The large diversity of cultivated and wild sorghum found in Africa creates no doubt as for its origin but its domestication might have taken place in different places through the action of humans who carried it inside the continent and all over the World.

Domestication of the guinea race was reported to have taken place 3,000 years ago in Nubia where the first archeologic remains were found (Chanterau et al., 2013). According to Chanterau et al. (2013), Southern and Western Africa might have been the secondary centres of sorghum domestication for guinea and kafir races while caudatum and durra races rose from the domestication of sorghum in its area of origin. In Asia, the first cultivated forms of sorghum, races bicolor and durra, appeared around 3,000 years BC in China, and 2,000 BC in India; this was followed by the introduction of guinea and guinea-kafir about 3,000 years ago (Chanterau et al., 2013). The introduction of sorghum in Europe took place during the first millennial after Christ (Piedallu, 1923). Sorghum was more recently introduced in America during the triangular triad and was reported in the 18th century (Chanterau et al., 2013). Sorghum’s adaptation to a wide range of environments might have resulted from the spreading of the crop to different parts of the World from the tropical climates of Africa to Asia followed by Europe and America.
2.1.3 Economic importance, production, and utilization

The top five largest producers of sorghum globally between 1994 and 2016 were in order: the United States of America (USA), Nigeria, India, Mexico, and Sudan (FAO, 2018). Since 1988, the USA and India have occupied the first and second places, however, sorghum cultivated land area has reduced in favour of sponsored crops like wheat and rice in India, and genetically modified maize in the USA leading to lower production of sorghum (Chanterau et al., 2013). Americas (39%) are the largest producer of sorghum followed by Africa (37.6%) where it constitutes the second most cultivated cereal crop after maize (FAO, 2018).

Compared to maize, rice, and barley, sorghum has a lower but stable and reliable yield (FAO, 2010). The low productivity of sorghum could be due to biotic stresses and low soil fertility. Other factors like poor management practices, non-availability of improved seeds, and fertilizers to farmers affect sorghum productivity, especially in developing countries. In Senegal, sorghum ranks fourth among the rainfed cultivated cereal crops after rice, pearl millet, and maize in terms of production, which was 139,691 Tons as an average of the last 5 years production (2013 to 2017) (ANSD, 2018). The land area attributed to sorghum production alone could be up to 50% of the total cereal land area in many West African countries where sorghum plays an important role as food and feed.

Sorghum is used in different ways depending on the country and the end-use goal. Approximately 40% of sorghum attributed to grain production is used to prepare food and beverages (example “dolo”) for human consumption (Dicko et al., 2006a). Due to its gluten-free status, sorghum could be used in a combination with 50-80% wheat flour to produce good quality bread and other pasta products (Hugo et al., 2003; Dicko et al., 2006b). Nonetheless, for people allergic to gluten, substituting wheat with sorghum flour will be a better choice. The grain can be popped like maize,
shredded and flaked to produce cereals for breakfast. Some varieties of sorghum have a sweet stem that can either be chewed or used to produce sugar. Sorghum is also used for animal feeding in emerging countries like Mexico, Brazil, or Argentina, but also in developed countries like the USA where it is primarily grown as a forage crop as well as for bioethanol production using sweet sorghum (Chanterau et al., 2013). In addition, sorghum stems can be used for roofing or to make brooms. In Senegal, sorghum is mainly consumed as granulated food (couscous) and/or porridge (“fondé”).

2.2 Composition and nutritional value of sorghum grain

Carbohydrates (starch and dietary fiber as main components), proteins, non-starch polysaccharides, and fat are the main components of sorghum grain (Duodu et al., 2003). The non-starch polysaccharides of sorghum are in the pericarp and endosperm cell walls and depending on the variety, can vary between 2 and 7% in the kernel (Verbruggen et al., 1993). The fat in sorghum grain, mostly made of polyunsaturated fatty acids and located in the germ, is similar to maize fat but is less saturated (Rooney, 1978).

Sorghum grain flour has an approximate energy value of 356 kcal/100g (BSTID-NRC, 1996). However, when cooked, some interactions between starch and protein lead to the formation of resistant complexes that may affect the digestibility of both (Spicer et al., 1982). This could be beneficial for feeding diabetic people or to fight obesity. In addition, food made from such low digestible protein or starch is known to exhibit a low glycaemic index and can give a long feeling of satiety (Awika and Rooney, 2004; Shin et al., 2004).

Sorghum is a good source of vitamins and minerals. Indeed, sorghum is rich in vitamins A, B, D, E, and K; in minerals phosphorous, potassium, iron, and zinc (Hopkins et al., 1998). The latter is
very important for pregnant women. Zinc deficiency is more prevalent in wheat and maize than in sorghum (Hopkins et al., 1998; Dicko et al., 2006a).

2.3 Factors reducing protein digestibility in sorghum grain

Sorghum, compared to other cereal grains like maize, wheat, and rice, has low protein digestibility. In fact, when sorghum is wet cooked, only about 46% of its protein is digestible (Axtell et al., 1981). However, it was reported that a pre-fermentation may increase the digestibility (Axtell et al., 1981; Taylor and Taylor, 2002). Many factors, such as endosperm structure, protein interactions, protein bodies, presence of a testa rich tannin layer (to name a few) are assumed to reduce the protein digestibility of sorghum grains (Duodu et al., 2003).

2.3.1 Endosperm structure, protein interactions, and protein bodies

Sorghum proteins have been classified into albumins, globulins, kafirins, cross-linked kafirins and glutelins based on the solubility-based grouping (Jambunathan et al., 1975). It is hypothesized that the decrease in digestible proteins in wild type sorghum with regular endosperm is largely caused by the kafirins (Oria et al., 2000). Kafirins, alcohol-soluble prolams, are the main storage proteins in the grain (Shull et al., 1991). They represent more than 70% of the total seed protein and are divided into three groups namely α, β, and γ kafirins based on their molecular weight and molecular sequence (Shull et al., 1991). The α-kafirins constitute about 80% of total protein content, the β- and γ-kafirins represent 15 and 5% respectively (Shull et al., 1991). It has been reported that all the 20 α-kafirin genes are located on chromosome 5 of the BTx623 genome (Xu and Messing, 2008). The α-kafirins are highly hydrophilic and easily digested, while the β- and γ-kafirins, rich in amino acid cysteine, are less soluble and not straightforwardly available for enzyme digestion because they can form enzyme-resilient structures in contact with water.
(Hamaker et al., 1987; Oria et al., 2000). Oria et al. (2000) suggested that the reduced digestibility of sorghum proteins is due to strong disulphide bonds between β- and γ-kafirins resulting in enzyme resistant structure on the periphery of storage proteins. The chemical bonds are assumed to block the accessibility of the highly digestible α-kafirins, enclosed inside the endosperm, to the proteolytic enzymes making them insoluble (Hamaker et al., 1987; Oria et al., 2000). In addition, during starch hydrolysis, the disulphide bonds, as well as protein-starch interactions, also limit access to starch granules by amylases (Wong et al., 2009). Therefore, grain end use can be affected by the structure and composition of the seed endosperm.

In the mutants P721Q and P850029, the increased digestibility is hypothesized to be due to structural rearrangement of β- and γ-kafirins in the endosperm and a reduction in the total amount of γ-kafirin in the endosperm (Tesso et al., 2006). The highly digestible mutants also exhibit a floury endosperm and modified protein bodies with invaginations that are assumed to allow proteolytic enzymes to reach the proteins located inside the endosperm increasing the digestibility of both protein and starch (Oria et al., 2000).

2.3.2 Testa, tannin, and grain colour

Sorghum is a source of unique phytochemical constituents that have important human health attributes (Awika and Rooney, 2004). Tannins might be the most widely studied phytochemicals in sorghum. Tannins, in sorghum cultivars, are located in a coloured testa layer whose development was found to be affected by environmental conditions (Kaufman et al., 2013). Tannins are polymeric phenolic compounds involved in plant defence mechanisms against pathogens and predator attacks. However, their content, composition and molecular weight are known to negatively impact the digestibility of sorghum proteins by binding to γ-kafirins (Awika and
Rooney, 2004; Kaufman et al., 2013) thereby reducing growth rate and productivity of livestock fed sorghums with high tannin contents. The effects of the tannins on proteins might be influenced by both genetic and environmental factors.

Depending on the tannin content extracted, and the product used, sorghum can be classified as type I, type II, or type III depending on the method of extraction (Price et al., 1978). In types II and III tannin sorghum varieties, the presence of tannin is associated with a pigmented testa located between the aleurone, and the endocarp layers (Price et al., 1978). In such sorghum types, the tannin is controlled by three major genes B1, B2 and the spreader gene (S) (Hahn and Rooney, 1986). The formation of the pigmented testa layer is controlled by B1 and B2, which must be all dominant while the spreader gene can be either dominant or recessive and when dominant, the seeds are brown in colour (Earp et al., 1983). Wu et al. (2012) characterized the B1 as Tannin1 or Tan1 gene using plant transformation on Arabidopsis thaliana and showed that the tannin biosynthesis is controlled by this gene that encodes a WD40 protein which is an approximately 40 amino acid repeat that can be sites for the assembly of protein complexes and for protein-protein interactions.

In addition, differences in molecular weight distribution of tannins showed a significant impact on protein digestibility, protein binding and antioxidant capacity (Duodu et al., 2003). Therefore, it is possible to have high-tannin sorghum consisting of mostly large polymers that may not have significant impacts on protein functionality. The reverse of that would also hold true: relatively low tannin sorghum could have drastic effects on protein digestibility, protein binding, and antioxidant capacity if the tannins present consisted mainly of the very active molecular weight range (Duodu et al., 2003). Therefore, in protein digestibility studies, it is important to consider not only the tannin content but also its composition and molecular weight.
High tannin containing sorghum varieties are also known to negatively affect chicken and animal growth in general. Tannin-free sorghum was reported to have the potential to replace maize in broilers feeding with no difference in performance while high tannin sorghums significantly reduced chicken growth (Gualtieri and Rapaccini, 1990). In addition, in some countries like Italy, white tannin-free sorghum is preferred to yellow maize because the chicken produced would have white skin which is preferred over yellow-skinned broilers (Gualtieri and Rapaccini, 1990).

Despite the negative effects of tannins, in some parts of Africa, farmers prefer tannin sorghum varieties to other cereals because the porridge made from it gives them a long sensation of satiety during field work (Awika and Rooney, 2004). This might be linked to the low digestibility of the nutrients due to the tannin content. Moreover, tannins are excellent antioxidants and when fed to animals (rabbits, pigs, or poultry) showed that tannins could play a key role in fighting against weight gain however, additional studies need to be done to prove the effect on humans (Jambunathan and Mertz, 1973; Featherston and Rogler, 1975; Al-Mamary et al., 2001). Tannin sorghum varieties are also appreciated and grown in areas where pests and diseases are common because they are more tolerant, and grains suffer less attack by birds (Awika and Rooney 2004).

According to Awika and Rooney (2004), there is white sorghum (white pericarp) with no detectable tannins, red sorghum (red pericarp) with no tannins, and brown sorghum with significant tannin content (different levels of pericarp coloration). White grained sorghum seems to be more susceptible to diseases such as grain mold even though they seem to be more appreciated by populations who prepare foods from sorghum. White sorghum is supposed to be more digestible because it lacks a pigmented testa. However, there are white sorghum grains with low protein digestibility.
2.4 Methods of improving protein digestibility

2.4.1 Fermentation and malting

Fermenting sorghum before or after wet cooking leads to modified prolamins and glutelins (non-soluble sorghum proteins) that become more available for digestion by pepsin enzymes hence improving the in vitro protein digestibility (Axtell et al., 1981; Taylor and Taylor, 2002). Natural fermentation applied traditionally in some African processed sorghum food like “injera” (traditional Ethiopian bread) and “ting” (traditional sorghum porridge from Southern Africa), is an efficient technique for improved cooked sorghum protein digestibility. The fermentation process involves a reduction of sorghum soluble proteins. On the contrary, malting, which is also reported to increase sorghum digestible proteins, induces the breakdown of proteins by enzymes to produce free amino acids and soluble proteins more prone to enzyme digestion (Bhise et al., 1988). Even though fermentation and malting have a positive impact on the digestibility of sorghum proteins, it is still important to find and/or develop sorghum that will be highly digestible after cooking since not all food and beverages made from sorghum are fermented or malted.

2.4.2 Mutagenesis

Mutagenesis, in general, is a process by which the genetic information of an organism is modified resulting in a mutation. It might occur spontaneously in nature, or because of exposure to mutagens (Singer and Kusmierek, 1982). It could also be achieved experimentally using laboratory procedures (Xin et al., 2008). In agriculture, there is mutation breeding which is the process of exposing seeds and other plant materials to chemicals or radiation to generate mutants with desirable traits. The mutants (seeds or plants) obtained could be used, after the mutations are fixed and stable through series of selfing, in breeding programmes to improve cultivars through the process of crossing. For example, ethyl methanesulfonate (EMS) mutagenesis in agriculture, is the
application of a chemical, called EMS, to seeds to generate diversity or generate desirable traits in
plants. EMS is a chemical mutagen with the formula CH₃SO₃C₂H₅. It induces random point
mutations in genetic materials by nucleotide substitution; particularly by guanine alkylation
(Krieg, 1963). The ethyl group of EMS reacts with the DNA base guanine and forms an abnormal
base O-6-ethylguanine which pairs with thymine instead of cytosine. The latter mostly results in a
transition of G-C to T-A which could be either harmful (induce disease) or beneficial in breeding

Indeed, in 1975, research focusing on improving the lysine content of sorghum led to the discovery
of a high lysine-high digestible mutant line, P721Q (Q for opaque endosperm) (Mohan, 1975). P721Q
was generated through the application of diethyl sulphate on seeds of the poorly digestible
P721N (N for normal endosperm), which revealed a unique folded protein body structure with
high invaginations (Oria et al., 2000). Diethyl sulphate was reported to show similar mutagenesis
effects as EMS (Hoffmann, 1980). P721Q mutant had a 25% increase of protein digestibility for
cooked samples because of a 36% increase of the lysine-rich kafirins in the seeds compared to the
wild type P721N (Mohan, 1975; Wu et al., 2013). In addition, the mutation in P721Q led to a more
stable peptide which indirectly augmented the lysine-rich kafirins which are reported to be
indirectly associated with the high digestible phenotype (Wu et al. 2013). The kernel of P721Q
mutant has a floury, soft texture which is not ideal for cooking applications, is attacked by birds
and susceptible to mold. However, the novel nutritional properties of this mutant have inspired
various research efforts aimed toward understanding and improving this highly digestible (HD)
line. With the aim of having a HD sorghum line with improved seed quality, additional mutants
with high protein digestibility were generated and are yet to be characterized to understand the
genetic parameters governing sorghum protein digestibility in those mutants which would certainly
help enhance sorghum breeding and have a positive impact, particularly in African breeding programmes.

2.4.3 Marker assisted selection

Plant breeders have a common goal which is to develop lines or varieties with higher value traits. This can be achieved either through conventional methods or with the application of molecular techniques. Conventional methods involve multiple crossing generations and precision phenotyping to identify the progeny with the trait(s) of interest which can be time-consuming not to mention potential linkage drag. Therefore, to help accelerate breeding programmes, it is good to combine conventional breeding with tools like marker-assisted selection (MAS). MAS is the use of DNA markers to accelerate and successfully identify genotypes, at the DNA level, carrying the marker or genes associated with the trait of interest (Kumar, 1999). This can be done at an early stage of development of the plant cutting on the need to phenotype at every generation hence is more time and cost effective. One of the strengths of MAS is its success in identifying QTLs for traits hard to phenotype and/or polygenic traits. However, MAS requires high-quality phenotypic data which is obtained using conventional methods. Wu et al. (2013) developed simple sequence repeat (SSR) markers linked to the floury phenotype in sorghum grain which was shown to be linked to the high digestible phenotype in P721Q mutant, however, there are no reports of any high protein digestibility sorghum variety developed using MAS. To this date, all high protein digestible sorghum lines were developed through conventional breeding from a cross using a mutant as donor parent hence the necessity to develop molecular tools to be implemented in breeding programmes for the development of high protein digestible sorghum varieties.
2.5 Gene discovery methods for sorghum improvement

2.5.1 Genetic Linkage analysis or QTL mapping

Alleles or markers that are closely located on the same chromosome have the tendency to be inherited together during sexual reproduction. This phenomenon is called genetic linkage. It is the non-random assortment of alleles or markers within the same chromosome. The closer the genes, the more likely they will be passed on together from parents to offspring and the smaller are the chance for genetic recombination or crossing over to occur (Collard et al., 2005). Genetic recombination underlies QTL mapping which is known as the process of identifying QTL associated with a phenotype. Quantitative trait loci (QTL) are regions on the chromosome associated with quantitative traits. The development of QTL mapping techniques has allowed the identification of chromosomal regions associated with complex traits (Seaton et al., 2002). This method, based on recombination occurring in a large segregating bi-parental population (F2, backcross, recombinant inbred lines (RILs), or doubled haploid, showed successful results with low-density marker coverage in different mapping populations (Collard et al., 2005). Furthermore, using QTL mapping technique on an F2 population developed from a highly digestible mutant (P850029) and low digestible inbred line (Sureno), Winn et al. (2009) reported the presence of two QTL linked to protein digestibility in sorghum grain, one locus from P850029 with a positive impact and the second which reduces digestibility.

However, the QTL mapping approach has a major limitation in identifying genes controlling key agronomical traits. Indeed, with this method, a small number of generations are needed to develop the bi-parental mapping population and a limited number of crossing-over events could occur. Hence, a large region is identified as encompassing a significant peak but may contain various candidate genes. This limitation can be overcome by using unrelated genotypes that have
accumulated a larger number of recombinations through the years. Such an approach is called genome-wide association studies (GWAS) which allowed the identification of candidate loci associated with many traits in animal as well as plant species (Appels et al., 2013; Korte and Farlow, 2013).

### 2.5.2 Genome wide association studies

Genome-wide association study (GWAS) is a method for finding significant associations between genotypes and phenotypes of interest. It is based on natural variations that exist in a collection of unrelated genotypes or association panels. The GWAS method can be considered as complementary to the classical bi-parental QTL mapping technique, especially for complex traits. In Arabidopsis thaliana, carrying out GWAS along with QTL mapping as well as GWAS followed by candidate gene identification have shown successful results (Sonah et al., 2015; Verslues et al., 2014).

GWAS has with its own limitations. In GWAS, there can be limited linkage disequilibrium between pairs of linked markers. Therefore, to be able to entirely cover the genome, there is a need to use a greater number of markers as well as genotypes. Sukumaran et al. (2012) reported eight significant marker-trait associations and three significant SNPs related to starch synthase and starch content in a population of 300 sorghum lines. In addition, Morris et al. (2013) applied GWAS for the identification of SNPs linked to plant height components and inflorescence architecture using a population of 971 sorghum genotypes. However, in the case of protein digestibility in sorghum, application of GWA methods could be challenging since no information is reported about the existence of natural variability for protein digestibility hence the need to study it and provide knowledge on this valuable trait.
2.5.3 Application of Next Generation Sequencing (NGS) and QTL-seq

In the early ages of biological science, identifying candidate genes was quite challenging. From the chromosome walking method (Bender et al., 1983) to the advent of genomic tools, discovering interesting mutations or polymorphisms is becoming much easier. Whole genome sequencing applied to the EMS-mutagenized population of Caenorhabditis elegans allowed the rapid discovery of causative genes without the development of mapping populations (Sarin et al., 2008; Zuryn et al., 2010). Thole and Strader (2015) showed that it was faster than a map-based cloning strategy in identifying causative mutations or single nucleotide polymorphisms (SNPs) by comparing sequences of EMS mutants from Arabidopsis thaliana and sequence of wild types using next-generation sequencing (NGS) in backcrossed and non-backcrossed populations. Unlike GWAS, which detects shared mutations among genotypes, sequencing allows the identification of rare mutations that occur at a very low frequency but show strong association with the phenotype of interest. Another advantage of using NGS is that it is less laborious than QTL mapping in which hundreds of plants, as a mapping population, must be generated and PCR genotyped to narrow the chromosomal region containing the causative mutation (Collard et al., 2005). In addition, it is possible to achieve positional sequencing of the regions identified under significant peaks during GWAS or linkage studies in order to identify rare alleles associated with phenotypes observed (Koboldt et al., 2013). Furthermore, it was possible to narrow down a large region found from GWAS of multiple sclerosis to three times smaller size using targeted capture and NGS (Koboldt et al., 2013).

On the other hand, QTL-seq is a new and rapid method for the identification of causative loci that combines the QTL method to whole genome sequencing of DNA. Takagi et al. (2013) compared DNA sequences from two bulked contrasting populations for the phenotype of interest. It was first
applied to rice and led to the identification of QTLs for important agronomic traits (Takagi et al., 2013).

2.5.4 Bulked segregant analysis

Bulked segregant analysis (BSA) is a mapping technique that involves comparing allele frequencies of two bulks or pools containing genetically similar samples, with contrasting phenotypes, developed from a segregating population to identify markers associated to a specific gene or genomic region (Michelmore et al., 1991). The bulk samples differ at the causative locus/loci: one bulk with the wild type phenotype (containing the wild type allele) and another one with the mutant phenotype (having the causative allele). This method was first successfully applied to identify markers linked to a gene for resistance to downy mildew in lettuce using RFLP and RAPD (Michelmore et al., 1991). Subsequently, BSA showed successful results for multiple crops through the use of AFLP (Asnaghi et al., 2004), SSRs (Shen et al., 2003), and SNPs (Trick et al., 2012) markers. Through BSA, Massafaro et al. (2016) identified a genomic region significantly associated with high protein digestibility in the high lysine-highly digestible mutant P721Q however, the exact gene mutated remained unknown. BSA along with fine mapping, applied to the high protein digestible sorghum EMS mutants, could help in the identification of the genes controlling the trait.
CHAPTER THREE

3.0 Characterization of a West African Sorghum Association Panel for Protein Digestibility and other Quality Traits

3.1 Introduction

Sorghum is a major source of food in the world’s semi-arid regions especially Asia and Africa where it is a staple crop for millions of people. Africa provides at least 58% of the total harvested land area of sorghum where it constitutes the second most important cereal crop, after maize, in terms of area and production per year (FAO, 2018). The presence of anti-nutritional factors like tannin, lignin, and phytic acids limits the availability of the nutrients present in this valuable crop (Duodu et al., 2003). Therefore, it is desirable to have a sorghum crop with higher nutritional value that will reduce the need for additional cost especially in areas with people with low income. It is in that perspective that the genetic variability of sorghum needed to be assessed to identify varieties with better nutritional value.

Genetic variability studies of some traits revealed that they are not easily found in nature hence must be induced chemically or through genetic engineering to generate more variability. It is the case of brown midrib and lysine traits on sorghum. Sorghum, just like most cereal grains, has low lysine content (Deyoe and Shellenberger, 1965). However, the first two naturally occurring high lysine sorghum lines were found from Ethiopia after screening a World collection of 9000 sorghum accessions (Singh and Axtell, 1973). This low occurrence led to the use of mutagenesis to generate more genotypes with high lysine (Mohan, 1975). Similarly, sorghum is reported to have significantly lower digestible proteins after wet cooking (Maclean Jr et al., 1981). However, no study was done for protein digestibility using a large collection of
sorghum lines to search for naturally occurring sorghum lines with high digestible proteins after wet cooking. Massafaro (2015), using a diversity panel of only 53 sorghum lines found none to be highly digestible. This number might not be a good representation of the diversity that exists in sorghum. Most studies related to high protein digestibility in sorghum grain involve the use of sorghum mutants (Weaver et al., 1998; Oria et al., 2000; Wu et al., 2012; Massafaro et al., 2016). Furthermore, no prior knowledge of the variability for protein, protein digestibility, tannin and other seed quality traits on collections of West African sorghum accessions is available. Therefore, it is important for breeders to first assess these valuable traits to provide key information on these lines. In fact, prior to starting a plant breeding programme, it is crucial to identify sources of breeding material for the trait(s) of interest as well as the available cultivars. In addition, knowing the interaction and correlation between the traits of interest is key for the indirect selection of traits difficult to screen. Consequently, understanding the variability for seed traits, composition, digestibility, and their correlations is beneficial for designing optimal breeding strategies as well as finding suitable germplasm to use in breeding programmes for the improvement of sorghum nutritional quality and yield components.

This study’s focus was to:

1. characterize phenotypically and genotypically a collection of sorghum accessions from four countries in West Africa (Mali, Niger, Senegal, and Togo) for protein digestibility and related traits.
2. examine the correlation between traits, and
3. identify promising genotypes to be used in breeding programmes for the improvement of local germplasm for protein digestibility after wet cooking.
3.2 Materials and methods

3.2.1 Plant material

This study was done using a West African Sorghum Association Panel (WASAP) comprising 385 lines collected from four countries: Senegal (73 lines), Mali (38 lines), Niger (150 lines) and Togo (124 lines). The collection was made of local and improved varieties assembled by Centre d’Etude Régional pour l’Amélioration de l’Adaptation à la Sécheresse (CERAAS) through the support of the USAID funded SMIL (Sorghum and Millet Innovation Lab) project. The accessions were planted in Senegal during 2015 rainy season to produce uniform seeds for laboratory experiments where specific experimental designs were used.

3.2.2 Phenotypic characterization

3.2.2.1 Phenotyping the WASAP for seed colour

The harvested seeds were phenotyped for seed colour. Scoring based on visual examination was done and genotypes were classified as white or coloured based on their pericarp colour. The later comprised all non-white seeds: red, yellow, and brown.

3.2.2.2 Determination of protein digestibility of the WASAP

The modified digestibility assay from Aboubacar et al. (2003) was used to measure protein digestibility in a row by column design on a 96 well plate. The modifications in this version involved wet cooking the samples with double distilled water. In addition, pepsin digestion was performed for 2 hours and digestibility was measured after 20 minutes of trichloroacetic acid (TCA) application. The purpose of the digestibility assay was to mimic the actual use of sorghum by humans through cooking and digestion using pepsin enzymes naturally present in the organism. Seeds were obtained from a single head for each entry and were tested in triplicate. The low
digestible BTx623 and the highly digestible mutant, P721Q, were used respectively as negative and positive controls for the assay performed as follows.

On the first day, 60 ± 2 mg of seeds were ground in 830 µL of double distilled water. Before grinding, the seeds were first placed in a 2 mL reinforced polypropylene tube with 6 ceramic beads of 2.8 mm size. The samples were ground in the Omni bead ruptor 24 from Omni International at 6.5 m/s for 60 seconds and grinding repeated three times or until the product was completely homogenized. Samples were then cooked into a porridge for 20 min at 95°C in a VWR rotating oven model 5420. The samples were left at room temperature to cool down and then vortexed one at a time to get a homogenous solution. Three hundred microliters of each sample were then put in two different wells of a 96 deep well plate: one well to be digested with pepsin and the other one to be used as a control. The plate was then stored at 4°C overnight.

The second day, 300 µL of a Porcine Pepsin solution (9 mg/mL of 3300 U/mg of pepsin J.T Baker) was added to each sample to be digested, 300 µL of 0.2 M KH₂PO₄ at pH 2 was added to each control well and mixed by shaking. Digestion was performed in a rotating oven at 37°C for two hours. The digestion was stopped by adding 200 µL of 2 N NaOH to all 96 wells. The samples were then centrifuged at 3700 rpm for 5 minutes. The supernatant was poured off and 500 µL of 0.1 M KH₂PO₄ pH 7 was added to each well to neutralize the pH, then vortexed for 5 minutes and centrifuged at 3700 rpm for 5 minutes. The supernatant was poured off and each sample was washed by adding 500 µL of double distilled water to each well and the plate vortexed for 5 minutes then centrifuged at 3700 rpm for 5 minutes. The supernatant was poured off and the pellets resuspended in 500 µL of extraction buffer pH 10 (1% SDS (w/v), 0.05% beta-mercaptoethanol (v/v), 0.0125M sodium tetraborate solution pH 10). Following an hour incubation in a cold room on a shaking incubator at 250 rpm at 25°C, the plates were centrifuged at 3700 rpm for 20 minutes.
Twenty microliters of the supernatant were then transferred to a microtiter plate with 200 µL autoclaved distilled water and 50 µL of 72% trichloroacetic acid (w/v) was added to all samples to precipitate the remaining protein. The plate was set aside at room temperature for 20 minutes before reading the absorbance at 562 nm (A_560) using an EL x 800 UV spectrophotometer (BioTek Instruments, Inc.).

**3.2.2.3 Phenotyping for tannin using a Bleach test**

To check the presence of tannin in the seeds, a bleach test adapted from Waniska *et al.* (1992) was used. Hundred millilitres of 3.5% sodium hypochlorite solution (commercial bleach) was used to dissolve 5 g of sodium hydroxide (NaOH). The liquid mixture was used to soak ten (10) seeds of sorghum per entry. Seeds were just covered on the surface to avoid overbleaching which could lead to false positive results. The tests were performed in duplicate. The mixture of bleach and NaOH work by melting away the pericarp layer and exposing the presence of a pigmented testa layer or not. When the testa layer turned black after 15 minutes incubation it revealed the presence of tannins through the appearance of darker coloured seeds hence the sample was recorded as having tannin. However, in the case of a tannin-free line, there was no change in colour in the seed. A count of the number of seeds with tannin i.e. displaying a dark colour after bleach test was done and samples were scored as follows: one (1) for tannin positive, and zero (0) for tannin-free.

**3.2.2.4 Determination of Starch, protein, total digestible nutrients and acid detergent fiber using near-infrared reflectance spectroscopy**

About 3 to 4 g (depending on the line) of whole seeds of each of the WASAP accessions were put in the micro mirror module sample holder provided by Perten, until it was almost full (seeds were not superimposed). The seeds were analysed using a Perten DA 7250 Near-Infrared Spectroscopy
(NIRS) analyser. Each sample was tested in triplicate with seeds been reshuffled after each measurement. The mean value was computed. To obtain the values recorded in this study, pre-calibrated standard curves were generated for each trait through partial linear square regression of wet chemistry data to spectra data to ensure accurate measure and repeatability of each trait applied to a wide range of samples (unpublished). Data was generated for the following traits/mareas: starch (1667-2500 nm), protein content (2052 nm), total digestible nutrients, and acid detergent fiber (1724, 1756, 2308, and 2348 nm). The moisture content of the seeds was also determined.

3.2.3 Genotyping by sequencing

DNA of the WASAP was previously extracted from young dried leaves using the Mixed Alkyl Trimethyl Ammonium Bromide (MATAB) DNA extraction protocol (Risterucci et al., 2000) modified at CERAAS, Senegal. The major modifications involved pre-heating the extraction buffer to 65°C instead of 75°C and DNA being resuspended in TE buffer or distilled water instead of NaCl. The DNA was sent to Kansas State University for genotyping by sequencing (GBS) using the method described by Elshire et al. (2011). The DNA was first digested by the enzyme ApeKI followed by unique barcoding for each sample. PCR amplification was performed on a 96 well plate followed by gel purification using Qiagen PCR purification kit. Quality of the amplified DNA was checked using a ThermoScientific NanoDrop spectrophotometer. Sequencing was done through Illumina HiSeq 2500 single end with 100 bp reads. The GBS data obtained were filtered removing the adapters then aligned to the sorghum reference genome. A total of 105,226 SNPs was obtained before site filtering.
3.2.4 Data analysis

3.2.4.1 Protein digestibility

Raw absorbance values (A_{560}) of the tested entries obtained from the digestibility assay were normalized against the absorbance of the TCA-buffer well. The percentage digestibility (PD) was calculated using the normalized values based on the following formula:

$$PD\ (\%) = \left[\frac{(A_{560\ Undigested} - A_{560\ of\ Digested})}{A_{560\ Undigested}}\right] \times 100$$

The threshold for high protein digestibility was set at 60%.

Analysis of variance (ANOVA) for protein digestibility was done with the following formula:

$$\text{lm (PD ~ Genotype + Row: Plate + Column: Plate + Plate)}$$

Variance components were calculated from the linear model fitted using the lme4 package in R version 3.5.1 with all significant parameters from the ANOVA being set as random.

$$\text{lmer (Protein digestibility ~ (1|Genotype) + (1|Column: Plate) + (1|Plate))}$$

In addition, using the variance components, broad sense heritability was calculated for PD as follows:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 / r}$$

Where $H^2$ is the broad sense heritability, $\sigma_g^2$ is the genotypic variance, $\sigma_e^2$ is the error variance, and $r$ is the number of replications.

3.2.4.2 Genetic diversity of WASAP

TASSEL 5.0 was used for GBS data filtering through the GBS v2 pipeline (Glaubitz et al., 2014). SNPs with more than 10% of missing data and with minor allele frequency (MAF) less than 5%
were removed. The files were first filtered for sites that had more than two alleles and too many missing data. The filter sites were as follows: minimum frequency 0.01, maximum frequency 0.99, minor allele frequency 0.05. Only 73,512 SNPs were kept after filtering.
3.3 Results

3.3.1 Distribution of seed colour of the collection

WASAP accessions comprised white (43.64%) and coloured (56.36%) seeds with different proportions based on the country of origin (Table 3.1). The coloured seeds comprised red, yellow and brown.

Table 3.1 Distribution of seed colour in WASAP

<table>
<thead>
<tr>
<th>Origin</th>
<th>Colour</th>
<th>Count</th>
<th>(%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mali</td>
<td>white</td>
<td>28</td>
<td>16.67</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>coloured</td>
<td>10</td>
<td>4.61</td>
<td></td>
</tr>
<tr>
<td>Niger</td>
<td>white</td>
<td>66</td>
<td>39.29</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>coloured</td>
<td>84</td>
<td>38.71</td>
<td></td>
</tr>
<tr>
<td>Senegal</td>
<td>white</td>
<td>54</td>
<td>32.14</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>coloured</td>
<td>19</td>
<td>8.76</td>
<td></td>
</tr>
<tr>
<td>Togo</td>
<td>white</td>
<td>20</td>
<td>11.90</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>coloured</td>
<td>104</td>
<td>47.93</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>white</td>
<td>168</td>
<td>43.64</td>
<td>385</td>
</tr>
<tr>
<td></td>
<td>coloured</td>
<td>217</td>
<td>56.36</td>
<td></td>
</tr>
</tbody>
</table>

3.3.2 Variation for protein content

Protein values ranged from 7 to 18 g/100 g of samples (Figure 3.1). Samples from Niger were found to have the largest variability for protein content (7 to 18 g) followed by Senegal (10 to 17 g) then Mali (10 to 16 g) and Togo had the least variability for protein (12 to 16 g) (Figure 3.3). The top 20 accessions of the WASAP for protein content are reported in table 3.2.
Figure 3.1 Distribution of protein content of all WASAP lines.

Table 3. 2 Top 20 accessions of the WASAP for protein content

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Protein</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ni133</td>
<td>18.22</td>
<td>Niger</td>
</tr>
<tr>
<td>S410</td>
<td>17.88</td>
<td>Niger</td>
</tr>
<tr>
<td>S173</td>
<td>17.34</td>
<td>Niger</td>
</tr>
<tr>
<td>S229</td>
<td>17.10</td>
<td>Niger</td>
</tr>
<tr>
<td>Sn110</td>
<td>17.06</td>
<td>Senegal</td>
</tr>
<tr>
<td>Ni120</td>
<td>16.94</td>
<td>Niger</td>
</tr>
<tr>
<td>S148</td>
<td>16.77</td>
<td>Niger</td>
</tr>
<tr>
<td>K72.1</td>
<td>16.58</td>
<td>Niger</td>
</tr>
<tr>
<td>Tg116</td>
<td>16.53</td>
<td>Togo</td>
</tr>
<tr>
<td>Sn47</td>
<td>16.51</td>
<td>Senegal</td>
</tr>
<tr>
<td>Sn108</td>
<td>16.45</td>
<td>Senegal</td>
</tr>
<tr>
<td>ML196</td>
<td>16.43</td>
<td>Mali</td>
</tr>
<tr>
<td>Ni57</td>
<td>16.41</td>
<td>Niger</td>
</tr>
<tr>
<td>Ni340</td>
<td>16.35</td>
<td>Niger</td>
</tr>
<tr>
<td>Sn63</td>
<td>16.34</td>
<td>Senegal</td>
</tr>
<tr>
<td>S227</td>
<td>16.33</td>
<td>Niger</td>
</tr>
<tr>
<td>S151</td>
<td>16.21</td>
<td>Niger</td>
</tr>
<tr>
<td>S483</td>
<td>16.16</td>
<td>Niger</td>
</tr>
<tr>
<td>Sn1</td>
<td>16.16</td>
<td>Senegal</td>
</tr>
<tr>
<td>S457</td>
<td>16.12</td>
<td>Niger</td>
</tr>
</tbody>
</table>
### 3.3.3 Variation for protein digestibility

There was a significant difference ($P < 0.001$) between genotypes, plate and columns within the 96 well plate in which the digestibility assay was performed (Table 3.2). Most of the variation was found within genotypes. Broad-sense heritability was 39.8%.

Table 3.3: Mean square, variance components and broad sense heritability of protein digestibility

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Mean Square</th>
<th>Variance components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>386</td>
<td>1146.12 ***</td>
<td>50.58</td>
</tr>
<tr>
<td>Plate</td>
<td>35</td>
<td>629.94 ***</td>
<td>28.09</td>
</tr>
<tr>
<td>Row: Plate</td>
<td>252</td>
<td>46.39 ns</td>
<td>N/A</td>
</tr>
<tr>
<td>Column: Plate</td>
<td>180</td>
<td>56.78 *</td>
<td>2.37</td>
</tr>
<tr>
<td>Residuals</td>
<td>733</td>
<td>44.41</td>
<td>45.96</td>
</tr>
<tr>
<td>Heritability</td>
<td></td>
<td></td>
<td>0.398</td>
</tr>
</tbody>
</table>

*Df: degree of freedom; ns: not significant; * significant at $P = 5\%$; *** significant at $P < 0.001$*

Protein digestibility (PD) of the WASAP accessions ranged from 1 to 55% (Figure 3.2) with samples from Niger having most variability, followed by Togo, Senegal, and Mali (Figure 3.4). None of the 385 accessions of the WASAP was found to be highly digestible i.e. with 60% digestibility or higher. The top 20 samples with the highest values of PD were from Niger and Togo (Table 3.4).
Figure 3.2 Distribution of protein digestibility of WASAP lines.

Table 3. 4 Top 20 accessions of the WASAP for protein digestibility

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Protein digestibility</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>K26.3</td>
<td>53.17</td>
<td>Niger</td>
</tr>
<tr>
<td>K100</td>
<td>50.38</td>
<td>Niger</td>
</tr>
<tr>
<td>K79</td>
<td>49.70</td>
<td>Niger</td>
</tr>
<tr>
<td>K83</td>
<td>46.65</td>
<td>Niger</td>
</tr>
<tr>
<td>K49.2</td>
<td>45.02</td>
<td>Niger</td>
</tr>
<tr>
<td>K34</td>
<td>43.66</td>
<td>Niger</td>
</tr>
<tr>
<td>K32.3</td>
<td>43.16</td>
<td>Niger</td>
</tr>
<tr>
<td>K42.1</td>
<td>43.01</td>
<td>Niger</td>
</tr>
<tr>
<td>K31.2</td>
<td>42.93</td>
<td>Niger</td>
</tr>
<tr>
<td>K38.1</td>
<td>42.38</td>
<td>Niger</td>
</tr>
<tr>
<td>K45.2</td>
<td>42.33</td>
<td>Niger</td>
</tr>
<tr>
<td>K46.3</td>
<td>40.98</td>
<td>Niger</td>
</tr>
<tr>
<td>K76.2</td>
<td>40.87</td>
<td>Niger</td>
</tr>
<tr>
<td>K36.1</td>
<td>40.38</td>
<td>Niger</td>
</tr>
<tr>
<td>Tg16</td>
<td>39.02</td>
<td>Togo</td>
</tr>
<tr>
<td>K40.1</td>
<td>38.58</td>
<td>Niger</td>
</tr>
<tr>
<td>K45.1</td>
<td>38.30</td>
<td>Niger</td>
</tr>
<tr>
<td>K28.2</td>
<td>37.21</td>
<td>Niger</td>
</tr>
<tr>
<td>K31.3</td>
<td>37.01</td>
<td>Niger</td>
</tr>
<tr>
<td>K19.1</td>
<td>36.51</td>
<td>Niger</td>
</tr>
</tbody>
</table>
3.3.4 Tannin in seeds of WASAP

There were 275 tannin-free and 110 tannin containing lines (Table 3.3). Different levels were seen between samples based on their country of origin. Accessions from Togo had the highest number of accessions containing tannin (40%), followed by Niger (30%), Senegal (21.82%), then Mali (8.18%).

White seeds without testa containing tannin as well as white seeds with coloured testa showing the presence of tannins (Figure 3.3) were identified. In addition, some coloured seeds without tannin and others with tannin (Figure 3.3) were also reported.

Table 3.5 Presence of tannins in WASAP seeds

<table>
<thead>
<tr>
<th>Origin</th>
<th>Bleach test</th>
<th>Count</th>
<th>(%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mali</td>
<td>Tannin-free</td>
<td>29</td>
<td>10.55</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>With Tannin</td>
<td>9</td>
<td>8.18</td>
<td></td>
</tr>
<tr>
<td>Niger</td>
<td>Tannin-free</td>
<td>117</td>
<td>42.55</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>With Tannin</td>
<td>33</td>
<td>30.00</td>
<td></td>
</tr>
<tr>
<td>Senegal</td>
<td>Tannin-free</td>
<td>49</td>
<td>17.82</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>With Tannin</td>
<td>24</td>
<td>21.82</td>
<td></td>
</tr>
<tr>
<td>Togo</td>
<td>Tannin-free</td>
<td>80</td>
<td>29.09</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>With Tannin</td>
<td>44</td>
<td>40.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Tannin-free</td>
<td>275</td>
<td>71.43</td>
<td>385</td>
</tr>
<tr>
<td></td>
<td>With Tannin</td>
<td>110</td>
<td>28.57</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.3: Bleach test analysis results. A- White Sorghum seed without tannin containing testa layer. B- White Sorghum seed with tannin-containing testa layer. C- Coloured Sorghum seed with tannin-containing testa layer. D- Coloured Sorghum seed without tannin containing testa layer. Seeds in the bags are before the bleach test. In each sample, both replicates displayed similar results.

3.3.5 Distribution of acid detergent fiber, total digestible nutrients, and starch in WASAP

Box plots of all traits showed normally distributed data (Figure 3.4). Acid detergent fiber (ADF) ranged from 3 to 9 and total digestible nutrients (TDN) from 80.5 to 89. Range variability for starch content for all 4 populations in the WASAP was from 49 to 65. Variation in samples for each trait and per country of origin showed that for all three traits measured there was no difference in mean values between the accessions from the four different countries; nonetheless, there was a difference in variability especially for total digestible nutrients. In fact, similar variation was
observed in samples from Niger and Togo in terms of TDN followed by Senegal then Mali (Figure 3.4). Samples with the highest TDN had the least ADF. Samples from Niger and Senegal displayed most of the variability for all traits measured (Figure 3.4). In addition, samples from Niger, even though not highly digestible, had higher digestibility values than other accessions and had lower values of ADF (Figure 3.4).
Figure 3.4: Boxplots for Protein digestibility, Protein content, Starch content, Total digestible nutrients, and Acid detergent fiber for accessions from Mali, Niger, Senegal, and Togo.
3.3.6 Correlations between traits

There were significant positive correlations between protein digestibility (PD) and total digestible nutrients ($r = 0.45$), and starch content ($r = 0.41$) ($P < 0.001$) (Figure 3.5). However, negative correlation coefficients were detected between PD and the following: protein content ($r = -0.59$), acid detergent fiber ($r = -0.34$), seed colour ($r = -0.17$) and tannin content ($r = -0.32$). A positive correlation was found between tannin and seed colour ($r = 0.1$) ($P = 0.05$). In addition, negative correlations between TDN and ADF ($r = -0.96$) as well as TDN and tannin ($r = -0.53$) were detected in the WASAP lines ($P < 0.001$).

Figure 3. 5: Correlation matrix of seven measured traits on WASAP accessions. Frequency distribution of each variable on the diagonal, the bivariate scatter plots with a fitted line on the bottom of the diagonal, values of the correlation plus the significance level as stars on the top. “*” Significant at $P = 0.05$, “**” Significant at $P = 0.01$, “***” Significant at $P < 0.001$. 

University of Ghana http://ugspace.ug.edu.gh
3.3.7 Genetic diversity in the West African sorghum collection

A high level of heterozygosity of up to 50% was found in the WASAP lines with most of the accessions (200) having 42% of heterozygosity (Figure 3.6). Approximately 130 samples had heterozygosity less than 42% and 55 accessions which heterozygosity greater than 42%.

![Heterozygosity level of lines in the West African Sorghum Association Panel.](http://ugspace.ug.edu.gh)

Figure 3.6: Heterozygosity level of lines in the West African Sorghum Association Panel.

WASAP accessions were clustered in three main groups: group 1 with only samples from Niger,
group 2 with only samples from Togo, and group 3 comprising all samples from Senegal and Mali as well as some samples from Niger and Togo (Figure 3.6). Several samples from Niger and Togo seem to be genetically unique to those locations.

Figure 3.7: Genetic distribution of WASAP accessions.
3.4 Discussion

Sorghum, unlike other cereal grains, becomes less digestible after wet cooking. Many factors are known to affect the digestibility of cooked sorghum. Some of the factors are endosperm texture, protein body structure (Oria et al., 2000), protein cross-linking, as well as grain composition to name a few. In this study, a negative correlation was found between protein digestibility, protein content, and acid detergent fiber, as well as protein digestibility and tannin content (Figure 3.5). These results confirmed previous findings from in vitro (Armstrong et al., 1974) and in vivo (Rostagno et al., 1973) studies which reported the negative impact of tannin (Chibber et al., 1980), and ADF (Bach and Munck, 1985) on digestibility of sorghum proteins.

Tannins are natural polyphenols found in many plants. The negative correlation between protein digestibility and tannin content could be explained by the fact that tannins are known to bind and precipitate fairly large, loose proteins with an open structure and rich in proline (Armstrong et al., 1974; Butler et al., 1984). In sorghum, tannins are one of three main phenolic compounds found in the grain along with phenolic acids and flavonoids (Hahn et al., 1984). Even though tannins have a protective role on the grain against birds, insects and fungi damage (Awika and Rooney, 2004), their presence in the grain leads to a reduction in nutrition and food quality (Serna-Saldivar and Rooney, 1995). Depending on their structure and abundance, they can have different levels of impact on the digestibility of proteins for animals and humans before and after wet cooking. A sensation of long satiety was reported by consumers of sorghum meals (Duodu et al., 2002) which could be due to the slow digestion of the nutrients present in tannin-rich varieties. Processing methods of sorghum grain such as grinding and/or boiling may enhance tannin-protein interactions limiting their availability to proteolytic enzymes either before or after cooking (Butler et al., 1984). Nonetheless, not only tannin containing varieties have less digestible proteins. This study showed
that among the 385 accessions tested, 275 (71.4%) had tannin (Table 3.3) implying that this phenomenon also occurs on tannin-free varieties (28.6%) confirming the results of Maclean Jr et al. (1981) and Mertz et al. (1984).

A weak but significant correlation ($r = 0.1$) was observed between tannin and seed colour which was also reported by Hahn and Rooney (1986). Two hundred and seventeen coloured-seeded accessions were present in the sorghum panel, however, the total number of entries with detectable tannins was only 110 and included white coloured seeds which implied that the presence of tannins in sorghum seeds is not best predicted by seed colour because phenolic compounds like tannins can modify the natural colouring of the upper layers of sorghum grain (Rooney and Miller, 1981; Waniska, 1992).

Other factors present in the grain or interactions of multiple factors could lead to the low digestibility of sorghum proteins. The negative impact of ADF on protein digestibility could be explained by the fact that ADF refers to the portion of the forage that consists of lignin and cellulose and is linked to how easily the plant material can be digested by animals (Van Soest and Robertson, 1985). Therefore, the more ADF a plant has, the least digestible it will be. ADF is measured from the fodder; however, Font et al. (2003) showed that it is possible to measure ADF on whole as well as ground seeds using NIRS. In addition, the negative correlation between protein content and protein digestibility could be due to the association between proteins, the pericarp, starch granules and protein bodies (Glennie, 1984; Bach and Munck, 1985). In sorghum, the starch granules are surrounded by spherical protein bodies. During the process of wet cooking, the gelatinized sorghum starch could limit the availability of the proteins to the proteolytic enzymes or lead to the formation of non-digestible complexes due to the organisation of the protein bodies (Shull et al., 1991). In addition, proteins are reported to be able to bind to ADF in cooked and
uncooked sorghum grains (Bach and Munck, 1985) due to their amino acid composition reducing the digestibility of the proteins.

Genotypes contributed most of the observed variation for protein digestibility (Table 3.2). A moderate broad sense heritability for protein digestibility (39.8%) was observed in the West African Sorghum Association Panel. The high genotypic variance observed can be exploited when breeding for protein digestibility in this set of germplasm.

Sorghum digestible protein after wet cooking was first reported to be 46% in a study involving malnourished children (Maclean Jr et al., 1981) however, Kurien et al. (1960) using the same methodology on under-nourished children, showed the digestibility of sorghum to be at 55%. Based on those results, the threshold for low digestibility in this study was set at below 60% to ensure a clear distinction between low and high protein digestible samples. Even though no highly digestible line was identified in the present study, the accessions from Niger presented more variability for protein and protein digestibility (Figure 3.4). Therefore, there is more potential for improvement of protein digestibility in the collection from Togo compared to the rest of the panel. Genetic variability is important for diversity in breeding since it helps in selecting promising material for the improvement of varieties for desirable traits.

The results of this study showed that the West African sorghum association panel is mainly divided into 3 groups. The first group with only samples from Niger, the second comprising Togolese germplasm, and the last group with a mixture of samples from all four countries (Mali, Niger, Senegal, and Togo) (Figure 3.7). The presence of a mixture of samples in the third group could be due to similarities in agroecological regions between Senegal and Mali and the sharing of plant material between neighbouring countries since all four countries of the panel are in West Africa.
and were involved in shared projects (example INTSORMIL). Though some accessions were found in all four countries, Togo and Niger had unique materials (groups 1 and 2) which could be explained by the fact that more than 50% of the samples from these two countries are local accessions obtained from farmers when the collection was being assembled. The latter could also explain the presence of a high level of heterozygosity (Figure 3.6) since the varieties have not been kept pure through selfing methods used by breeders on stations. The collection was made of improved varieties used in breeding programmes, of cultivars, and of local accessions.
3.5 Conclusion

This study confirmed that factors like tannins and acid detergent fibre reduced protein digestibility while seed colour was not a good indicator of protein digestibility even though it was slightly negatively correlated with protein digestibility. Samples from Niger had the largest variability for protein and protein digestibility, however, none of the WASAP accessions tested was found to have high digestible proteins after wet cooking hence cannot be used in breeding programmes as a source of highly digestible proteins. Nonetheless, knowledge of the characteristics of the WASAP accessions is provided by this study to facilitate their use in breeding programmes. The accessions were classified into three main groups: group 1 with entries from Niger only, group 2 with entries from Togo, and group 3 comprising all germplasm from Senegal and Mali as well as a portion of accessions from Niger and Togo.
CHAPTER FOUR

4.0 Characterization of highly digestible sorghum EMS mutants using Bulked Segregant Analysis

4.1 Introduction

Sorghum is a very important crop grown worldwide for various uses. Sorghum is used for human consumption, animal feeding and ethanol production to name a few. However, compared to other cereal grains like wheat, rice, or maize, when sorghum is wet cooked, its protein digestibility decreases to about 46% or less. This phenomenon leads to malnutrition, especially for people who rely on sorghum as one of their primary sources of protein (Maclean Jr et al., 1981). It is therefore important to solve this problem to contribute to food security.

Previous research on sorghum mutants led to the identification of a high lysine, highly digestible mutant named P721Q (Q for opaque endosperm) (Mohan, 1975; Weaver et al., 1998) generated from P721N (N for normal endosperm) through the application of diethyl sulphate. Although P721Q has 60% more lysine (Mohan, 1975) and is 25% more digestible than its parent line (Weaver et al., 1998), it has small soft seeds with floury endosperm, making it very susceptible to bird attacks and not very suitable for some African foods, thus limiting consumer acceptance. Therefore, it is desirable to combine the high protein digestible trait with better seed quality. To achieve that goal, mutant lines of BTx623 (the reference sorghum genome) that had been treated with ethyl methanesulfonate (EMS) were screened using a protein digestibility (PD) assay (described later). Three highly digestible mutants, SbEMS1613, SbEMS1227, and SbEMS3324 were identified (Massafaro, 2015) These mutants were genome re-sequenced with the support of The Bill and Melinda Gate Foundation and the EMS-induced mutations within predicted protein-
coding sequences previously identified (Addo-Quaye et al., 2018). The mutants are known to be highly digestible however the genetic mechanisms underlying this key trait remains unknown.

Few mapping studies were done to identify genomic regions associated with protein digestibility and involved quantitative trait loci (QTL) mapping (Winn et al., 2009), DNA amplification and sequence alignment followed by plant transformation (Wu et al., 2013) and bulked segregant analysis (Massafaro et al., 2016). These studies led to the identification of two QTLs (specific to P850029 developed from P721Q through breeding) on chromosome 1 with contrasting effect on digestibility: one significantly reducing digestibility and another one favourably increasing PD (Winn et al., 2009). However, additional studies (Wu et al., 2013; Massafaro et al. 2016) reported regions on chromosome 5 containing kafirin genes to be significantly linked to high protein digestibility, in P721Q, which is a simply inherited trait governed by a recessive allele (Wu et al., 2013). The increase in protein digestibility was hypothesized to be associated with the invaginated protein bodies phenotype in P721Q which probably resulted in the increase of the accessible area to the proteolytic enzymes hence probably augmenting the total digestible proteins (Oria et al., 2000). Wu et al. (2013) reported that the highly digestible protein phenotype in P721Q is linked to a point mutation on a kafirin gene that makes the protein resistant to processing. However, the exact kafirin gene copy mutated remained unknown. Therefore, investigating the genetics of this valuable agricultural trait in the EMS mutants has become very important.

The objectives of this study were to

1. characterize highly digestible EMS mutants,
2. investigate the molecular mechanisms underlying the increase in protein digestibility in the EMS mutants,
3. develop molecular markers that could facilitate rapid introgression of the digestibility trait into low digestible sorghum lines.
4.2 Material and methods

4.2.1 Plant materials

Sorghum EMS mutants (SbEMS1227, SbEMS1613, SbEMS3324 and SbEMS932) were developed at Purdue University from seeds of BTx623 treated with ethyl methanesulfonate (Krothapalli et al., 2013). The mutant seeds used in this study were at the M5 generation and identified as highly digestible from a population of mutants (Massafaro, 2015).

P721Q is a high lysine, high protein digestible mutant generated from P721N at Purdue University (Mohan, 1975). Seeds of maize varieties (W22, B73, and Mo17), EMS mutants, as well as sorghum lines (BTx623, Maia, CE151-262A, Tx430, TARG1, 03MN952, and MR732) were provided by Professors Cliff Weil and Mitch Tuinstra from Purdue University. The maize varieties were inbred lines from the USA that were used to assess how digestible sorghum proteins are compared to other cereals like maize.

4.2.2 Phenotypic characterization of the EMS mutants

4.2.2.1 Determination of percent protein digestibility

Protein digestibility of the EMS mutants and wild types (P721N and BTx623) was measured based on a modified version of the turbidity assay by Aboubacar et al. (2003) described in Chapter 3 of this thesis with the main differences being that in this version the samples were wet cooked with double distilled water and the digestion was performed with pepsin enzyme for 2 hours and digestibility was measured after 20 minutes after trichloroacetic acid (TCA) application. Each sample was treated under two conditions: uncooked and cooked. The purpose of the digestibility assay was to mimic real usage of sorghum grain as a feed and staple food. P721Q was used as a positive control for the assay.
Three lines of sorghum (P721N, BTx623, and SbEMS932) and 3 lines of maize (W22, B73, and Mo17) were tested in 5 replications, both uncooked and cooked.

4.2.2.2 Protein body morphology of sorghum mutants

Protein bodies of the three highly digestible EMS mutants and their low digestible parent, BTx623, were phenotyped at Purdue University using transmission electron microscopy. Samples were prepared following the procedure described by Oria et al. (2000). Seeds were grown at Purdue Agronomy Center for Research and Education (ACRE), West Lafayette, the USA during summer of 2016 and harvested 30 days after half bloom and imaged with a transmission electron microscope (Philips CM-100, FEI Company). The microscope was run at accelerating voltage of 100 kV with the following settings: spot 3,200 µm condenser aperture and 70 µm objective aperture. Photos were taken with a Gatan digital camera with diverse magnifications.

4.2.2.3 Determination of protein, total lysine and tryptophan contents

Five grams of seeds of P721Q, BTx623, SbEMS1613 and SbEMS3324 were sent to the University of Missouri, the USA for amino acid analysis. Total lysine and tryptophan contents were measured. Tryptophan content was measured using the enzymatic hydrolysis - colorimetric determination method described by Draher and White (2017). Total lysine was determined with the Association of Official Agricultural Chemists (AOAC) approved acid hydrolysis method.

4.2.2.4 Measurement of seed diameter

Seed diameters of P721N, BTx623, P721Q, SbEMS1227, SbEMS1613, and SbEMS3324 were measured using a digital calliper. Five mature dried seeds were randomly selected for each sample. The measurements were conducted transversely in the middle part of the seed. Each measurement was recorded, and a mean value was calculated.
4.2.1.5 Determination of seed hardness

Seed hardness measurements of the highly digestible sorghum mutants (P721Q, SbEMS1227, SbEMS1613, and SbEMS3324) and wild types (BTx623 and P721N) were conducted at the Purdue Mechanical Engineering Department using an MTS 858 Mini Bionix instrument designed for static testing of material strength. The measurements were expressed as the amount of force (in Newtons) needed to break a single seed fed to the instrument. Each genotype was measured two times and an average hardness value was calculated.

4.2.3 Identification of allele(s) and gene(s) controlling high protein digestibility through bulked segregant analysis (BSA)

4.2.3.1 Mapping population development

To investigate the genetic mechanism controlling high protein digestibility, mapping populations were developed from two EMS mutants derived from BTx623. Two highly digestible white midrib mutants showing contrasting protein body phenotype, SbEMS1613 and SbEMS3324, were each crossed as males onto SbEMS932, a low digestible, brown midrib (bmr6) line from BTx623. The white midrib is dominant over brown therefore, all true F1 had a white midrib. Crosses were initiated in summer 2014 at ACRE, West Lafayette (Indiana, USA). The seeds obtained were planted the next summer at ACRE and successful F1 plants, identified by having a white midrib, were self-fertilized to generate F2 seeds. F2 seeds were then planted in summer 2016 at ACRE and five hundred and sixty resulting F2 plants were self-fertilized to generate F3 seeds, which were then tested for protein digestibility in Dr Weil’s lab at Purdue University. The result was used for bulked segregant analysis.
4.2.3.2 Development of F₃ recombinant populations

Since alleles are passed from parent to offspring, additional F₃ recombinant populations were developed by crossing each highly digestible EMS mutant to four low digestible lines each starting summer 2015 using the plastic bags method. SbEMS1613 was crossed to CE151-262A, MR732, Tx430, and BTx623. These populations were used to search for conserved mutations between progeny from crosses made with the mutant as a shared parent.

SbEMS3324 was crossed to varieties Macia, TARG1, 03MN952, and MR732. Plant colour, as well as hybrid vigour, were the parameters used to select the successful F₁ plants which were self-fertilized to generate F₂ seeds in 2016. F₂ seeds were then planted during summer 2017 at ACRE and each head was covered prior to anthesis to ensure the production of selfed seeds. Thirty panicles from each generated population were tested for digestibility and the highly digestible progeny were Sanger sequenced (Sanger et al., 1977) using primers designed from the BSA results in this study. Sequence alignment was performed between sequences of the highly and lowly digestible F₂ of the mapping population and highly digestible F₃ recombinants.

4.2.3.3 Measurement of protein digestibility for the mapping populations

To check the percentage of digestible proteins in each sample of the mapping and F₃ validation populations, a protein digestibility assay was performed in three replicates within a panicle for each F₃ plant (sorghum single head) according to the methodology described in section 4.2.2.1. A total of 506 F₃ panicles were tested for SbEMS1613 mapping population and 455 F₃ panicles for SbEMS3324 mapping population.

Samples with consistently ≥ 60% protein digestibility were considered highly digestible while those with less than 60% digestibility were classified as lowly digestible. Samples that had a
mixture of high and low digestible values when the F$_3$ seeds were tested were scored as coming from a heterozygous F$_2$ plant. They were not included in the bulked segregant analysis. All the high and low digestible samples were tested two more times to confirm digestibility status.

### 4.2.4 Mapping of protein digestibility genes in EMS mutants

#### 4.2.4.1 DNA extraction and sequencing

One to two inches of the youngest fully expanded leaf were placed in an envelope loaded with one tablespoon of silica gel changed frequently to allow leaves to dry until DNA extraction could take place. Using a scaled-down version of the CTAB DNA extraction method (Allen et al., 2006), DNA was extracted as a pool for each mapping population’s samples separately. The concentration and purity of the DNA were checked on a Thermo Scientific Nanodrop 1000 Spectrophotometer and samples with 260/280 nm absorbance value between 1.8 and 2.0 were considered pure. A total of 2 ng of DNA in a volume of 120 µL for each pool was sent for whole genome sequencing at Purdue Genomics Core Center.

#### 4.2.4.2 Bulked segregant analysis for protein digestibility

To perform the bulked segregant analysis, samples with contrasting phenotype for protein digestibility were pooled into two bulks: bulk 1 with the highly digestible samples and bulk 2 with the lowly digestible samples. This was done for each mapping population. To make the low and high digestible bulks of SbEMS1613, the 50 lowest and the 50 highest digestible samples of the mapping population were pooled. The low and high bulks of SbEMS3324 were made of DNA from the 30 lowest and the 30 highest digestible samples of that mapping population. Extracted DNA was sent to the Purdue Genomics Core Facility for whole genome sequencing. Hundred base pairs paired-end Illumina HiSeq 2500 was used for SbEMS1613 bulks
while SbEMS3324 DNA bulks were sequenced through Illumina NovaSeq platform.

4.2.5 Analysis of conserved mutations between progeny with one shared parent

To verify results from BSA, DNA markers were designed using primer-BLAST (Ye et al., 2012) and ordered from Integrated DNA Technologies (IDT). They were designed based on the reference genome sequence region around the identified SNPs from BSA (see primer sequences in table 4.1) and were used to amplify the genes in the selected progeny of F2 mapping population and F3 recombinant populations.

After DNA extraction of each sample individually, PCR amplification was performed with the parameters described in Tables 4.2 and 4.3 following conditions from the high fidelity master mix manufacturer. Following PCR amplification, DNA was purified using a Qiagen gel purification kit. A total of 50 µL PCR product was loaded on a 1.5% agarose gel and electrophoresis was performed at 100 Volts for 1 hour. The DNA bands were cut from the gel and cleaned using a Qiagen gel purification kit according to the manufacturer’s instructions. Quality of purified DNA was checked using a ThermoScientific NanoDrop spectrophotometer. All DNA samples with an A260/A280 absorbance ratio between 1.8 and 2 were considered high quality and were sent to the Purdue Genomics Core Center for Sanger sequencing. The resulting DNA sequence was aligned for all samples using BioEdit Sequence Alignment Editor version 7.2.6.
Table 4.1 Primer information per mutant

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Primer name</th>
<th>Orientation</th>
<th>Primer sequence</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SbEMS1613</td>
<td>Pro-F</td>
<td>Forward</td>
<td>5’-TTAGGTGGCAGACGAAGCAG-3’</td>
<td>68.2</td>
</tr>
<tr>
<td></td>
<td>Pro-R</td>
<td>Reverse</td>
<td>5’-TCATCACCTGCTGGCAACCAT-3’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kaf-F</td>
<td>Forward</td>
<td>5’-AACGTCTTGGCAAATACTGGC-3’</td>
<td>65.8</td>
</tr>
<tr>
<td></td>
<td>Kaf-R</td>
<td>Reverse</td>
<td>5’-CTGAGCGCTGGTAGGATCTG-3’</td>
<td></td>
</tr>
</tbody>
</table>

*Tm* = hybridization temperature

Table 4.2: PCR reaction volumes per samples according to high fidelity master mix manufacturer’s conditions (items put in order).

<table>
<thead>
<tr>
<th>Component</th>
<th>50 µL final reaction volume</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2O</td>
<td>10 µL</td>
<td>-</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>2.5 µL</td>
<td>0.5 µM</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>2.5 µL</td>
<td>0.5 µM</td>
</tr>
<tr>
<td>Template DNA</td>
<td>10 µL</td>
<td>250 ng</td>
</tr>
<tr>
<td>2X Phusion Green HSII High-Fidelity Master Mix</td>
<td>25 µL</td>
<td>1 X</td>
</tr>
</tbody>
</table>

Table 4.3: PCR cycling steps and conditions according to the master mix manufacturer’s conditions.

<table>
<thead>
<tr>
<th>Cycle step</th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>98°C</td>
<td>30 s</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>98°C</td>
<td>10 s</td>
<td>40</td>
</tr>
<tr>
<td>Annealing</td>
<td>Tm</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>5 min</td>
<td>1</td>
</tr>
<tr>
<td>Hold</td>
<td>4°C</td>
<td>∞</td>
<td></td>
</tr>
</tbody>
</table>
4.2.6 Data analysis

R version 3.3 based packages xlsx, Rmisc, agricolae and Lme4 were used for the digestibility data analysis. Package xlsx was used to import and extract digestibility values from the excel template used for the protein digestibility assay data. Rmisc package was used to calculate means and standard deviations while Lme4 was used for the best linear unbiased estimators (BLUEs) of protein digestibility values used to rank the entries. Agricolae was used for Duncan mean separation test (p = 5%).

4.2.6.1 Phenotypic data

Raw absorbance values of the tested entries obtained from the digestibility assay were normalized against the absorbance of the TCA-buffer well. The percentage digestibility (PD) was calculated as follows using the adjusted absorbance values ($A_{560}$) of digested and undigested samples for each entry:

$$PD = \frac{((A_{560\text{ Undigested}} - A_{560\text{ Digested}})/A_{560\text{ Undigested}}) * 100}{1}$$

Analysis of variance for protein digestibility (PD) was performed as follows:

$$\text{lm (PD ~ Genotype + Row: Plate + Column: Plate + Plate)}$$

BLUEs were calculated from the linear mixed model with genotypes set as fixed, and other significant parameters from the ANOVA set as random.

4.2.6.2 Sequencing data

Whole genome sequencing reads contained in a FastQ file were mapped against version 3.1.1 of the sorghum reference genome (Paterson et al., 2009) which was obtained online from Phytozome web portal version 12.1.6 (Goodstein et al., 2012). The alignment was achieved through the
Burrows-Wheeler Aligner (BWA) software package designed for short reads alignment (Li and Durbin, 2009). SNPs were called using the command multiallelic-caller of SAMtools module version 1.4 (Li et al., 2009). Only homozygous SNPs with a minimum quality of 20 and a maximum depth of 250 were kept in Variant Call Format (VCF) files.

After obtaining the VCF files containing the filtered SNPs, all subsequent analyses were performed in R version 3.3 using the packages VariantAnnotation and Zoo. Allele frequencies were calculated and used to compute a rolling average difference between allele frequencies in the highly and lowly digestible groups for every 21 bp for SbEMS1613 and for every 11bp for SbEMS3324 mapping populations. The results were then plotted to show the BSA.

Sanger sequencing DNA sequence reads of the progenies were compared to the DNA sequences of the EMS parents using the sequence alignment editor BioEdit version 7.2.6 (Hall, 1999).
4.3 Results

4.3.1 Protein digestibility of sorghum

Sorghum was 50 to 60% less digestible than maize after wet cooking (Figure 4.1). Sorghum when wet cooked showed an average of 30% lower digestibility than uncooked. The results also showed that even the uncooked maize is more digestible than sorghum and there was more variability in sorghum samples (Figure 4.1).

![Bar chart showing protein digestibility of sorghum and maize](http://ugspace.ug.edu.gh)

**Figure 4.1**: Comparison between protein digestibility of uncooked and cooked maize (B73, M017, and W22) and sorghum (BTx623, SbEMS932, and P721N).

The standard error is shown by the error bars.

Mutants, as well as wild types, showed a significant decrease in digestibility after wet cooking (Figure 4.2). However, compared to their low digestible parent line BTx623, the EMS mutants were 25 to 40% more digestible after wet cooking. P721Q served as a positive control. Using the
modified digestibility assay, it was found that P721Q had a 20% increase in digestibility compared to its progenitor.

![Graph showing protein digestibility of different samples](attachment://graph.png)

Figure 4.2: Comparison between average (5 replicates) protein digestibility of P721N and its mutant P721Q; between BTx623 and its EMS mutants SbEMS1227, SbEMS1227, SbEMS3324. The standard errors are shown in black bars.

### 4.3.2 Protein body morphology of the EMS mutants

The highly digestible EMS mutants SbEMS1227 and SbEMS3324 have invaginated protein bodies just like P721Q (Figure 4.3). In contrast, the protein bodies of SbEMS1613 are round, like the wild type BTx623.
Figure 4.3: Transmission electron microscopy images showing the protein body structure of the non-mutagenized parent BTx623 (A) and the 3 highly digestible mutants SbEMS1613 (B), SbEMS1227 (C), and SbEMS3324 (D). Arrows point to invaginated PB.

4.3.3 Crude protein, lysine and tryptophan contents of sorghum mutants

The highly digestible sorghum mutants contained higher proportions of crude protein as well as higher lysine, and tryptophan contents than the wild type BTx623 (Table 4.4). Compared to the latter, SbEMS1613 contained 50% more lysine, 30% more protein, and 33% more tryptophan
while SbEMS3324 has double the lysine content of BTx623, 40% more protein, and 56% more tryptophan. SbEMS3324 has 12% more lysine and 9% more protein than the known high lysine, highly digestible sorghum mutant, P721Q.

Table 4.4: Crude protein, total lysine and tryptophan contents of wild type (BTx623) and sorghum mutants.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Crude protein (g/100 g of samples)</th>
<th>Total lysine (g/100 g of samples)</th>
<th>Tryptophan (g/100 g of samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTx623</td>
<td>10.66</td>
<td>0.24</td>
<td>0.09</td>
</tr>
<tr>
<td>SbEMS1613</td>
<td>13.38</td>
<td>0.36</td>
<td>0.12</td>
</tr>
<tr>
<td>SbEMS3324</td>
<td>14.84</td>
<td>0.48</td>
<td>0.14</td>
</tr>
<tr>
<td>P721Q</td>
<td>13.62</td>
<td>0.43</td>
<td>0.13</td>
</tr>
</tbody>
</table>

4.3.4 Seed diameter of EMS mutants

Significant differences were noted between the genotypes tested (Table 4.5). Duncan means separation test was performed to compare the means (P < 0.001). BTx623 had significantly larger seeds than P721N, P721Q as well as the EMS mutants (Table 4.6). The mutation seems to have induced a smaller seed size compared to the wild type parents BTx623 and P721N. However, seeds of the high digestible EMS mutants were significantly larger than those of the high lysine P721Q.
Table 4.5: Mean square for seed diameter in F3 families.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>4</td>
<td>0.07644 ***</td>
</tr>
<tr>
<td>Rep</td>
<td>4</td>
<td>3.28 e-3 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>20</td>
<td>0.00127</td>
</tr>
</tbody>
</table>

DF: degree of freedom; ‘***’ significant at $P < 0.001$.

Table 4.6: Seed diameter of sorghum mutants (SbEMS3324, SbEMS1613, P721Q) and their parent lines (BTx623 and P721N).

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Seed diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTx623</td>
<td>3.98 a</td>
</tr>
<tr>
<td>P721N</td>
<td>3.86 b</td>
</tr>
<tr>
<td>SbEMS3324</td>
<td>3.83 b</td>
</tr>
<tr>
<td>SbEMS1613</td>
<td>3.77 c</td>
</tr>
<tr>
<td>P721Q</td>
<td>3.64 d</td>
</tr>
</tbody>
</table>

LSD=0.08; Means with the same letter are not significantly different ($P < 0.001$).

4.3.5 Seed hardness of EMS mutants

Wild type sorghum lines, P721N and BTx623, had a similar level of seed hardness while all high digestible mutants had significantly softer seeds than their low digestible progenitors (Figure 4.4). However, SbEMS1613 had significantly harder seeds than P721Q and SbEMS3324. The latter two display similar seed hardness.
Figure 4.4: Seed hardness of mutants P721Q, SbEMS1613 and SbEMS3324 compared to their parent lines P721N and BTx623 respectively. The standard errors are shown as black bars.

4.3.6 Whole genome sequencing

For SbEMS1613, a total of 299,495 high-quality SNPs were identified in the highly digestible pool; however, after filtering, only 104,253 G:A and C:T EMS-induced changes were kept. From a total of 295,458 SNPs, only 102,663 EMS-induced SNPs were obtained in the low digestible pool for SbEMS1613. In SbEMS3324 mapping population bulks, after filtering, a total of 181,744 SNPs was identified for the highly digestible bulk and 175,899 SNPs for the lowly digestible pool of SbEMS3324.

4.3.7 Identification of causative alleles in EMS932 x EMS1613 mapping population

4.3.7.1 Variation in protein digestibility in SbEMS1613 mapping population

Significant differences (P < .001) were found for the following parameters: genotypes, plate ID,
and columns in the 96 well plate (Table 4.8). Rows were significantly not different between plates. Most of the variation was found between genotypes.

Table 4.7: Mean squares for protein digestibility in F₃ families.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>485</td>
<td>1232.09 ***</td>
</tr>
<tr>
<td>Plate</td>
<td>41</td>
<td>371.81 ***</td>
</tr>
<tr>
<td>Column: Plate</td>
<td>210</td>
<td>406.89 ***</td>
</tr>
<tr>
<td>Row: Plate</td>
<td>294</td>
<td>93.46 ns</td>
</tr>
<tr>
<td>Residuals</td>
<td>925</td>
<td>93.98</td>
</tr>
</tbody>
</table>

DF: degree of freedom; ‘***’ Significant at P < 0.001; ‘ns’ not significant.

Distribution of protein digestibility followed a bimodal distribution with a mean around 35% and 65% for each mode representing the low and the high bulks as two distinct populations and include the two parents of the mapping population (Figure 4.5). Furthermore, individuals with lower digestibility values than the low digestible parent as well as progenies with higher protein digestibility than the HD parent were observed in the mapping population.
Figure 4.5: Distribution of the BLUEs of protein digestibility.

The red arrows point the position of the low (SbEMS932) and high (SbEMS1613) controls corresponding to the female and the male parents of the mapping population.

4.3.7.2 Bulked segregant analysis in SbEMS1613 mapping population

The low and high digestible bulks used for bulked segregant analysis (BSA) were made of 50 entries each with mean protein digestibility values ranging from 60.12 to 86.8% for the high bulk (Table 4.9) and 11.96 to 30.08% for the low bulk (Table 4.10).
Table 4.8 Protein digestibility means per entries of high digestible bulk for BSA.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein digestibility (%)</th>
<th>SE</th>
<th>Entry</th>
<th>Protein digestibility (%)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>86.8</td>
<td>0.19</td>
<td>1</td>
<td>69.38</td>
<td>3.79</td>
</tr>
<tr>
<td>262</td>
<td>84.37</td>
<td>2.43</td>
<td>395</td>
<td>68.6</td>
<td>1.85</td>
</tr>
<tr>
<td>316</td>
<td>83.45</td>
<td>2.06</td>
<td>51</td>
<td>68.59</td>
<td>0.1</td>
</tr>
<tr>
<td>513</td>
<td>82.63</td>
<td>6.85</td>
<td>247</td>
<td>68.58</td>
<td>1.69</td>
</tr>
<tr>
<td>432</td>
<td>81.45</td>
<td>1.06</td>
<td>59</td>
<td>68.24</td>
<td>4.62</td>
</tr>
<tr>
<td>64</td>
<td>80.53</td>
<td>7.79</td>
<td>121</td>
<td>68.22</td>
<td>6.95</td>
</tr>
<tr>
<td>171</td>
<td>79.99</td>
<td>0.67</td>
<td>40</td>
<td>67.24</td>
<td>5.13</td>
</tr>
<tr>
<td>55</td>
<td>79.09</td>
<td>3.53</td>
<td>290</td>
<td>67.2</td>
<td>5.74</td>
</tr>
<tr>
<td>58</td>
<td>77.64</td>
<td>7.29</td>
<td>365</td>
<td>67.05</td>
<td>1.12</td>
</tr>
<tr>
<td>99</td>
<td>77.59</td>
<td>6.25</td>
<td>57</td>
<td>65.97</td>
<td>0.42</td>
</tr>
<tr>
<td>12</td>
<td>76.39</td>
<td>4.66</td>
<td>397</td>
<td>64.94</td>
<td>2.4</td>
</tr>
<tr>
<td>184</td>
<td>76.15</td>
<td>1.93</td>
<td>258</td>
<td>64.68</td>
<td>3.94</td>
</tr>
<tr>
<td>127</td>
<td>74.43</td>
<td>1.26</td>
<td>79</td>
<td>64.07</td>
<td>3.8</td>
</tr>
<tr>
<td>172</td>
<td>73.93</td>
<td>3.95</td>
<td>230</td>
<td>63.89</td>
<td>3.6</td>
</tr>
<tr>
<td>50</td>
<td>73.69</td>
<td>5.76</td>
<td>129</td>
<td>63.65</td>
<td>3.56</td>
</tr>
<tr>
<td>15</td>
<td>72.94</td>
<td>3.44</td>
<td>433</td>
<td>63.57</td>
<td>1.57</td>
</tr>
<tr>
<td>295</td>
<td>71.95</td>
<td>2.71</td>
<td>255</td>
<td>63.15</td>
<td>5.3</td>
</tr>
<tr>
<td>72</td>
<td>71.22</td>
<td>1.1</td>
<td>332</td>
<td>62.96</td>
<td>2.92</td>
</tr>
<tr>
<td>54</td>
<td>71.11</td>
<td>2.44</td>
<td>19</td>
<td>62.94</td>
<td>3.29</td>
</tr>
<tr>
<td>363</td>
<td>70.97</td>
<td>4.84</td>
<td>361</td>
<td>62.15</td>
<td>4.79</td>
</tr>
<tr>
<td>114</td>
<td>70.73</td>
<td>6.66</td>
<td>7</td>
<td>61.61</td>
<td>6.1</td>
</tr>
<tr>
<td>357</td>
<td>70.59</td>
<td>6.04</td>
<td>446</td>
<td>61.35</td>
<td>2.32</td>
</tr>
<tr>
<td>8</td>
<td>70.47</td>
<td>2.88</td>
<td>496</td>
<td>61.07</td>
<td>1.17</td>
</tr>
<tr>
<td>16</td>
<td>70.06</td>
<td>3.86</td>
<td>74</td>
<td>60.92</td>
<td>4.02</td>
</tr>
<tr>
<td>3</td>
<td>69.56</td>
<td>2.75</td>
<td>484</td>
<td>60.12</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Table 4. 9 Protein digestibility means per entries of low digestible bulk for BSA.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein digestibility (%)</th>
<th>SE</th>
<th>Entry</th>
<th>Protein digestibility (%)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>459</td>
<td>11.96</td>
<td>3.85</td>
<td>470</td>
<td>20.87</td>
<td>5.6</td>
</tr>
<tr>
<td>420</td>
<td>12.14</td>
<td>3.05</td>
<td>265</td>
<td>21.27</td>
<td>0.94</td>
</tr>
<tr>
<td>263</td>
<td>12.48</td>
<td>0.57</td>
<td>4</td>
<td>21.53</td>
<td>6.28</td>
</tr>
<tr>
<td>222</td>
<td>12.84</td>
<td>5.19</td>
<td>66</td>
<td>21.54</td>
<td>6.08</td>
</tr>
<tr>
<td>125</td>
<td>13.15</td>
<td>4.97</td>
<td>168</td>
<td>21.59</td>
<td>1.56</td>
</tr>
<tr>
<td>98</td>
<td>13.94</td>
<td>0.9</td>
<td>377</td>
<td>21.69</td>
<td>2.07</td>
</tr>
<tr>
<td>407</td>
<td>14.79</td>
<td>1.15</td>
<td>485</td>
<td>22.06</td>
<td>8.26</td>
</tr>
<tr>
<td>514</td>
<td>14.98</td>
<td>3</td>
<td>150</td>
<td>22.11</td>
<td>5.54</td>
</tr>
<tr>
<td>289</td>
<td>15.59</td>
<td>4.23</td>
<td>378</td>
<td>22.31</td>
<td>2.04</td>
</tr>
<tr>
<td>277</td>
<td>16.15</td>
<td>4.13</td>
<td>22</td>
<td>22.57</td>
<td>1.11</td>
</tr>
<tr>
<td>285</td>
<td>16.95</td>
<td>3.15</td>
<td>82</td>
<td>22.92</td>
<td>4.8</td>
</tr>
<tr>
<td>48</td>
<td>17.15</td>
<td>4.41</td>
<td>335</td>
<td>23.55</td>
<td>2.11</td>
</tr>
<tr>
<td>154</td>
<td>17.19</td>
<td>4.04</td>
<td>156</td>
<td>23.59</td>
<td>7.3</td>
</tr>
<tr>
<td>425</td>
<td>17.67</td>
<td>1.9</td>
<td>65</td>
<td>23.74</td>
<td>5.15</td>
</tr>
<tr>
<td>449</td>
<td>18.07</td>
<td>2.77</td>
<td>597</td>
<td>23.75</td>
<td>1.9</td>
</tr>
<tr>
<td>233</td>
<td>18.1</td>
<td>4.85</td>
<td>227</td>
<td>24.84</td>
<td>1.83</td>
</tr>
<tr>
<td>188</td>
<td>18.6</td>
<td>2.31</td>
<td>89</td>
<td>25.36</td>
<td>4.32</td>
</tr>
<tr>
<td>457</td>
<td>19.41</td>
<td>4.77</td>
<td>20</td>
<td>25.58</td>
<td>5.81</td>
</tr>
<tr>
<td>468</td>
<td>19.84</td>
<td>4.23</td>
<td>67</td>
<td>26.01</td>
<td>3.12</td>
</tr>
<tr>
<td>164</td>
<td>20.1</td>
<td>2.64</td>
<td>47</td>
<td>26.4</td>
<td>6.08</td>
</tr>
<tr>
<td>383</td>
<td>20.24</td>
<td>1.24</td>
<td>73</td>
<td>26.63</td>
<td>4.33</td>
</tr>
<tr>
<td>226</td>
<td>20.27</td>
<td>3.57</td>
<td>252</td>
<td>26.83</td>
<td>2.78</td>
</tr>
<tr>
<td>422</td>
<td>20.35</td>
<td>2.01</td>
<td>284</td>
<td>27.66</td>
<td>1.59</td>
</tr>
<tr>
<td>403</td>
<td>20.45</td>
<td>4.21</td>
<td>259</td>
<td>28.45</td>
<td>1.96</td>
</tr>
<tr>
<td>415</td>
<td>20.85</td>
<td>4.09</td>
<td>68</td>
<td>30.08</td>
<td>5.53</td>
</tr>
</tbody>
</table>

BSA results revealed a strong association with protein digestibility on chromosome 5. While the entire chromosome 5 seemed to be biased towards the high digestible allele(s), a strong peak between 39 Mb and 43 Mb and a minor peak around 11 Mb and 12 Mb were identified on chromosome 5 (Figure 4.6). There were no SNPs found within genes in the region under the major peak. However, a C:T change was found at position 1100 within the coding region of the 26S
proteasome complex regulatory subunit PSMD10 (Sobic.005G0833340) (source: Phytozome version 12.1.6).

C to T change in the DNA coding region induced a non-synonymous missense mutation of Histidine (H) to Tyrosine (Y) at position 234 in the signal peptide of the 26S proteasome regulatory complex (see below highlighted).

BTx623  LAAHKNQIEVLRLQLQHDPSLGYFISTDGSPLLCIAATEGHVGVERLLRHCDDPPYCDA
SbEMS1613  LAAHKNQIEVLRLQLQHDPSLGYFISTDGSPLLCIAATEGYVGVERLLRHCDDPPYCDA

Figure 4.6: Bulked segregant analysis in F2 segregants of SbEMS932 x SbEMS1613 mapping population.

4.3.7.3 SNP validation

Figure 4.7 shows that the SNP occurring on the 26S proteasome complex regulatory subunit PSMD10 gene was found in 16 F2 highly digestible genotypes of the mapping population and 20
F3 highly digestible recombinants having in common only the highly digestible parent, SbEMS1613. The results showed that all highly digestible samples, regardless of the cross, had the mutant allele while no presence of the mutation could be detected in the lowly digestible progenies (Figure 4.7).

Figure 4.7: Partial alignment of the 26S proteasome complex regulatory subunit PSMD10 gene sequence amplified from the genomic DNA of 14 low protein digestible (LD) and 36 high protein digestible (HD) sorghum samples. C:T change, highlighted in red, occurs at position 1100 of the genomic sequence of the 26S PMSD proteasome subunit gene of the HD EMS mutant as well as the high digestible samples from 5 different crosses but absent from the LD progenies.
4.3.8 Identification of causative alleles in SbEMS3324 x SbEMS932 mapping population

4.3.8.1 Distribution of protein digestibility in SbEMS3324 mapping population

Significance differences (P < 0.001) between genotypes, unique plate IDs, columns and rows in a 96 well plate were noted (Table 4.10). Most of the variation observed was within genotypes.

Table 4.10: Analysis of Variance for protein digestibility in F3 families of SbEMS3324 x SbEMS932.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>456</td>
<td>2412.67 ***</td>
</tr>
<tr>
<td>Plate</td>
<td>38</td>
<td>678.35 ***</td>
</tr>
<tr>
<td>Column: Plate</td>
<td>195</td>
<td>165.08 ***</td>
</tr>
<tr>
<td>Row: Plate</td>
<td>273</td>
<td>140.61 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>870</td>
<td>99.23</td>
</tr>
</tbody>
</table>

DF: degree of freedom; ‘***’ Significant at P < 0.001

Protein digestibility in the mapping population follows a bimodal distribution showing the presence of two distinct groups with means around 30 and 70% corresponding to the low and high digestible samples respectively. Each group comprises one parent from which the population was generated (Figure 4.8).

A total of 119 (26%) highly digestible and 336 (74%) lowly digestible F3 progenies were obtained from the digestibility assay of the mapping population which implies a ratio of 2.82:1 which conforms to the expected 3:1 ratio. From these results, a recessive gene controls the trait.
Figure 4.8: Distribution of the corrected protein digestibility values (BLUEs). The blue arrows point the position of the low (SbEMS932) and high (SbEMS3324) controls corresponding to the female and the male parents of the mapping population respectively.

4.3.8.2 Bulked segregant analysis in SbEMS3324 mapping population

The low and high digestible bulks used for bulked segregant analysis (BSA) were made of 30 samples each with mean protein digestibility ranging from 62.7 to 91.87% with a mean of 80.84% for the high bulk and 11.39 to 30.84% for the low bulk with a mean of 16.64% (Table 4.11).
Table 4.11: Protein digestibility means per entry of high and low bulks chosen for BSA

<table>
<thead>
<tr>
<th>Entry</th>
<th>High digestible bulk</th>
<th></th>
<th>Low digestible bulk</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entry</td>
<td>Protein digestibility (%)</td>
<td>SE</td>
<td>Entry</td>
</tr>
<tr>
<td>81</td>
<td>91.87</td>
<td>1.08</td>
<td>71</td>
<td>11.39</td>
</tr>
<tr>
<td>181</td>
<td>91.23</td>
<td>1.94</td>
<td>184</td>
<td>12.55</td>
</tr>
<tr>
<td>157</td>
<td>89.23</td>
<td>1.73</td>
<td>182</td>
<td>12.79</td>
</tr>
<tr>
<td>138</td>
<td>85.87</td>
<td>0.98</td>
<td>160</td>
<td>13.13</td>
</tr>
<tr>
<td>140</td>
<td>84.95</td>
<td>3.75</td>
<td>23</td>
<td>13.26</td>
</tr>
<tr>
<td>70</td>
<td>84.92</td>
<td>3.05</td>
<td>69</td>
<td>13.42</td>
</tr>
<tr>
<td>178</td>
<td>84.82</td>
<td>4.01</td>
<td>95</td>
<td>13.63</td>
</tr>
<tr>
<td>156</td>
<td>84.68</td>
<td>3.27</td>
<td>6</td>
<td>13.7</td>
</tr>
<tr>
<td>162</td>
<td>84.36</td>
<td>4.89</td>
<td>148</td>
<td>13.8</td>
</tr>
<tr>
<td>15</td>
<td>84.09</td>
<td>2.06</td>
<td>170</td>
<td>14.25</td>
</tr>
<tr>
<td>22</td>
<td>83.79</td>
<td>2.68</td>
<td>72</td>
<td>14.27</td>
</tr>
<tr>
<td>8</td>
<td>82.73</td>
<td>3.00</td>
<td>68</td>
<td>15.16</td>
</tr>
<tr>
<td>28</td>
<td>82.54</td>
<td>0.32</td>
<td>168</td>
<td>15.21</td>
</tr>
<tr>
<td>74</td>
<td>82.45</td>
<td>4.56</td>
<td>124</td>
<td>15.24</td>
</tr>
<tr>
<td>118</td>
<td>82.38</td>
<td>2.28</td>
<td>30</td>
<td>15.52</td>
</tr>
<tr>
<td>35</td>
<td>81.29</td>
<td>3.51</td>
<td>33</td>
<td>15.99</td>
</tr>
<tr>
<td>167</td>
<td>81.2</td>
<td>3.98</td>
<td>169</td>
<td>16.1</td>
</tr>
<tr>
<td>56</td>
<td>80.98</td>
<td>1.34</td>
<td>125</td>
<td>16.15</td>
</tr>
<tr>
<td>63</td>
<td>80.93</td>
<td>5.41</td>
<td>122</td>
<td>16.44</td>
</tr>
<tr>
<td>58</td>
<td>79.36</td>
<td>4.08</td>
<td>180</td>
<td>17.05</td>
</tr>
<tr>
<td>60</td>
<td>79.13</td>
<td>4.07</td>
<td>188</td>
<td>18.17</td>
</tr>
<tr>
<td>87</td>
<td>78.86</td>
<td>3.32</td>
<td>128</td>
<td>18.24</td>
</tr>
<tr>
<td>105</td>
<td>78.28</td>
<td>2.96</td>
<td>141</td>
<td>18.89</td>
</tr>
<tr>
<td>119</td>
<td>77.52</td>
<td>3.34</td>
<td>192</td>
<td>18.98</td>
</tr>
<tr>
<td>102</td>
<td>76.57</td>
<td>2.86</td>
<td>121</td>
<td>19.11</td>
</tr>
<tr>
<td>154</td>
<td>75.69</td>
<td>6.07</td>
<td>164</td>
<td>19.2</td>
</tr>
<tr>
<td>21</td>
<td>75.03</td>
<td>2.82</td>
<td>16</td>
<td>19.31</td>
</tr>
<tr>
<td>18</td>
<td>74.45</td>
<td>3.57</td>
<td>41</td>
<td>19.36</td>
</tr>
<tr>
<td>53</td>
<td>63.35</td>
<td>1.11</td>
<td>129</td>
<td>28.01</td>
</tr>
<tr>
<td>172</td>
<td>62.7</td>
<td>1.68</td>
<td>54</td>
<td>30.84</td>
</tr>
</tbody>
</table>

BSA results showed that the entire chromosome 5 was biased towards the allele in the high pool.

One major peak was found on chromosome 5 around 67 Mb showing a strong association between that region of the chromosome and protein digestibility trait (Figure 4.9). Only a Kafirin PSKR2 precursor-like gene (Sobic.005G189000) is located under the peak bearing a G:A change in its
coding region at position 61 (source: Phytozome version 12.1.6).

The mutation on the coding region of the kafirin gene led to a missense mutation that induced a change from alanine (Ala) being substituted with threonine (Thr) at position 21 in the signal peptide.

Figure 4.9: Bulked segregant analysis in F3 segregants of SbEMS932 x SbEMS3324 mapping population.
4.3.8.3 SNP validation

Following PCR amplification, alignment of the DNA sequence generated from Sanger sequencing results revealed the presence of the mutant SNP (A) in all 35 high protein digestible entries (17 F3 from mapping population and 18 F3 recombinants) while all 15 low protein digestible samples showed only the wild type SNP (G) (Figure 4.10).

Figure 4.10: Partial alignment of the Kafirin PSKR2 precursor-like protein amplified from the genomic DNA of 15 low and 35 high protein digestible (HD) progenies. G:A change, highlighted in red, occurs at position 61 of the genomic sequence of the Kafirin PSKR2 precursor-like gene of the HD SbEMS3324 as well as the HD samples from five different crosses (1 mapping population and 4 additional) but absent from the low protein digestible progenies.
4.4 Discussion

Sorghum EMS mutants had significantly higher protein, lysine and tryptophan than their wild type parent BTx623. Similar results were reported by Singh and Axtell (1973) where natural and induced high lysine sorghum showed higher nutrient content than normal sorghum lines. However, SbEMS3324 had more lysine and protein than the high lysine, highly digestible sorghum mutant, P721Q. The increase in total lysine in mutants could be a result of an increase in free-lysine or in lysine-rich proteins or even a reduction in prolamins (lysin-poor proteins) along with an increase in albumins, globulins and glutelins (Singh and Axtell, 1973; Wu et al, 2013). Lysine and tryptophan are essential amino acids meaning that they cannot be synthesized by non-ruminant organism hence must be supplied in the diet. Tryptophan is the second limiting amino acids, after lysine, in most cereals proteins. In sorghum and/or maize-based diets, a disease called Pellagra in humans could prevail due to low tryptophan content especially when only very few legumes are consumed (Singh and Axtell, 1973). Therefore, combining the high digestible trait with high lysine in sorghum lines would be very beneficial.

Because in the high lysine and highly digestible mutant, P721Q, generated about 4 decades ago (Mohan, 1975), the increase in lysine and amount of digestible protein was accompanied with a floury endosperm (Wu et al., 2013), soft and smaller seed size, one might expect the new mutants to have the same characteristics. Therefore, these important traits were investigated in the newly developed and identified highly digestible mutants. The results revealed that the EMS mutants had larger seeds than P721Q. In addition, SbEMS1613 showed harder seeds than P721Q. However, compared to their wild type progenitors, a significant decrease in seed hardness and seed size was seen in all mutants even though they had up to 26 and 37% more digestible proteins in SbEMS1613 and SbEMS3324 respectively. The decrease in seed size and hardness could be due to allelic
interactions or pleiotropic effects of the genes controlling these traits and those controlling the increase in digestibility. Seed hardness is a very important trait especially in areas where birds’ attack prevails. In fact, varieties with soft seeds are more susceptible to damage from birds because the softer the seeds, the easier it is to break through the pericarp. All mutants appear to have a floury endosperm (data not shown) but the advantage with the EMS mutants is the harder seeds they display which would be less susceptible to grain mold and to bird attack and more suitable for hand or industrial processing of the whole grain. In fact, during the process of removing the pericarp (dehulling) by pounding before cooking, if the seeds are too soft, they will break and produce powder in the mortar. It was therefore beneficial to have harder kernels to overcome these limitations.

As previously reported (McLean Jr et al., 1981) sorghum is less digestible than other cereal grains when wet cooked. This study reported 50 to 60% decrease in digestible proteins of sorghum compared to maize which was reported to be influenced by many factors such as protein crosslinking, endosperm structure, presence of antinutrients (Duodu et al., 2003) among others. For more than 40 years since the creation of the first highly digestible sorghum mutant (Mohan, 1975), the focus had been on trying to understand the factors affecting the decrease in digestibility of sorghum after wet cooking but also what are the genetic mechanisms controlling the increase in digestible protein in the mutants compared to other sorghum lines (Mertz et al., 1984; Oria et al., 2000; Winn et al., 2009; Wu et al., 2013). In P721Q, the increase in protein digestibility was hypothesized to be associated with the invaginated protein bodies phenotype (Oria et al., 2000). However, in this study, among the three highly digestible EMS mutants from BTx623, only SbEMS1227 and SbEMS3324 displayed the invaginated protein body phenotype while SbEMS1613 showed a phenotype close to the wild type i.e. round proteins. Therefore, even if the
invaginated protein body phenotype is linked to the increase in digestibility, it might not be the only factor that is responsible for high protein digestibility. Sorghum protein bodies are in the endosperm and contain the kafirins, which are alcohol soluble prolamins and the main storage of protein in the seed.

Distribution of the best linear unbiased estimators (BLUEs) of protein digestibility in SbEMS1613 and SbEMS3324 followed a bimodal distribution showing the presence of two populations with means at 30 and 70% respectively. The contrasting phenotype observed reflects the presence of a low and a high digestible population comprising the low and high digestible parents respectively. The difference between the two populations could be mainly due to the presence of the mutant allele(s) increasing protein digestibility in the high digestible population. In addition, a chi-square test revealed a ratio close to 3:1 (2:82) which would imply that the trait is controlled by one recessive allele in each EMS mutant.

The increase in digestible proteins in SbEMS1613 after wet cooking was mapped to the 26S proteasome complex regulatory subunit PSMD10 (Sobic.005G083340) present on chromosome 5. The mutation in the 26S proteasome regulatory complex (26SP) was present with an allele frequency of 1 or 100% in the high bulk and 0.16 in the low bulk. The presence of the alternative allele in the low bulk could be due to the selection of heterozygous genotypes as low digestible. In fact, the chi-square test showed a ratio close to 3:1 which implies that the trait is controlled by a recessive gene. Therefore, all high digestible samples would be homozygous recessive while the low digestible could be either homozygous dominant or heterozygous. The presence of the mutant allele in all highly digestible progeny from four different crosses, which shared the mutant as a parent, in addition to the samples from the mapping population, strongly suggests the causative effect of the allele and that the increase in digestibility is linked to the mutation in the proteasome.
So far, this mutation has never been reported to be linked to an increase in digestibility of proteins. Proteasomes are complexes that, together with ubiquitin, are involved in degrading non-functional, misfolded or unwanted proteins by the breakdown of their peptide bonds which result in smaller peptides (Bard et al., 2018). The 26S proteasome is the most important protease present in the nucleus and cytoplasm of eukaryotic cells. Through a ubiquitination process, proteins are marked for degradation by the 26SP (Bard et al., 2018). The 26SP is therefore known to regulate the presence, abundance and stability of many proteins implicated in different metabolic processes and biosynthesis pathways (Bard et al., 2018). A mutation in such a complex could lead to a change or loss of function. The potential change in proteasome function could also relate to the differences observed in protein body morphology in this study. In fact, SbEMS1613 has round protein bodies almost like its wild type BTx623 counterpart. It is hypothesized that wet cooking induces protein cross-linking and disulphide bonds in round protein bodies genotypes (Duodu et al., 2002). Therefore, a mutation in a gene encoding a protein involved in the breakdown of misfolded proteins could result in a thermostable mutant form of the enzymes due to the presence of the mutation. The hypothesis was that proteasomes, instead of only degrading the misfolded proteins, could also break down other peptide bonds or even disulphide bonds that could have been formed during the process of wet cooking sorghum flour. By breaking down the bonds, peptides with smaller amino acid chains would be obtained and exposed to proteolytic enzymes hence making them more available to the organism. It is important to further investigate this by studying the effect of the mutation in the highly digestible EMS mutant for a better understanding of the mechanism.

Mapping results of SbEMS3324 revealed a G to A SNP on the kafirin PSKR2 precursor-like gene (Sobic.005G189000) located on chromosome 5. The mutation was found with an allele frequency
of 1 in the high bulk while none of the lowly digestible F3 genotypes from SbEMS3324 displayed the mutation which could mean that they were all homozygous for the wild type allele since it was an advanced generation (F3) compared to F2 in the first mapping population studied. The G:A mutation led to the replacement of the Alanine at the 21st position by Threonine in the signal peptide. In maize, the Alanine at the 21st position is known to play a key role in the cleavage of peptide bonds (Kim et al., 2004) and was shown to play a similar role in sorghum kafirins inside the protein bodies (Wu et al., 2013). Therefore, the replacement of the amino acid Alanine with Threonine in the α-kafirin would probably result in the peptide bond not being cleaved properly, hence induce the invaginated protein body phenotype (Lending and Larkins, 1992). The later will allow the proteolytic enzymes to easily access and digest the α-kafirins which are the most abundant (80%) of all kafirins. A similar phenotype was reported in maize proteins (Lending and Larkins, 1992). The irregular shape of the protein bodies in maize mutant, floury2, was reported to be due to the substitution of Alanine at position 21 by Valine in the signal peptide of the α-zein (Lending and Larkins, 1992). Results of this study corroborate with previous findings by Wu et al. (2013) in which the invaginated protein bodies, hypothesized to be linked to the increase in digestibility in P721Q, were due to a mutation in a food storage kafirin gene but the exact copy of kafirin gene remained unknown. In fact, all 20 kafirin genes are located on chromosome 5 and are highly similar to one another which makes it difficult to differentiate them using the methodology by Likpa et al. (2012). Nevertheless, using bulked segregant analysis, the exact copy of kafirin (PSKR2, Sobic.005G189000) mutated in SbEMS3324 was identified in this study. In addition, sequence alignment revealed that it is the same mutation occurring in P721Q. These findings support and augment the results reported by Wu et al. (2013) by showing the recessive nature of the allele controlling high protein digestibility and provide clear and precise
information about the mutated gene controlling high protein digestibility in sorghum after wet cooking.

4.5 Conclusion

This study showed that the sorghum EMS mutants have more digestible proteins after wet cooking as well as higher lysine, tryptophan and crude protein content than their wild type parent. In addition, the increased protein digestibility was controlled by a single recessive allele and linked to a point mutation on a kafirin gene (Sobic.005G189000) in the mutant with invaginated protein bodies in SbEMS3324. The results of this study complement the work previously reported by allowing the identification of the exact copy of kafirin genes mutated that cause the high protein digestibility phenotype. Nonetheless, it was discovered that not only invaginated protein bodies led to high protein digestibility since one of the mutants in this study, SbEMS1613, had round protein bodies like the wild type. Consequently, the high digestibility phenotype in the later mutant was linked to a point mutation on a 26S proteasome PSMD10 subunit complex (Sobic.005G083340) located on chromosome 5. Two pairs of SNP markers: ProF and ProR, KafF and KafR were developed and confirmed to be tightly linked to the gene controlling the digestibility phenotype in each of the two mutants. These primers will play a very important role in accelerating the introgression of the high digestibility trait into sorghum cultivars through marker-assisted selection hence providing the market with sorghum with higher nutritional value.
CHAPTER FIVE

5.0 Development and Evaluation of a Highly Digestible Sorghum Population in Senegal

5.1 Introduction

In Senegal, sorghum is grown in a wide range of environments from the dryland of Eastern Senegal during the rainy season, to the flooded areas of Senegal river valley in the North. It is also cultivated in the Casamance region, and in the central zone where groundnut predominates also known as the groundnut basin. Local or traditional varieties are still cultivated and preferred by farmers in some areas of Senegal because of low adaptation of improved varieties. In that perspective, significant efforts have been made by research institutes to provide high yielding varieties with better grain quality. In fact, the Senegalese Institute of Agricultural Research (ISRA) released white grained, tannin-free varieties yielding between 2 to 3.5 T/ha. These varieties, namely Faourou, Nganda, Darou, and Nguinthe, are adapted to a wide range of agroecological regions of Senegal and well appreciated by farmers because of the relatively short to medium cycle (90 to 110 days), and grain quality (white with no tannin, vitreous endosperm). These varieties are rich in carbohydrates and their protein content ranges between 11.78 and 13.62% (ISRA, 2012) however, the proteins are poorly digested after wet cooking. Therefore, improving the digestibility of the proteins will enhance the nutritive value of these Senegalese varieties and could further become a good source of high protein digestibility in breeding programmes for other ecological zones in Africa.

In Senegal, wheat is imported every year since the early 1960s and is being used in industries to make flour for bread and other baking products. Senegal wheat importation went from $68,241,000 in 2005 to $133,166,000 in 2016 with a peak of $191,129,000 in 2013 (FAO 2016). With the aim of reducing this excessive cost, the Food Technology Institute (ITA) is researching ways to reduce
the quantity of wheat flour in the bread by substituting with more affordable and accessible cereal flours. So far, ITA successfully substituted wheat with up to 50% sorghum flour (unpublished) in bread. The challenge remains with the lack of gluten in sorghum which is an essential element in bakery. With collaboration with ITA that also focuses on using sorghum for infant feeding, incorporating the high lysine, high protein digestibility trait into locally-adapted sorghum varieties will provide a higher nutritional value product for infant nutrition.

The main objective of this study was to integrate food quality traits into locally adapted sorghum varieties, thus improving the availability of high-quality grains for new market development. Specifically, the study aimed to:

1. improve the digestibility of proteins of Faourou, a Senegalese white grained sorghum variety, and

2. identify promising BC$_3$F$_3$ families combining high protein digestibility and desirable agronomic traits.
5.2 Material and methods

5.2.1 Plant material

In this study, the third generation of a backcross (BC$_3$F$_3$) population was used. It was generated from a cross between a highly digestible mutant, P721Q, and a lowly digestible Senegalese inbred line, Faourou (ISRA-621B). P721Q has erect leaves while Faourou has a more open canopy (Figure 5.1).

5.2.1.1 Description of the parents used

Faourou is a tannin-free white grained inbred line developed and released in 2011 by the Senegalese Institute of Agricultural Research (ISRA) from a cross between Sorvato-1 and CE151-262. Faourou is rich in carbohydrates (72.83%), in phosphorous (369 mg/100mg), and has 2.26% of cellulose and performs well in environments with 600 to 800 mm of rainfall (ISRA, 2012). P721Q is a high lysine - highly digestible sorghum mutant developed at Purdue in 1975 (Mohan, 1975). It has white grains and contains anthocyanin on the stem and grains. The presence of anthocyanin is a monogenic trait displaying a dominant phenotype (Reddy et al., 2008) in P721Q hence was used in this study to check successful crosses at early stages of plant development.

Table 5.1: Comparison between the donor (P721Q) and the recurrent (Faourou) parents.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Plant type</th>
<th>Plant Height (cm)</th>
<th>Days to Maturity (j)</th>
<th>Grain yield (t/ha)</th>
<th>1000 grain weight (g)</th>
<th>Protein content (N*5.7) %</th>
<th>Protein digestibility (%)</th>
<th>Lysine content (g/100g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faourou</td>
<td>Tan</td>
<td>175</td>
<td>105</td>
<td>2.5 - 3</td>
<td>22</td>
<td>12.22</td>
<td>39</td>
<td>-</td>
</tr>
<tr>
<td>P721Q</td>
<td>Anthocyanin</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>6.21</td>
<td>70</td>
<td>0.24</td>
</tr>
</tbody>
</table>
5.2.1.2 Population development

The low digestible Senegalese variety, Faourou was used as a female recurrent parent and the high digestible mutant, P721Q as the donor parent in a backcrossing scheme (Figure 5.3). Since Faourou is a fertile inbred line, plastic emasculation was used to kill viable pollen and ensure production of true F1 hybrids. To achieve this, panicles of Faourou were covered with plastic bags, for at least 48 hours, just before anthesis. The heat produced by covering with plastic bags triggered pollen release and caused its abortion. With non-viable pollen, the plant was successfully emasculated. During the same period, selfing bags were used to cover P721Q panicles, just before anthesis, to collect pollen that was used to fertilize emasculated Faourou panicles. The crosses were performed
at the National Centre for Agronomic Research (CNRA) at Bambey, Senegal, during 2015 rainy season. The crossing blocks were laid as follows: two rows of the recurrent parent (Faourou) followed by two rows of the donor parent (P721Q) repeated with three different sowing dates to ensure matching flowering times.

The resulting seeds were planted during the cold off-season of 2015 (from mid-November). Successful F\textsubscript{1} plants were identified as plants with anthocyanin on the stem (Figure 5.2). The presence of anthocyanin is controlled by genes with dominant alleles and can be used as a phenotypic marker to identify successful hybrids in a cross between a tan plant and a plant with anthocyanin provided the tan plant is the female. All successful F\textsubscript{1}s carried the dominant allele and displayed the anthocyanin phenotype present in the donor parent, P721Q (Figure 5.2). Based on the latter information, all tan plants were discarded because they resulted from self-pollination of the recurrent female parent Faourou which is tan. In addition, genotyping was performed using a pair of SSR markers to confirm the presence of the mutant allele (see section quality control).
Seeds of five identified F1s were bulked and the plants crossed to Faourou as female during the cold off-season of 2015-2016. The resulting BC₁F₁ seeds were planted during the hot off-season of 2016 and leaf samples were collected for PCR genotyping. Successful BC₁F₁ plants were identified carrying the mutant allele from P721Q using a pair of simple sequence repeats (SSR) molecular marker linked to the mutant phenotype (Wu et al., 2013) (see section quality control). From the BC₁F₁ generation, the presence of anthocyanin only could not be used to select true crosses since it was not clear that the allele controlling high protein digestibility (eg. “a”) in the donor parent was linked to the one coding for anthocyanin (eg. “B”). Since the F1s were “Aa” for protein digestibility and “Bb” for anthocyanin, due to recombination, the “a” allele could have
combined with a genotype “Bb” with the “B” allele coming from Faourou; hence there could be BC₁F₁s with the high digestible allele “a” but no anthocyanin. Therefore, no plant was discarded based on phenotypic marker.

Due to the absence of background markers, two additional generations of backcrosses to the recurrent parent, Faourou, were done to ensure an average of 93.75% recovery of the recurrent parent genome in the progeny. The resulting BC₃F₁ seeds obtained during the rainy season of 2016 were self-fertilized twice to obtain the BC₃F₃ seeds used in this study. Since the digestibility trait is controlled by a recessive allele, and no phenotyping for protein digestibility was possible before an F₃ generation due to the destructive nature of the screening, a total of 128 BC₃F₃ families were selected to expect at least 32 highly digestible families (25% of the total). The 128 families, as well as the donor and recurrent parents, were evaluated in Bambey during 2017 rainy season.
5.2.1.3 Quality control using foreground markers

Wu et al. (2013) showed that the mutation in P721Q is linked to a point, missense mutation on one copy of the 20 kafirin genes located on chromosome 5 of the sorghum genome. A pair of simple sequence repeats (SSR) primers was developed and shown to be linked to the phenotype observed in the donor parent, P721Q (Wu et al., 2013). The SSR primers (see below) were used to check the presence of the highly digestible mutant allele in the progeny.
SSR-F: 5’-AGTCAACAACTCCCTCCACC-3’

SSR-R: 5’-ATCGGCTGGTCGTCGACTGAG-3’

In the F1 generation, the presence of anthocyanin in the seedlings, as well as the presence of the mutant allele, were used as an indicator to select the successful crosses. From BC1F1 to BC3F1, only the pair of SSR primers was used to select successful backcrosses. No protein digestibility phenotyping was performed at early stages because the phenotyping method is destructive hence would need at least an F3 generation’s seeds which can be treated as families.

5.2.1.4 DNA extraction and quantification

DNA extraction was carried out based on the methodology described by (Risterucci et al., 2000). Approximately 20 mg of dried leaves were placed into 2 mL Eppendorf tubes with 3 steel beads of 5 mm each. The leaves were ground for 3 minutes at 30 beats per second using a RETSCH grinder. To release the DNA, 750 μL of Mixed Alkyl Trimethyl Ammonium Bromide (MATAB) buffer (pre-heated at 65 °C) was added to each ground sample to allow lysis of the cell membrane. The resulting mixture was vortexed, and the tubes were incubated for 20 min in a water bath at 65 °C with series of shaking every 5 min to facilitate the release of the DNA into the buffer solution. The samples were then cooled to room temperature for 5 minutes. To separate the DNA from the residues, 750 μL of chloroform-isoamyl alcohol (CIAA, 24: 1) was added to the mixture which was then homogenized by inversion followed by centrifugation for 20 min at 13000 rpm. A volume of 600 μL of supernatant was transferred to a 1.5 μL Eppendorf tube and an equal volume of cold isopropanol (-20 °C) was added. The mixture was gently stirred until the precipitated DNA pellet appeared and the tubes placed at -20 °C for 2h. After incubation, the samples were centrifuged at 13000 rpm for 20 min at 4 °C and the supernatant discarded. The pellet was then
washed in 500 μL of 70% ethanol. After centrifugation at 13000 rpm at 4 °C for 20 min, the supernatant was discarded. The tubes were dried under a ventilated hood for 30 min to 1 hour. The pellet was resuspended in 1 X TE overnight at room temperature.

Estimation of the quantity of extracted DNA was carried out through electrophoresis migration at 100 volts for 30 min using 1% agarose gel (Lee et al., 2012). The size and intensity of the DNA bands were compared to the bands of a StepLadder of known concentration and molecular weight. The electrophoresis migration was performed at 100 volts at room temperature for 30 minutes using an EPS 300 generator (Pharmacia Biotech). DNA was diluted to a final concentration of 5 ng/μL before amplification.

### 5.2.1.5 PCR amplification and sequencing

A two-step polymerase chain reaction (PCR) amplification was performed with the following parameters (Table 5.3) and components according to the Taq DNA polymerase manufacturer (Table 5.2).

Table 5.2: PCR reaction components and concentration per sample.

<table>
<thead>
<tr>
<th>Component</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer</td>
<td>1 X</td>
</tr>
<tr>
<td>dNTPs</td>
<td>0.5 μM</td>
</tr>
<tr>
<td>Taq DNA polymerase</td>
<td>0.5 μM</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>0.5 μM</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>0.5 μM</td>
</tr>
<tr>
<td>Template DNA</td>
<td>250 ng</td>
</tr>
<tr>
<td>H2O</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 5.3: PCR cycling steps and conditions.

<table>
<thead>
<tr>
<th>Cycle step</th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94 °C</td>
<td>4 min</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94 °C</td>
<td>45 s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>58 °C</td>
<td>60 s</td>
<td>35</td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>75 s</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72 °C</td>
<td>5 min</td>
<td>1</td>
</tr>
<tr>
<td>Hold</td>
<td>4°C</td>
<td>∞</td>
<td></td>
</tr>
</tbody>
</table>

The amplified DNA fragments were separated on a LI-COR (Lincoln, Nebr.) model 4300 DNA sequencer with fluorescent dye 700 (Yomano and Scopes, 1993). To each 3 μL of amplified DNA, 8 μL of loading dye (6 X blue urea) were added followed by one cycle of denaturation at 94 °C for 3 min in a Thermocycler 18 (MWG AG Biotech Primus 96). After denaturation, 2 μL were migrated on a 6.5% polyacrylamide gel until the appearance of the fluorescent bands detected using an infrared camera.

5.2.2 Description of the experimental site

The BC₃F₃ families were evaluated at the National Centre for Agronomic Research (CNRA) at Bambey. It is in the Northern part of the peanut basin of Senegal (14°42’N; 16°28’W). Bambey is characterized by a Sudano-Sahelian climate with sandy-clay soil (86-90% sand, 2-5% silt loam, and 6-10% clay) (Tovignan et al., 2016). The typical growing season in Bambey is from July to October however, a long period of drought was observed in August 2017. Rainfall and weather conditions at Bambey during 2017 growing season are shown in table 5.4.
Table 5. Precipitation, temperature and relative humidity at Bambe research station during period of trials

<table>
<thead>
<tr>
<th>Precipitation</th>
<th>Temperature (°C)</th>
<th>Relative Humidity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mm</td>
<td>Days</td>
<td>Min.</td>
</tr>
<tr>
<td>425.7</td>
<td>36</td>
<td>24.6</td>
</tr>
</tbody>
</table>

_mm=_millimetre; Min= minimum; Max= maximum._

### 5.2.3 Field experimental design and management

One hundred and twenty-eight BC$_3$F$_3$ families along with the two parents (Faourou and P721Q) were evaluated for agronomic performance in Bambe during the rainy season of 2017. A 10 by 13 alpha lattice design was used as the field layout. Thirteen entries were randomly assigned to each of the 10 blocks. Three replications were done however, one replication was removed from the analysis due to missing data. The experimental unit was a two-row plot 3.6 m long with a spacing of 0.4 m between hills in the same row and 0.8 m between rows. Approximately 4 to 6 seeds were sown per hill and three weeks after planting the seedlings were thinned to three plants. Field management followed the standard guidelines recommended by ISRA for optimum sorghum production. A dose of 150 kg/ha of compound fertilizer (NPK 15-15-15) was applied to the field before planting. An insecticide (carbofuran) was used to treat the seeds just before planting. Two equal doses of 50 kg/ha of urea were applied after plant thinning and at boot stage. Manual and mechanical weeding were done when necessary to keep the experimental field weed-free.
5.2.4 Data collection

The BC$_3$F$_3$ families and their two parents were subjected to detailed phenotypic characterization using the descriptor for sorghum (IBPGR and ICRISAT, 1993). Six quantitative traits were measured:

- time to 50% flowering (DSFLO): from emergence to when 50% of the plants started flowering;
- plant height (HPL) of the main stem measured at 50% flowering from the base of the plant to the panicle end;
- peduncle length (LPED): from the attachment to the main stalk to the starting point of seeds on the panicle;
- panicle length (LPAN) measured longitudinally at panicle maturity excluding the peduncle;
- panicle width (WPAN) transversally measured in the middle of the panicle;
- protein digestibility (PDigest): measured using the digestibility assay described in chapter three. Matured dried seeds obtained after self- fertilizing individual panicles of BC3F3 families were tested.

Some families did not produce enough seeds to measure 1000 grain weight. Therefore, no yield data were recorded due to more than 10% missing data.
5.2.5 Data analysis

- Analysis of variance

Analysis of variance (ANOVA) was performed on all progenies using “aov” command of R software version 3.5 (R studio version 1.1.44). The following linear model was used for the ANOVA.

\[ Y_{ijk} = \mu + R_i + B(R)_{ij} + G_k + \varepsilon_{ijk} \]

Where \( Y_{ijk} \) = measured trait, \( \mu \) = overall population mean, \( R_i \) = effect of replicate \( i \), \( B(R)_{ij} \) = effect of block \( j \) within replicate \( i \), \( G_k \) = effect of genotype \( k \) (BC3F3 families), \( \varepsilon_{ijk} \) = experimental error.

Based on the protein digestibility results, the BC3F3 families were divided into two groups: a low digestible group with digestibility values < 60%, and a high digestible group with protein digestibility values \( \geq 60\% \). ANOVA was performed to compare the low and high protein digestible groups using the following model:

\[ Y_{ijk} = \mu + R_i + B(R)_{ij} + G_k + \varepsilon_{ijk} \]

Where \( Y_{ijk} \) = measured trait, \( \mu \) = overall population mean, \( R_i \) = effect of replicate \( i \), \( B(R)_{ij} \) = effect of block \( j \) within replicate \( i \), \( G_k \) = effect of groups \( k \), \( \varepsilon_{ijk} \) = experimental error.

- Correlation test

Pearson correlation coefficients between protein digestibility and agronomic traits were calculated using the R-based package PerformanceAnalytics.
• **Broad-sense heritability**

Broad-sense heritability was calculated using the following formula for one year, one location. The variance components were calculated using the function VarCorr from the lme4 package in R.

\[ H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 / r} \]

Where \( H^2 \) is the broad sense heritability, \( \sigma_g^2 \) is the genotypic variance, \( \sigma_e^2 \) is the error variance, and \( r \) is the number of replications.

• **Classification of BC3F3 progenies**

Lambda Wilk test was performed using XLSTAT (trial version) to check the best discriminant trait to classify the entries.

Agglomerative hierarchical clustering followed by a discriminant analysis were performed using the software XLSTAT (trial version) to classify the genotypes based on their similarities for the traits measured.
5.3 Results

5.3.1 Variation for measured traits

A highly significant difference was observed between BC3F3 families for protein digestibility (PDigest, P<0.001), panicle width (WPAN, P < 0.001), panicle length (LPAN, P = 1%), peduncle length (LPED P = 5%), and plant height (HPL, P = 5%) (Table 5.5). However, there was no significant difference between genotypes for days to 50% flowering (DSFLO). All the measurements were obtained from one location (Bambey) and during one experimental year (2017).

Table 5.5: Mean square values for protein digestibility and agronomic traits for BC3F3 families.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>PDigest</th>
<th>WPAN</th>
<th>DSFLO</th>
<th>HPL</th>
<th>LPAN</th>
<th>LPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Families</td>
<td>129</td>
<td>375.4 ***</td>
<td>0.59 ***</td>
<td>36.1 ns</td>
<td>358 *</td>
<td>13.93 **</td>
<td>20.3 *</td>
</tr>
<tr>
<td>Rep</td>
<td>1</td>
<td>1108.6 **</td>
<td>2.60 ***</td>
<td>1762.8 ***</td>
<td>19819 ***</td>
<td>1.39 ns</td>
<td>828.1 ***</td>
</tr>
<tr>
<td>Bloc:Rep</td>
<td>18</td>
<td>119.3 ns</td>
<td>0.44 *</td>
<td>77.8 **</td>
<td>660 **</td>
<td>28.73 ***</td>
<td>24.2 *</td>
</tr>
<tr>
<td>Residuals</td>
<td>111</td>
<td>36.3</td>
<td>0.22</td>
<td>36.3</td>
<td>257</td>
<td>7.89</td>
<td>13.7</td>
</tr>
</tbody>
</table>

"***" Significant at P = 0.05, "**" Significant at P = 0.01, "***" Significant at P < 0.001, "ns" non significant. DSFLO = days to 50% flowering, HPL = plant height, LPED = peduncle length, LPAN = panicle length, WPAN = panicle width, PDigest = protein digestibility.

5.3.2 Performance of BC3F3 families

Based on the threshold for digestibility, 18 entries had protein digestibility ≥ 60% and formed the high digestible group. The remaining 110 were classified in the low digestible group. The high and
low protein digestible groups significantly differed ($P < 0.001$) for protein digestibility (Table 5.6).

No significant difference was observed between the groups for all the agronomic traits measured (days to flowering, panicle length, peduncle length, plant height, and panicle width) (Table 5.5).

Table 5.6: Analysis of variance between high digestible and low digestible groups.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>PDigest</th>
<th>WPAN</th>
<th>DSFLO</th>
<th>HPL</th>
<th>LPAN</th>
<th>LPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>1</td>
<td>27893.3***</td>
<td>0.08 ns</td>
<td>27.3 ns</td>
<td>254.8 ns</td>
<td>0.31 ns</td>
<td>1.52 ns</td>
</tr>
<tr>
<td>Rep</td>
<td>1</td>
<td>1108.6 **</td>
<td>2.60 *</td>
<td>1762.8 ***</td>
<td>19818.8 ***</td>
<td>1.39 ns</td>
<td>828.06 ***</td>
</tr>
<tr>
<td>Bloc: Rep</td>
<td>18</td>
<td>123.9 ns</td>
<td>0.75 *</td>
<td>89.1 ***</td>
<td>1003.4 ***</td>
<td>46.73 ***</td>
<td>42.11 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>239</td>
<td>159.8</td>
<td>0.40</td>
<td>35.4</td>
<td>286.0</td>
<td>9.83</td>
<td>15.97</td>
</tr>
</tbody>
</table>

“***” Significant at $P < 0.001$, “ns” non significant. $PDigest =$ protein digestibility; $WPAN =$ panicle width; $DSFLO =$ Days to 50% flowering; $HPL =$ plant height; $LPAN =$ panicle length; $LPED =$ peduncle length.

The top 20 entries had a protein digestibility range of 58 to 91.1% which included 18 high digestible families i.e. have 60% protein digestibility or more (Table 5.6). Among the latter, eight families (E8, E20, E143, E34, 49, E113, E99, and E56) outperformed the mutant parent P721Q in terms of protein digestibility which ranged from 70.25 to 91% for the 8 families. The lowest 10 show protein digestibility values lower than the low digestible parent Faourou (Table 5.7) suggesting transgressive segregation.
Table 5.7: Mean performance of twenty highest and ten lowest digestible entries.

<table>
<thead>
<tr>
<th>Entries</th>
<th>PDigest</th>
<th>HPL</th>
<th>DSFLO</th>
<th>LPED</th>
<th>LPAN</th>
<th>WPAN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8</td>
<td>91.1</td>
<td>113.0</td>
<td>88</td>
<td>30.0</td>
<td>23.0</td>
<td>2.0</td>
</tr>
<tr>
<td>E20</td>
<td>84.0</td>
<td>108.5</td>
<td>94</td>
<td>30.5</td>
<td>20.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E143</td>
<td>79.4</td>
<td>136.0</td>
<td>84</td>
<td>36.0</td>
<td>27.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E49</td>
<td>78.2</td>
<td>134.5</td>
<td>88</td>
<td>35.5</td>
<td>23.0</td>
<td>3.5</td>
</tr>
<tr>
<td>E113</td>
<td>77.0</td>
<td>134.5</td>
<td>91</td>
<td>33.5</td>
<td>24.0</td>
<td>4.0</td>
</tr>
<tr>
<td>E99</td>
<td>76.8</td>
<td>120.0</td>
<td>97</td>
<td>35.0</td>
<td>21.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E56</td>
<td>75.0</td>
<td>135.0</td>
<td>95</td>
<td>37.0</td>
<td>26.0</td>
<td>5.0</td>
</tr>
<tr>
<td>E40</td>
<td>70.3</td>
<td>142.0</td>
<td>85</td>
<td>38.0</td>
<td>23.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E55</td>
<td>69.3</td>
<td>139.0</td>
<td>90</td>
<td>38.5</td>
<td>26.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E1</td>
<td>68.8</td>
<td>129.0</td>
<td>101</td>
<td>37.0</td>
<td>23.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E23</td>
<td>66.4</td>
<td>151.5</td>
<td>91</td>
<td>45.0</td>
<td>22.5</td>
<td>2.0</td>
</tr>
<tr>
<td>E45</td>
<td>64.4</td>
<td>156.0</td>
<td>81</td>
<td>45.0</td>
<td>25.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E38</td>
<td>63.8</td>
<td>126.0</td>
<td>95</td>
<td>32.0</td>
<td>25.5</td>
<td>2.0</td>
</tr>
<tr>
<td>E3</td>
<td>62.6</td>
<td>130.0</td>
<td>95</td>
<td>34.5</td>
<td>25.5</td>
<td>3.5</td>
</tr>
<tr>
<td>E48</td>
<td>61.3</td>
<td>118.5</td>
<td>92</td>
<td>32.5</td>
<td>24.0</td>
<td>2.5</td>
</tr>
<tr>
<td>E111</td>
<td>60.6</td>
<td>129.0</td>
<td>93</td>
<td>32.5</td>
<td>24.0</td>
<td>3.5</td>
</tr>
<tr>
<td>E138</td>
<td>60.1</td>
<td>121.5</td>
<td>95</td>
<td>33.5</td>
<td>24.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E2</td>
<td>60.0</td>
<td>139.0</td>
<td>78</td>
<td>37.0</td>
<td>25.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E140</td>
<td>58.3</td>
<td>138.5</td>
<td>93</td>
<td>34.5</td>
<td>29.0</td>
<td>4.0</td>
</tr>
<tr>
<td>E149</td>
<td>58.0</td>
<td>124.0</td>
<td>92</td>
<td>35.5</td>
<td>23.5</td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Lowest 10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E117</td>
<td>27.5</td>
<td>138.0</td>
<td>93</td>
<td>33.0</td>
<td>23.5</td>
<td>3.5</td>
</tr>
<tr>
<td>E102</td>
<td>27.5</td>
<td>133.0</td>
<td>90</td>
<td>34.0</td>
<td>27.0</td>
<td>2.5</td>
</tr>
<tr>
<td>E93</td>
<td>25.4</td>
<td>128.0</td>
<td>88</td>
<td>33.5</td>
<td>24.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E52</td>
<td>24.8</td>
<td>125.0</td>
<td>83</td>
<td>37.0</td>
<td>25.0</td>
<td>2.0</td>
</tr>
<tr>
<td>E92</td>
<td>24.1</td>
<td>122.5</td>
<td>94</td>
<td>36.5</td>
<td>23.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E70</td>
<td>22.7</td>
<td>135.5</td>
<td>84</td>
<td>39.5</td>
<td>24.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E135</td>
<td>22.3</td>
<td>119.0</td>
<td>92</td>
<td>36.0</td>
<td>20.0</td>
<td>3.0</td>
</tr>
<tr>
<td>E136</td>
<td>21.9</td>
<td>126.0</td>
<td>97</td>
<td>34.5</td>
<td>25.0</td>
<td>3.5</td>
</tr>
<tr>
<td>E83</td>
<td>19.3</td>
<td>109.0</td>
<td>90</td>
<td>32.0</td>
<td>22.5</td>
<td>3.5</td>
</tr>
<tr>
<td>E60</td>
<td>14.4</td>
<td>120.0</td>
<td>92</td>
<td>32.0</td>
<td>22.0</td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Checks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P721Q</td>
<td>70.0</td>
<td>110.5</td>
<td>96</td>
<td>35.0</td>
<td>20.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Faourou</td>
<td>39.0</td>
<td>137.5</td>
<td>96</td>
<td>37.5</td>
<td>24.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**PDigest** = protein digestibility; **WPAN** = panicle width; **DSFLO** = Days to 50% flowering; **HPL** = plant height; **LPAN** = panicle length; **LPED** = peduncle length.
5.3.3 Performance of high and low protein digestible entries

Protein digestibility ranged from 7 to 58% for the low protein digestible families and 60 to 90% for the high protein digestible families with a mean of 40.05% and 73.54% respectively. A wide range of difference existed for plant height and panicle length. Plant height varied between 99 and 167 cm for the high protein digestible, 74 and 174 cm for the low protein digestible families while panicle length was 18-30 cm, and 15-48 cm respectively (Table 5.8). Panicle width for both groups of families had the same range (2-5 cm).

Table 5.8: Comparison between the performance of high and low protein digestible BC3F3 families

<table>
<thead>
<tr>
<th>Traits</th>
<th>Range</th>
<th>Mean ± SE</th>
<th>Range</th>
<th>Mean ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDigest (%)</td>
<td>60-90</td>
<td>70.46 ± 2.87</td>
<td>7-58</td>
<td>41.13 ± 0.80</td>
</tr>
<tr>
<td>WPAN (cm)</td>
<td>2-5</td>
<td>3.06 ± 0.13</td>
<td>2-5</td>
<td>3.08 ± 0.04</td>
</tr>
<tr>
<td>DSFLO (days)</td>
<td>77-114</td>
<td>90.50 ± 1.23</td>
<td>77-116</td>
<td>89.74 ± 0.45</td>
</tr>
<tr>
<td>HPL (cm)</td>
<td>99-167</td>
<td>131.28 ± 2.51</td>
<td>74-174</td>
<td>132.95 ± 1.42</td>
</tr>
<tr>
<td>LPAN (cm)</td>
<td>18-30</td>
<td>24.11 ± 0.42</td>
<td>15-48</td>
<td>24.01 ± 0.25</td>
</tr>
<tr>
<td>LPED (cm)</td>
<td>26-48</td>
<td>35.89 ± 0.82</td>
<td>24-50</td>
<td>35.89 ± 0.31</td>
</tr>
</tbody>
</table>

PDigest = protein digestibility; WPAN = panicle width; DSFLO = Days to 50% flowering; HPL = plant height; LPAN = panicle length; LPED = peduncle length; SE = Standard error.

5.3.4 Variance components and heritability estimates

For all the traits measured, most of the variation was attributed to error (Table 5.9). The genotypes contributed less to the variation observed. Broad-sense heritability estimated for all traits at Bambey during 2017 rainy season was 58.9% for protein digestibility, and 61.3% for panicle width.
Plant height was the least heritable trait with 18.3%, followed by peduncle length (24.2%), and panicle length (29.6%).

Table 5.9: Variance components and heritability estimates at Bambey research station

<table>
<thead>
<tr>
<th>Source</th>
<th>PDigest</th>
<th>WPAN</th>
<th>DSFLO</th>
<th>HPL</th>
<th>LPAN</th>
<th>LPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypes</td>
<td>110.7</td>
<td>0.17</td>
<td>0</td>
<td>28.65</td>
<td>1.71</td>
<td>2.19</td>
</tr>
<tr>
<td>Error</td>
<td>154.1</td>
<td>0.22</td>
<td>35.22</td>
<td>256</td>
<td>2.85</td>
<td>13.72</td>
</tr>
<tr>
<td>Heritability</td>
<td>58.9</td>
<td>61.3</td>
<td>0</td>
<td>18.3</td>
<td>29.6</td>
<td>24.2</td>
</tr>
</tbody>
</table>

*PDigest = protein digestibility; WPAN = panicle width; HPL = plant height; LPAN = panicle length; LPED = peduncle length.*

5.3.5 Phenotypic correlation between protein digestibility and agronomic traits

Positive significant correlations (P < 0.001) were found between peduncle length, panicle length, and plant height. Days to 50% flowering was significantly negatively correlated to panicle length, plant height and peduncle length (P < 0.001). In addition, panicle width was positively correlated to panicle length (P = 1%), and peduncle length (P = 5%). No significant correlation was observed between protein digestibility and any of the agronomic traits (Table 5.10).
Table 5.10: Correlation coefficients between protein digestibility and agronomic traits. N=130

<table>
<thead>
<tr>
<th></th>
<th>LPAN</th>
<th>WPAN</th>
<th>DSFLO</th>
<th>HPL</th>
<th>LPED</th>
</tr>
</thead>
<tbody>
<tr>
<td>WPAN</td>
<td></td>
<td></td>
<td>-0.048 ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSFLO</td>
<td>-0.2***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPL</td>
<td>0.4***</td>
<td>0.23***</td>
<td></td>
<td>-0.59***</td>
<td></td>
</tr>
<tr>
<td>LPED</td>
<td>0.33***</td>
<td>0.12*</td>
<td>-0.52***</td>
<td></td>
<td>0.76***</td>
</tr>
<tr>
<td>PDigest</td>
<td>0.035 ns</td>
<td>0.05 ns</td>
<td>0.043 ns</td>
<td>0.089 ns</td>
<td>0.047 ns</td>
</tr>
</tbody>
</table>

“*” Significant at P = 0.05, “**” Significant at P = 0.01, “***” Significant at P < 0.001, “ns” non significant.

PDigest = protein digestibility; WPAN = panicle width; DSFLO = Days to 50% flowering; HPL = plant height; LPAN = panicle length; LPED = peduncle length.

5.3.6 Classification of BC₃F₃ families

The 130 families used in this study (128 BC₃F₃ and 2 parents) were classified into three groups based on their characteristics for the 6 variables measured (Table 5.11). Group 1 encompassed 22 lines among which the 18 high digestible progenies and the mutant parent P721Q. The group 1 was characterized by lines with the highest digestibility values with a mean of 67.77%. Group 2 comprises 63 lines with the lowest protein digestibility mean (39.43%). Group 1 and group 2 did not differ much for the agronomic traits. However, group 2 has the smallest panicle width (3.02 cm) while group 1 and group 3 have the same panicle width. Group 3 is made of 45 lines comprising the low digestible recurrent parent Faourou. Group 3 has the tallest plants (144.76 cm), early flowering (88.5 days), with the longest peduncle and panicle.
Table 5.11: Classification of the lines by Agglomerative Hierarchical Clustering (AHC)

<table>
<thead>
<tr>
<th>Class \ Variable</th>
<th>Number</th>
<th>PDigest</th>
<th>HPL</th>
<th>DSFLO</th>
<th>LPED</th>
<th>LPAN</th>
<th>WPAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>22</td>
<td>67.77</td>
<td>127.14</td>
<td>91.34</td>
<td>34.8</td>
<td>23.86</td>
<td>3.11</td>
</tr>
<tr>
<td>Group 2</td>
<td>63</td>
<td>39.43</td>
<td>125.78</td>
<td>90.47</td>
<td>34.23</td>
<td>22.63</td>
<td>3.02</td>
</tr>
<tr>
<td>Group 3</td>
<td>45</td>
<td>42.82</td>
<td>144.76</td>
<td>88.5</td>
<td>38.69</td>
<td>26</td>
<td>3.11</td>
</tr>
</tbody>
</table>

PDigest = protein digestibility; WPAN = panicle width; DSFLO = Days to 50% flowering; HPL = plant height; LPAN = panicle length; LPED = peduncle length.

Discriminant analysis performed to assess the correctness of the classification obtained from the Agglomerative Hierarchical Clustering (AHC) showed a true clustering for all the genotypes except entry number 26 and number 23 (Figure 5.4). In addition, the Lambda Wilk test revealed that days to flowering and panicle width, with lambda values of 0.934 and 0.993 respectively, were the least discriminant variables for clustering of the genotypes (Table 5.12).

Table 5.12: Lambda Wilk test for the traits measured

<table>
<thead>
<tr>
<th>Variables</th>
<th>Lambda</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pdigest</td>
<td>0.440</td>
<td>80.774</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>HPL</td>
<td>0.557</td>
<td>50.477</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>DSFLO</td>
<td>0.934</td>
<td>4.479</td>
<td>0.013</td>
</tr>
<tr>
<td>LPED</td>
<td>0.578</td>
<td>46.436</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>LPAN</td>
<td>0.667</td>
<td>31.711</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>WPAN</td>
<td>0.993</td>
<td>0.426</td>
<td>0.654</td>
</tr>
</tbody>
</table>

PDigest = protein digestibility; WPAN = panicle width; DSFLO = Days to 50% flowering; HPL = plant height; LPAN = panicle length; LPED = peduncle length.
Figure 5.4: Discriminant Analysis of the study population showing entries grouped in 3 main clusters.
5.4 Discussion

Significant differences were noted between genotypes for plant height, peduncle length, panicle length, panicle width, and protein digestibility (p<0.001). However, no significant difference was observed between genotypes for days to 50% flowering. These findings would imply that the families have a similar life cycle and due to that similarity, the families can be selected based on their differences in protein digestibility and could be well adapted to the same range of environments.

A high broad sense heritability for protein digestibility (58.9%) was obtained at Bambey, Senegal for the BC$_3$F$_3$ which implied that the phenotypic variation for protein digestibility observed in the population is mostly controlled by genotypic effects. Therefore, improvement for this trait can be made through breeding and selection. However, this study was limited by the number of years and locations where the entries were tested. Therefore, the true improvement for protein digestibility in sorghum that can be made through breeding could be better assessed by replicating the study in different years and locations. No heritability studies had been reported before for this valuable trait.

A positive phenotypic correlation was detected between panicle width, panicle length, plant height, and peduncle length. However, days to 50% flowering was significantly negatively correlated with panicle length, plant height, and peduncle length. This implies that the early flowering genotype families had shorter plants, shorter panicles and peduncles. No correlation was observed between protein digestibility and the agronomic traits measured. Therefore, these traits cannot be used for early identification and phenotypic selection of genotypes carrying the improved protein digestibility characteristics.
Marker-assisted selection was used during early generations of breeding to accelerate the introgression of high protein digestibility into a locally-adapted Senegalese sorghum variety. Among the 128 BC3F3 families evaluated in the field, 18 were found to be highly digestible after running the digestibility assay on the resulting BC3F4 seeds (Table 5.7). These results imply that the allele controlling the increase in sorghum protein digestibility after wet cooking was successfully introgressed into the background of the white grained tannin-free Senegalese variety, Faourou. Comparison between highly digestible and lowly digestible groups revealed no significant differences between the genotypes for the agronomic traits. The BC3F3 families were not significantly different from the recurrent parent, Faourou. These results imply that the presence of the desirable high protein digestibility allele did not negatively affect the background of the local variety used in this study. The results of this study show that it is possible to introgress the high protein digestibility allele into local germplasm which can be a source of high protein digestible germplasm for breeding programmes.

The protein digestibility values ranged between 60 and 91% for the 18 highly digestible progenies among which eight families (E8, E20, E143, E49, E113, E99, E56, E40) outperformed the highly digestible donor parent P721Q which could be explained by transgressive segregation where many favourable alleles from the parents would combine in the progenies causing the extreme phenotype observed. Winn et al. (2009) reported a similar phenomenon where the progeny of a RIL population, developed from a cross between a high and a low digestible sorghum lines, showed lower digestibility values than the low digestible parent. These findings are very important and beneficial since sorghum is naturally of low digestibility after wet cooking (Axtell et al., 1981), and no West African sorghum accessions tested were found to be highly digestible (see chapter 3). These highly digestible genotypes, in addition to the potential high lysine trait from the donor
parent (need further analysis), their tannin-free stage, could be a good source of nutrients for humans especially for the formulation of infant food to fight disease like Kwashiorkor which is a result of protein-poor diet. Multiple studies reported that chickens and beef cattle fed with tannin-free varieties showed significantly higher growth rate compared to tannin-rich sorghums and that sorghum could be used to replace maize as animal feed (Spicer et al., 1983; Gualtieri and Rapaccini, 1990). Therefore, these tannin-free, highly digestible sorghum progenies could be used to enhance production and help lower the cost of animal feed.

Furthermore, in Senegal, more than 100 million US dollars is invested per year in wheat importation (source FAO). To cut on this cost, institutions like ITA are currently working on substituting wheat with more affordable and available cereal crops like white-grained sorghum. Sambe and Tounkara (2017) showed that bread made from a mixture of 80% wheat flour and 20% Faourou flour (sorghum recurrent parent used in this study) has the same texture and taste as bread made from 100% wheat. Hence, promoting these not only white-grained sorghums but also more nutritious types would have a huge economic impact for developing countries and constitute a significant step towards promoting locally-adapted crops for achieving food security in Africa and in the rest of the World.

Initially, breeders were focused on yield improvement with little attention given to food quality or nutritional value of the grain. Despite the higher yielding varieties developed, farmers would still prefer their locally-adapted varieties because the grains have better taste and are suited for their traditional cooking system (Witcombe et al., 2005). Therefore, with this study, providing farmers with highly digestible protein sorghum varieties and adapted to local environments will contribute to fighting malnutrition which can be a real challenge especially for families who have a sorghum-based diet.
5.5 Conclusion

This study revealed no correlation between protein digestibility and agronomic traits which implies that the latter cannot be used for early-stage indirect selection for high protein digestible sorghum genotypes. In Bambey, 18 highly digestible BC$_3$F$_3$ families were identified among which, 8 outperformed the highly digestible mutant/donor parent. The mutant allele controlling high protein digestibility in P721Q was successfully introgressed into the background of Faourou, a white grained highly preferred sorghum variety in Senegal. This shows that it is possible to develop varieties with higher digestible proteins adapted to local environments. Additional work will be conducted in more contrasting environments to assess the yield performance of the 18 high digestible families and release the best performing lines. These improved varieties will constitute a good source of highly digestible sorghum and contribute to fighting malnutrition.
CHAPTER SIX

6.0 GENERAL CONCLUSIONS AND RECOMMENDATIONS

6.1 General conclusions

It was observed in this study that anti-nutritional factors like tannins, acid detergent fiber had a negative effect on protein digestibility. In addition, even though seed colour was negatively correlated with the level of digestibility and the presence of tannin, seed colour cannot be used as an indicator to select for lines with high digestibility and tannin.

All accessions of the West African Sorghum Association Panel (WASAP) were of low protein digestibility after wet cooking i.e. less than 60% digestibility. However, variability for protein digestibility was observed which ranged from 6 to 55%. Therefore, it was important to find sources of high protein digestibility in sorghum to contribute to the fight against malnutrition and contribute to food security.

Sorghum EMS mutants (SbEMS1613 and SbEMS3324) were identified with higher lysine and 25 to 35% more protein digestibility than their wild type parent line BTx623. Contrary to previous studies, protein body structure cannot be used as the only indicator for high protein digestibility because SbEMS1613 had similar protein body structure to the wild type BTx623. Two mutations on chromosome 5 were identified to be associated with high protein digestibility in the mutants. In SbEMS3324, a G:A missense mutation on the PSK like kafirin gene led to the substitution of alanine by threonine which induced non-proper cleavage of the polypeptide chain and produced invaginated protein bodies. A single recessive allele controlled the change in phenotype. Similarly, in SbEMS1613, a point missense mutation (C:T) on a 26S proteasome subunit complex was linked
to high digestibility phenotype. The mutation caused a substitution of histidine by tyrosine in the polypeptide change.

Additionally, two pairs of SNP markers were developed and confirmed to be linked to the genes controlling high protein digestibility in each mutant. These primers can be used to accelerate the introgression of the high protein digestibility trait into sorghum cultivars through marker-assisted selection since phenotyping for the trait cannot be done at early stages of development due to its destructive nature.

Among the 128 BC3F3 families evaluated in the field and tested for protein digestibility, only 18 were highly digestible. Eight out of those 18 had higher protein digestibility values than the donor parent, P721Q. In addition, no significant difference was observed between the high digestible and the low digestible families for all the agronomic traits. These results showed that the allele controlling high protein digestibility in P721Q was successfully introgressed into the background of the white grained locally-adapted low digestible variety, Faourou. Hence, it is possible to breed for more digestible sorghum lines using locally-adapted and farmers’ preferred varieties. The lines developed in this study will undergo a multi-environmental evaluation to assess the yield performance and stability of genotypes as well as genotype by environment study and release the best performing lines. In addition, the lines will be a good source of highly digestible sorghum for breeding programmes across other countries and will aid fight malnutrition and decrease the cost of wheat importation in Senegal.

There was no correlation between the agronomic traits measured in this study and protein digestibility. Consequently, no indirect selection of highly digestible genotypes in a breeding
programme could be done based on the agronomic traits therefore, molecular markers would be a better choice.

6.2 Recommendations

None of the West African sorghum accessions had high digestible proteins. Therefore, the improvement of these lines for protein digestibility should be done for the improvement in nutrition in areas where sorghum is a staple food.

The two sorghum EMS mutants SbEMS1613 and SbEMS3324, characterized in this study could be used as good sources of high protein digestibility as well as lysine and tryptophan.

Since none of the agronomic traits could be used for early indirect selection for protein digestibility, the improvement of locally-adapted sorghum lines could be speed up with the SNP primer pairs (ProF and ProR, KafF and KafR) developed around the mutated genes (Sobic.005G189000 and Sobic.005G083340) identified to be causing the increase in digestibility in the EMS mutants.

Additional studies should be conducted to understand the pathway of increased protein digestibility in sorghum EMS mutants especially in SbEMS1613.

A high broad sense heritability of 58.9% was estimated for protein digestibility in the BC₃F₃ population. However, to better estimate how much improvement can be made through breeding, testing the entries across multiple locations and years should be considered.

The 18 high protein digestible backcross progenies developed in this study should be further tested for yield performance, lysine and tryptophan content for the release of the first ever reported West African sorghum varieties with highly digestible proteins.
This material could be used as a good nutrient supplement in sorghum-based diets or in weaning food for infants in many parts of Africa and Asia where it would be needed.

For countries where wheat production is not viable, blending sorghum flour from highly digestible varieties and wheat flour should be considered to help reduce the cost of wheat importation.
REFERENCES


sorghum varieties grouped according to food end-use properties. *Journal of the Science of Food and Agriculture*, 86(6), 953–963. https://doi.org/10.1002/jsfa.2443


APPENDICES

APPENDIX 5.1: Principal component analysis showing the contribution of variables measured on 18 highly digestible BC$_3$F$_3$ families

Principal component analysis showing contribution of variables measured on 18 highly digestible BC$_3$F$_3$ families
APPENDIX 5.2: Discriminant analysis for only highly digestible BC3F3 with parents

Discriminant analysis for only highly digestible BC3F3 with parents