QTL MAPPING OF HIGH PROTEIN DIGESTIBILITY TRAIT IN *Sorghum bicolor* (L.) Moench

A Thesis

by

JENNIFER ANN WINN

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

MASTER OF SCIENCE

December 2007

Major Subject: Plant Breeding

QTL MAPPING OF HIGH PROTEIN DIGESTIBILITY TRAIT IN *Sorghum bicolor* (L.) Moench

A Thesis

by

JENNIFER ANN WINN

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

MASTER OF SCIENCE

Approved by:

Chair of Committee, Dirk B. Hays
Committee Members, William L. Rooney
Stephen R. King
Head of Department, David Baltensperger

December 2007

Major Subject: Plant Breeding

ABSTRACT

QTL Mapping of High Protein Digestibility

Trait in *Sorghum bicolor* (L.) Moench. (December 2007)

Jennifer Ann Winn, B.S., Texas A&M University

Chair of Advisory Committee: Dr. Dirk B. Hays

As compared with other cereal grains, *Sorghum bicolor* shows very low levels of protein digestibility when exposed to proteolytic enzymes. Protein digestibility further decreases when sorghum is cooked. It is speculated that low digestibility is the result of extensive disulfide crosslinking in the β- and γ-kafirins (storage proteins) surrounding the endosperm protein bodies. The degree of crosslinking increases as sorghum is cooked, causing the highly digestible α-kafirins found at the interior of protein bodies to be locked within a tightly bound capsule, inaccessible to digestive enzymes. In this research project, two major QTLs were found to be associated with protein digestibility—one QTL unfavorably affecting digestibility and one QTL favorably affecting digestibility. By identifying the QTLs and the linked markers corresponding to the highly digestible trait, breeders will be able to use marker-assisted selection to quickly and accurately identify highly digestible lines to advance in a breeding program.

ACKNOWLEDGEMENTS

I would like to first of all sincerely thank my advisor, Dr. Dirk Hays, for his support and wisdom. I am very grateful for the opportunity to work in such a wonderful, welcoming lab. I would also like to thank Dr. C. Wayne Smith and my committee members, Dr. William Rooney and Dr. Stephen King, for their help along the way. I could not have succeeded without my infinitely patient and sweet labmates, who each taught me something valuable and new, in lab and out—Esten Mason, Arlene Pacheco, Suchismita Mondal, Ostilio Portillo, Adriana Robbins, Jung Hwa Do, and my friend Candace Seeve. Of course, I would also like to thank my cheerleading parents and brothers, Mike and Andy, for all their support.

Above all, I would like to thank God for more blessings than I deserve.

ABBREVIATIONS

HD = highly digestible

LD = lowly digestible

LOD = likelihood of odds

MAS = marker-assisted selection

PCR = polymerase chain reaction

QTL = quantitative trait loci

SSR = simple sequence repeat (microsatellite) marker

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Millions of people in Africa, Asia, and other semi-arid regions throughout the world depend on sorghum as a staple crop. In many households, sorghum is the primary source for energy, protein, vitamins, and minerals (Klopfenstein and Hoseney1995). As the fifth most abundant crop worldwide (Doggett 1988) and the third most economically important crop in the US (Shantharam 1995), sorghum plays a huge role on the world market as a means of livelihood for millions of subsistence farmers and as a important part of food security. Each year 27 million tons of sorghum are consumed as food, mostly in the form of flat breads, porridge, and deep-fried or boiled products. The majority of sorghum used as food is consumed in Africa and Asia. In Sudan, for example, the annual per capita consumption reaches 90-100kg and comprises approximately a third of total caloric intake. Furthermore, sorghum is used as an important feed source, particularly in developed countries such as the US. Worldwide, 31 million tons, or 48% of all sorghum grown, is used for livestock feed (ICRISAT 1996).

Although sorghum is used by millions of people worldwide for food and feed, the grain is low in protein digestibility. When cooked, only 46% of the total protein found in sorghum is digestible, as compared with 81% in wheat, 73% in maize, and 66% in rice (MacLean et al. 1981). The primary use for sorghum as food is in the form of cooked porridge. To prepare the porridges, sorghum flour is mixed with boiling water to achieve

---

This thesis follows the style of *Crop Science*.

a consistency ranging from thick to thin (Murty 1995 and Kumar; Taylor et al. 1997). Most types of preparation include heating of sorghum flour. Unfortunately, the protein digestibility of sorghum dramatically decreases when heated, as compared with other cereal grains. Cooking reduces sorghum's digestibility by more than half. However, the digestibility of cooked maize, for example, has been shown to be reduced only by a negligible amount when cooked. Digestibility of uncooked sorghum ranges from 78-100% but reduces to 45-55% after cooking (Axtell et al. 1981; MacLean et al. 1981).

Several theories have been proposed as to the cause of sorghum's reduced digestibility. Studied extensively by Duodu et al. (2003), these theories can be classified into exogenous factors (including protein interactions with non-protein components of the grain) and endogenous factors (like changes within the protein itself). The exogenous factors studied include ways in which protein associates with lipids, phytates, starch, and polysaccharides. Although some of these factors proved to correlate with protein digestibility, the endogenous factors—such as disulfide crosslinking and alterations in the secondary structure of the protein—appeared to play the greatest role in affecting protein digestibility (Duodu et al. 2003).

Sorghum protein is found mostly in the form of prolamins, which are storage molecules of protein in the grain endosperm. These prolamins, called kafirins, are characterized into three distinct categories: α-, β-, and γ-kafirins. All three kafirins are aqueous alcohol-soluble but differ in terms of solubility, molecular weight, and structure (Shull et al. 1991). At the periphery of the spherical protein bodies are β- and γ-kafirins, making up a combined 20% of kafirin content, most of which are γ-kafirins. Both β- and

γ-kafirins have high concentrations of the amino acid cysteine. Of total amino acid content in β-kafirins, 5% is cysteine; in γ-kafirins, 7% is cysteine (Shull et al. 1992). α-kafirins are found at the interior of the protein body and make up ≈80% of kafirin content and ≈60-70% of total protein content within the grain (Hamaker et al. 1995).

The high cysteine concentration in β- and γ-kafirins (as compared with α-kafirins' low cysteine content) causes extensive disulfide crosslinking when cooked. As a result, the kafirins (particularly the γ-kafirins) form polymers that create a tightly-bound structure encapsulating the α-kafirins. The crosslinked γ-kafirin barrier is resistant to proteolytic enzymes, which in turn causes the normally highly-digestible α-kafirins to be inaccessible to proteolytic enzymes (Duodu et al. 2003). When sorghum is cooked with a reducing agent, such as 2-mercaptoethanol—which breaks disulfide crosslinking—protein digestibility is higher than when cooked in water, verifying the observation that it is disulfide crosslinking that causes inhibited digestibility (Hamaker et al. 1995; Oria et al. 1995). Much of the protein in wet, cooked sorghum remains locked up due to the γ-kafirin crosslinking.

Weaver et al. (1998) identified highly digestible (HD) sorghum lines derived from a high-lysine chemical mutant (P721 Opaque, also known as P721Q) (Mohan 1975). The highly digestible lines were found to have ≈10-15% higher protein digestibility when uncooked and ≈25% higher digestibility when cooked. More specifically, the α-kafirins increased to ≈90-95% digestibility following pepsin digestion, as compared with ≈45-60% digestibility in normal cooked lines (Weaver et al. 1998).

The cause for the higher digestibility in specific lines is due to rearrangement of the kafirins, particularly the γ-kafirins, located at the exterior of the protein bodies. Instead of the γ-kafirins being located around the periphery of a protein body as they are in normal lines, highly digestible lines possess γ-kafirins that are found only in pockets of folds within the total protein body (Oria et al. 2000). Fig. 1 shows the relative distribution of kafirins within a protein body.

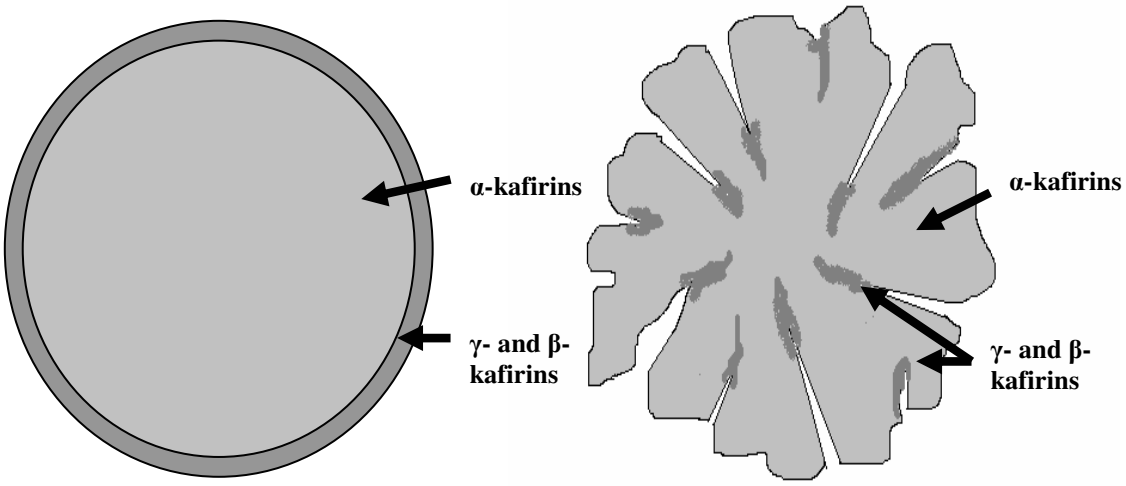


Fig. 1: Kafirin distribution in protein storage bodies. At left is a cross-section of a normal sorghum protein body with γ- and β-kafirins encapsulating α-kafirins. At right is a highly digestible protein body with the γ- and β-kafirins found only in small inclusions throughout the highly folded protein body. Figures are representative of the TEM results found by Oria et al. (2000).

As a result of the physiological changes of the HD endosperm protein body structure, the interior α-kafirins are exposed, making them susceptible to proteolytic enzymes. Furthermore, with the invaginated structure of the highly digestible lines, there is more total surface area available for hydrolysis by digestive enzymes.

In this research project, the genes controlling kafirin rearrangement are analyzed using molecular markers. Marker-assisted selection (MAS) to supplement crop breeding programs is becoming increasingly more common. As more linkage maps are generated to locate specific traits, genomewide marker precision increases and the efficiency of MAS improves. MAS is beneficial to breeders because it allows for the rapid identification of traits which may be controlled by many genes (making them difficult to breed for in a conventional scheme) or traits which previously could only be identified using time-consuming, expensive, or highly-specialized procedures. With the protein digestibility trait, MAS would be particularly useful because it would permit the breeder to circumvent the cumbersome turbidity assay to distinguish between HD and LD lines.

To use MAS in a breeding program, the markers associated with the trait of interest must first be identified. Marker saturation throughout the genome is ideal to locate the gene(s) controlling the trait of interest—genes in a region referred to as the quantitative trait loci (QTL). QTL analysis is based upon the principle that markers closely linked to a gene of interest will show linkage disequilibrium less frequently than markers not linked to the gene (Tanksley 1993). That is, the closer a marker lies to the gene, the less often the marker and the gene will be separated during meiotic recombination.

Simple sequence repeats (SSRs) were ideal markers to use in this project for several reasons. SSRs are inherited in a codominant Mendelian manner, so they are able to provide information about heterozygosity. Length variability in SSR markers makes them highly polymorphic, and SSR markers are stable in somatic cells (Morgante and Olivieri 1993). Furthermore, many SSR markers have been previously mapped in the sorghum genome, and there is fairly dense coverage across the genome (Bhattramakki et al. 2000; Wu and Huang 2007).

MATERIALS AND METHODS

The objective of this research is to identify molecular markers linked with the protein digestibility trait in *Sorghum bicolor* so that a marker-assisted selection procedure can be effectively used by breeders. The central hypothesis is that there is one gene or a few genes regulating the arrangement of γ-kafirins around protein bodies in sorghum, which influences total protein digestibility. The specific aims of this research are to:

1. Phenotypically differentiate sorghum lines that are highly and lowly digestible.

2. Correlate high-protein digestibility to a QTL and identify linked markers.

Research Objective 1: Phenotypically differentiate sorghum lines that are highly and lowly digestible.

- **Rationale**

To identify a QTL, it was first necessary to phenotype individual lines. To distinguish between highly and lowly digestible lines, a turbidity assay was used, derived from the procedure described by Aboubacar et al. (2003).

- **Methods**

Seed in the F_4 generation was obtained from 277 recombinant inbred lines resulting from a cross between the highly digestible line P850029 and the wild type line Sureno. Many individuals had sib lines used in the study. An individual and its sib (that is, two $F_{3:4}$ lines) were from the same parents and grown in the same location. Sib lines were valuable because they could approximate a replication in statistical analyses since seed from multiple years, replications, and environments was not available.

To begin pepsin digestion, 50mg of seed from each sample was ground and added to 1mL pepsin solution (20mg pepsin/mL 0.1M KH_2PO_4, pH 2) in a 1.5mL tube. The mixture was vortexed and incubated at 37°C for 2 hours, with shaking at 150 rpm. The reaction was stopped by adding 100µL of 2N NaOH to each sample. Samples were centrifuged at 14,000 rpm for 10 minutes. Supernatants were discarded and samples were resuspended in 1mL 0.1M potassium phosphate buffer, pH 7. The samples were again centrifuged at 14,000 rpm for 10 minutes, supernatants were discarded, and the

pellets were washed with 1mL ddH_2O. Samples were centrifuged at 14,000 rpm for 10 minutes, supernatants were discarded, and protein extraction was carried out.

To extract proteins, the pellets after pepsin digestion were incubated at 37°C in a water bath with 0.5mL extraction buffer (0.0125M sodium tetraborate pH 10; 1% SDS (sodium dodecyl sulfate) W/V; 2% mercaptoethanol). Samples were centrifuged at 10,000 rpm for 10 minutes at 4°C. From the middle layer of supernatants, 200µL was transferred to a clean 1.5mL tube to continue to the turbidity assay.

To begin the turbidity assay, 25µL per sample was transferred to another new 1.5mL tube, to which 1mL of H_2O and 200µL of 72% TCA (trichloroacetic acid) were added. A blanking solution was prepared with 25µL protein extraction buffer, 1mL of H_2O, and 200µL of 72% TCA. The spectrophotometer was set to record turbidity at 562nm. Readings were taken at 15 and 30 minutes after adding TCA to the protein solution. Originally, measurements were also taken at 45 and 60 minutes, as well, but turbidity readings across time were not found to vary. The samples with the highest and lowest turbidity readings were rerun through the entire process (from the flour stage) to ensure the measurements were reliable in determining whether each sample could be considered LD or HD. To ensure reliability in readings, samples with any two readings differing by more than a value of 0.4 were eliminated from subsequent analyses. Averages of all measurements were taken for each sample, which were then used to rank the lines in terms of digestibility—the averages are what will from now on be referred to as the "turbidity values."

Research Objective 2: Correlate high-protein digestibility to a QTL and identify linked markers.

- **Rationale**

After phenotyping, each individual line's genotypic data was evaluated using a PCR procedure with simple sequence repeat (SSR) primers. With the genotypic data from the population, a QTL map was constructed using the computer programs Mapmaker 3.0 and QTL Cartographer.

- **Methods**

Each seed line (277 lines total, including parents) was grown out for 15 days, and plant tissue was collected. DNA was extracted based on the procedure described by Dellaporta et al. (1983).

To begin, 0.33g of plant tissue was ground into powder with mortar and pestle and added to a 15mL conical tube with 5mL extraction buffer (100mM Tris, pH 8; 50mM EDTA, pH8; 500mM NaCl; 10mM mercaptoethanol). After adding 333µL of 20% SDS, the samples were incubated for 30 minutes in a 65°C water bath. To remove proteins, 1.69mL of 5M potassium acetate was added and incubated for 20 minutes on ice. The samples were centrifuged at 4,500 rpm for 20 minutes and the supernatant was filtered through a Miracloth filter into a new 15mL tube holding 3.33mL isopropanol. The samples were incubated for 30 minutes at -20°C. Pellets were formed by centrifugation for 15 minutes at 4,500 rpm. Supernatants were discarded and pellets

were inverted to allow pellets to dry for 10 minutes. 233µL of 50mM Tris-10mM EDTA pH 8 was added to resuspend the pellets and the mixtures were transferred to new 1.5mL tubes. To each sample, 1.5µL of RNase (2mg/µL) was added, and tubes were incubated for 30 minutes at 37°C. 25µL of 3M NaOAC pH 5.2 and 167µL phenol/chloroform were added and samples were spun for 5 minutes at 10,000 rpm. The upper aqueous phases were transferred to new tubes with 167µL phenol/chloroform/IAA, and the solutions were centrifuged for 3 minutes at 10,000 rpm. The upper aqueous phases were again transferred to new tubes with 167µL isopropanol and spun for 5 minutes at 12,000 rpm. Supernatants were discarded, and the pellets were washed with 200µL 80% ethanol, resuspended, and spun again to repellet. After removing supernatants, the pellets were inverted to air dry for 15 minutes. Pellets were redissolved in 33µL 10mM Tris-1mM EDTA. The concentrations of the extracted DNA were determined at A_{260}. Each sample was diluted with ddH_20 to a final concentration of 10ng/µL.

Bulked segregant pools were created using the procedure developed by Michelmore et al (1991). The highly digestible (HD) DNA bulk was pooled from the five lines showing the lowest turbidity values. Each line was added in equal concentration, and the final bulked DNA was diluted to 10ng/uL. Similarly, the lowly digestible (LD) DNA bulk was pooled from the five lines showing the highest turbidity averages.

Next, 355 SSR primers were screened for polymorphisms using DNA from the 2 parents. PCR reactions were set up as follows for one 10µL reaction: 1µL buffer, 0.1µL dNTPs (10mM), 1.5µL each of forward and reverse primer (2µM), 0.5µL $MgCl_2$, 0.1µL

BSA, 2µL DNA (10ng/µL), 0.1µL Taq Polymerase, and 3.2µL H_2O. The PCR conditions were: 94°C for 5 minutes; 40 cycles of: 1 minute at 94°C, 1 minute at the primer annealing temperature, 1 minute at 72°C; 10 minutes at 72°C; hold at 4°C. All PCR products were run on 3% agarose gels for 2 hours. Approximately 100 of the 300 original SSR primers showed polymorphisms between the parents. The 100 primers were then run using the 2 DNA bulks (HD and LD), and the number of primers found to be polymorphic between the bulks was 8 (see Appendix B for primer information). Finally, the 8 primers were run across 70 randomly chosen individuals from the population. Only one individual from each sib pair was included in primer analyses to prevent overrepresentation of a genotype.

Genotypes from each individual were scored as being one of the 2 parental genotypes or as being heterozygous (Appendix A). Individual genotypes and corresponding phenotypes (as evaluated by turbidity values) were entered into MapMaker 3.0 and QTL Cartographer. A QTL map and marker linkage data were generated using segregation data. Distances are calculated using the Kosambi function (Kosambi 1944). Measurements are given in centiMorgans (cM).

RESULTS AND DISCUSSION

The results from the turbidity assay show a surprisingly gradual, not step-wise, increase in turbidity across the population of 277 individuals, including parents and sibs. If protein digestibility was truly controlled by a single, simply-inherited gene, as was previously theorized, the turbidity assay should show distinct cut-offs between LD, heterozygous, and HD individuals. However, as shown by Fig. 2, that was not the case; the turbidity values appeared as a gradual curve, not step-wise. As expected, the HD parent falls in the range of low turbidity, while the LD parent has higher turbidity. It is interesting to note that there are many individuals with turbidity values higher than the LD parent. It could be speculated that transgressive segregation or highly favorable multi-allelic combinations could be contributing to the phenotypes of these extreme individuals.

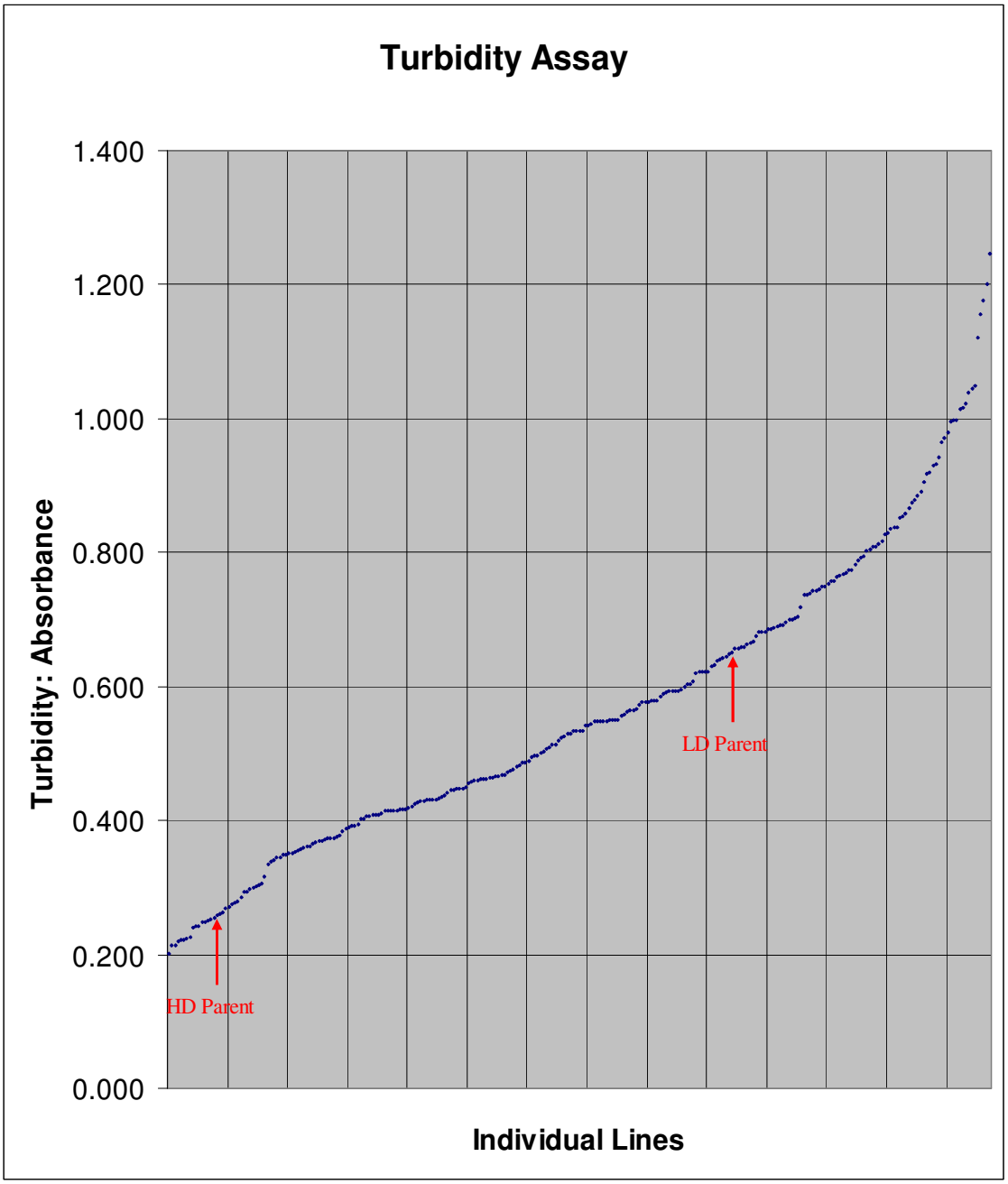


Fig. 2: Turbidity assay results. The turbidity of each sample at 562 nm correlates to protein digestibility. Samples with high protein digestibility are represented by low turbidity values. The distribution of turbidity values is surprisingly gradual (not step-wise). It is also interesting to note there are many lines showing transgressive segregation.

As consistent with the even slope in turbidity values (Fig. 2) the phenotypic values for the entire population, including parents and sib lines, display a slightly skewed distribution (Fig. 3).

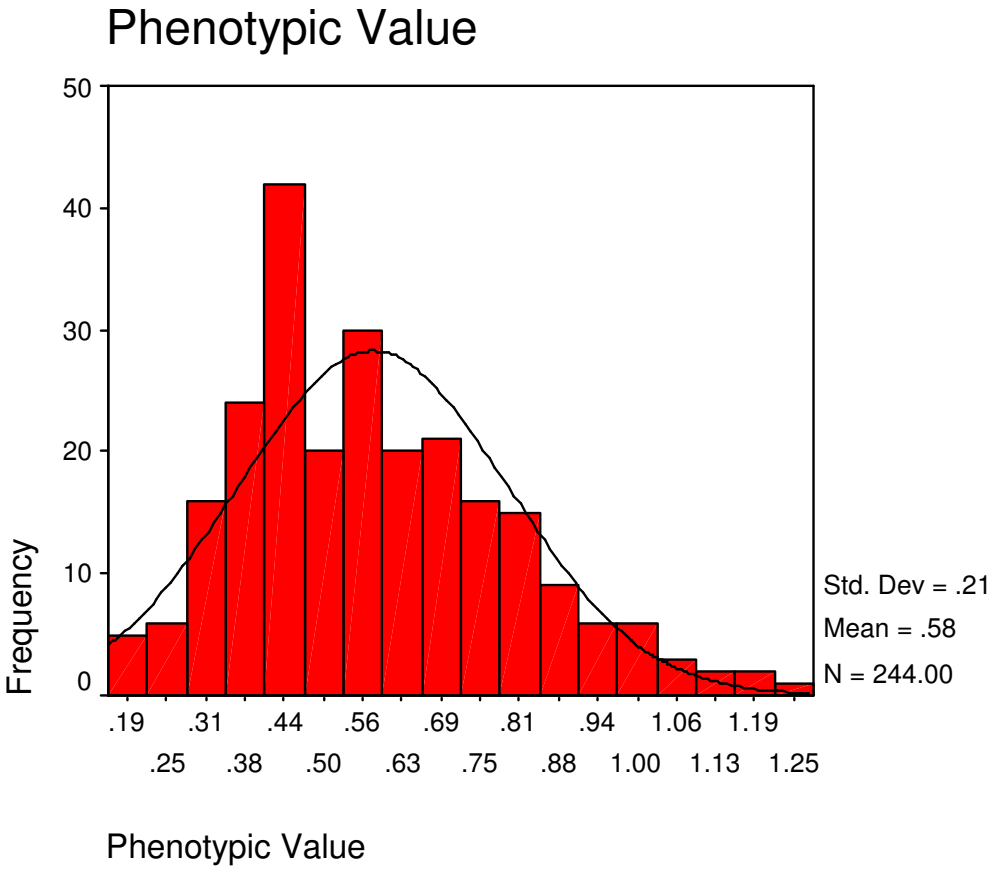


Fig. 3: Phenotypic distribution of entire population. The even distribution of phenotypic values—turbidity averages—suggests the possibility of digestibility being a multigene trait.

Similarly, the 70 lines (not including parents and sibs) used for the analysis of the 8 markers found to be polymorphic in screenings with the parents and the bulks showed a slightly skewed distribution for phenotypic values (Fig. 4).

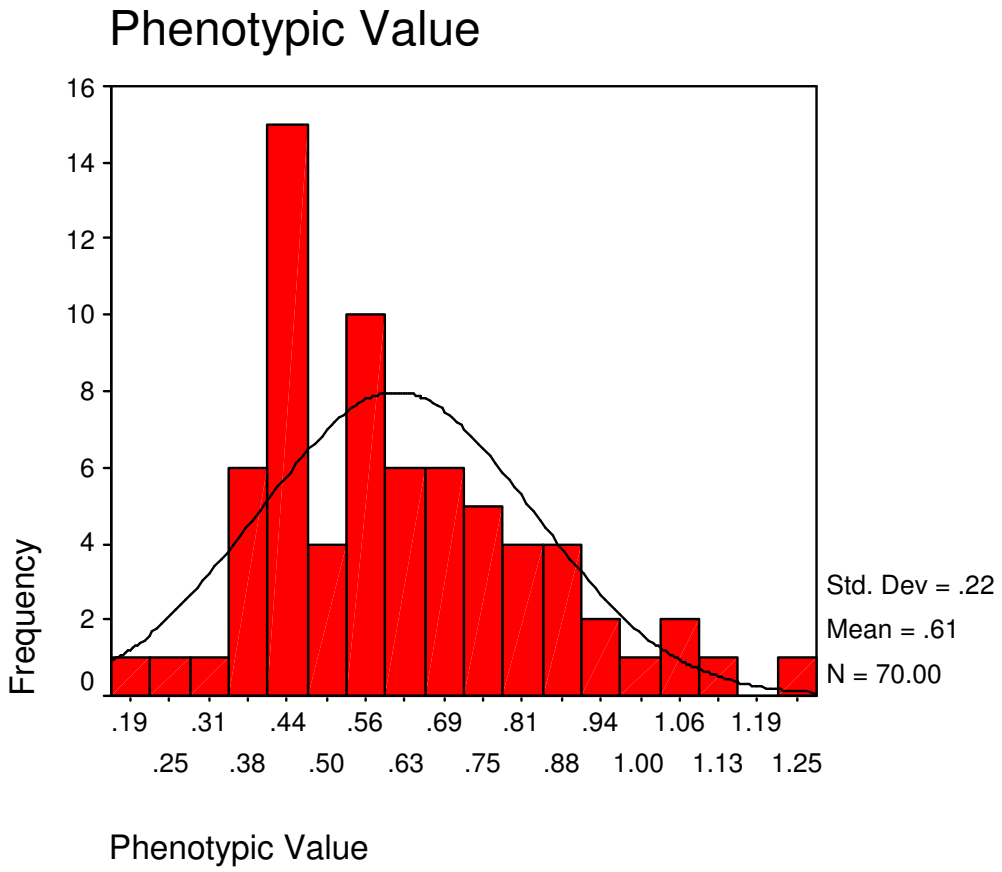


Fig. 4: Phenotypic distribution of the individuals used for the final marker analysis. Phenotypic values are averages from the turbidity assay. The phenotypic distribution is similar to the distribution of the entire population (Fig. 3). Thus, the 70 individuals used for marker analysis can be considered an accurate representation of the whole population during linkage mapping.

An ANOVA (analysis of variance) of phenotypic values was calculated using sib lines as a blocking effect to approximate replications (Table 1). Genotypes were found to be highly significant at α=0.001.

Table 1: ANOVA for phenotypic values.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.477(a)	108	.078	7.414	.000
Intercept	73.515	1	73.515	6944.184	.000
Genotype	8.477	108	.078	7.414	.000
Error	1.154	109	.011		
Total	83.146	218			
Corrected Total	9.631	217			

a R Squared = .880 (Adjusted R Squared = .761)

A summary of the phenotypic data according to Mapmaker 3.0/QTL Cartographer is as follows:

Table 2: Phenotypic variation—turbidity values.

Trait: Digestibility (Phenotypic Values)	
Sample Size	70
Mean Trait Value	0.6109
Variance	0.0478
Standard Deviation	0.2187
Coefficient of Variation	0.3579

Upon running the 8 polymorphic markers against 70 individuals, 6 markers were shown to segregate together by analysis with Mapmaker 3.0/QTL Cartographer (Fig. 5). According to previous marker analyses of the sorghum genome (Bhattramakki et al.

2000; Wu and Huang 2007), the six markers are located on Chromosome 1. A linkage map of the markers is shown below. Distances are given in centiMorgans (cM).

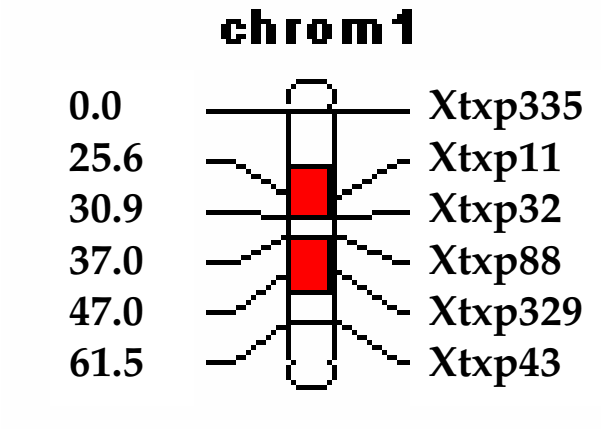


Fig. 5: Linkage map of primers segregating with protein digestibility trait. Shaded areas are significant QTLs. The region—Locus 1—near Xtxp11 unfavorably impacts digestibility, while the region—Locus 2—near Xtxp329 favorably impacts digestibility.

Two major QTLs (Fig. 6) were found to be significant at LOD>2.5. One major QTL (which will now be referred to as "Locus 1") occurs near marker Xtxp11 and shows a LOD score of 3.1. The QTL at this locus is surprising in that it displays dominance and additive effects that act unfavorably in terms of protein digestibility, as shown in Fig. 7 and Fig. 8. The percent of phenotypic variation (R^2) explained by the alleles at this locus accounts for approximately 29% of the total variation seen.

Conversely, only approximately 20cM away lays a second QTL (which will now be referred to as "Locus 2") located between Xtxp88 and Xtxp329. This locus has a

LOD score of 2.7 and an R^2 value of 18%. As opposed to the first QTL, this locus favorably affects protein digestibility. That is, an increase in favorable alleles at this locus serves to increase protein digestibility (decreases turbidity value).

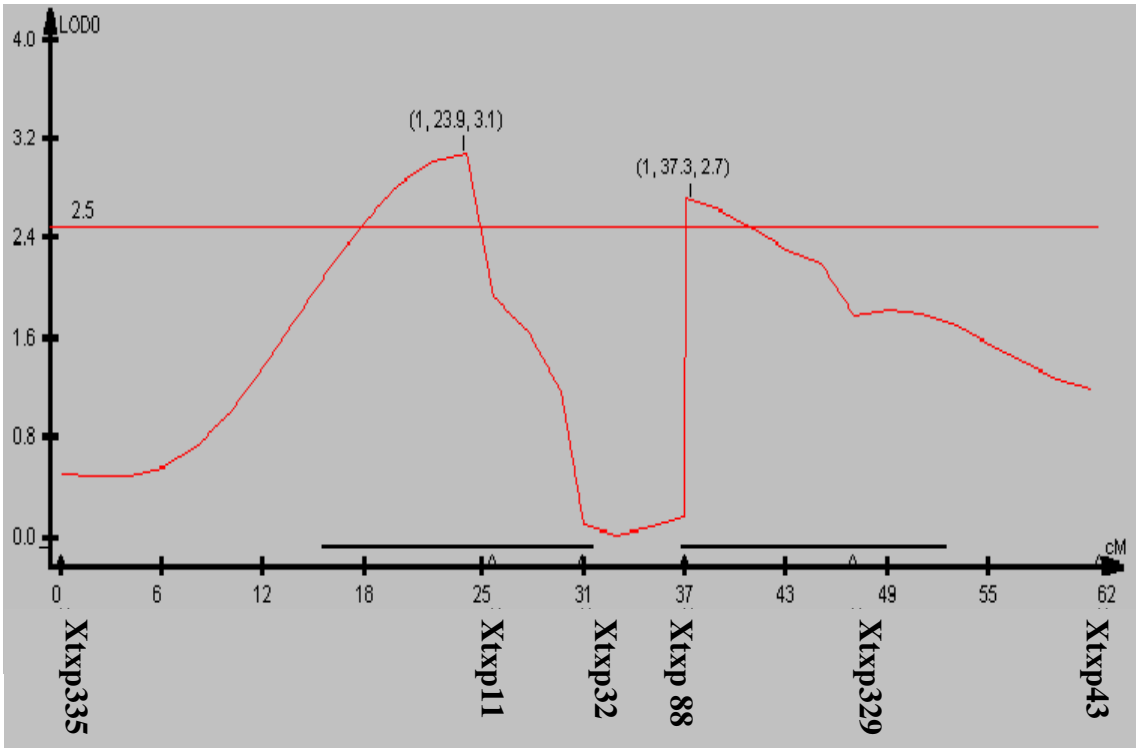


Fig. 6: QTL positions with LOD scores. Two QTLs were found to associate with high protein digestibility. The QTL on the left (Locus 1) contributes unfavorably to digestibility, while the QTL located on the right (Locus 2) contributes favorably. LOD scores over 2.5 were considered significant.

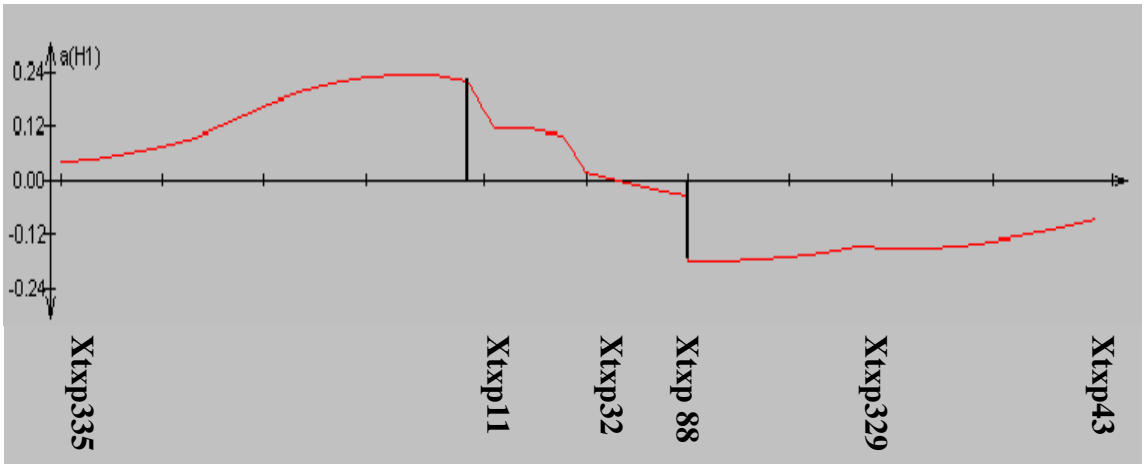


Fig. 7: Additive effect of QTLs. Units are phenotypic values of the turbidity assay. The QTL on the left decreases digestibility, thus it is shown to have a positive additive affect—it increases turbidity; the QTL on the right works conversely.

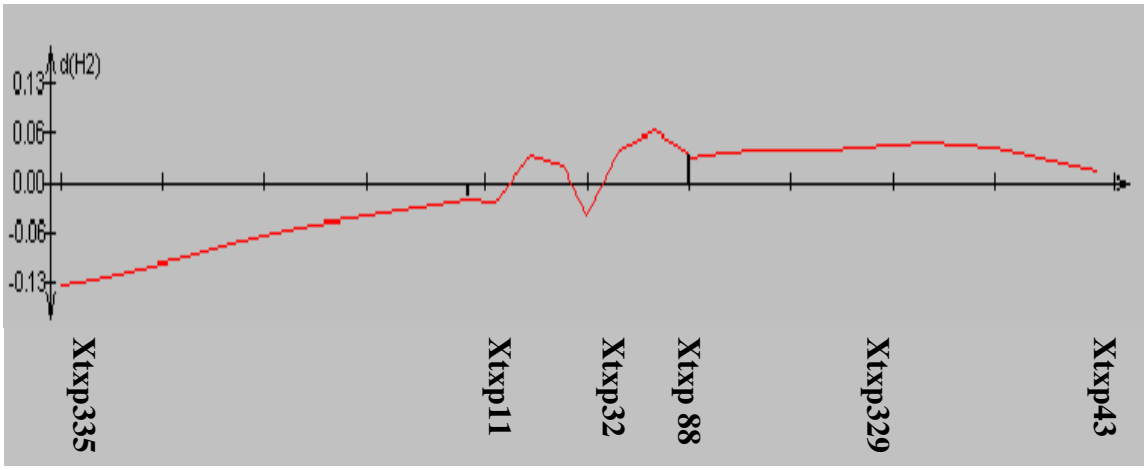


Fig. 8: Dominance effect of QTLs. Units are phenotypic values of the turbidity assay. The QTL on the left (Locus 1) has a stronger dominance effect than the QTL on the right (Locus 2). This leads to the left QTL contributing more to overall phenotypic variation.

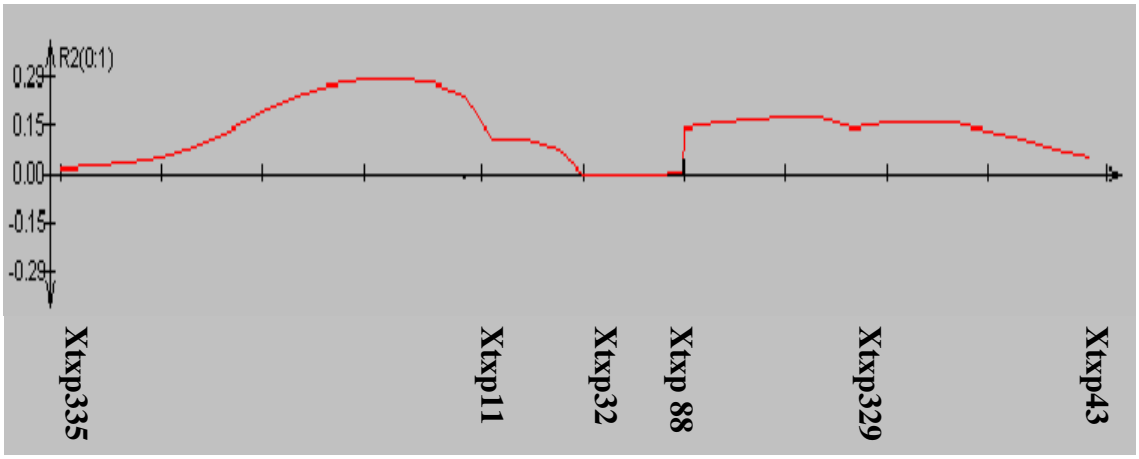


Fig. 9: R^2 for the QTLs. The QTL on the left (Locus 1) correlates stronger to phenotypic values—turbidity averages—than the QTL on the right (Locus 2).

Although two significant QTLs were found, no individual marker was found to be significant (Table 3).

Table 3: Summary of marker segregation. The number of individuals with informative results is given, along with the LOD score according to the QTL distribution (Fig. 6) and the probability that each marker segregates independently of the protein digestibility trait (using Mapmaker 3.0/QTL Cartographer Single Marker Analysis).

Chromosome 1

Name	n	LOD	pr(F)
Xtxp335	70	0.6	0.710
Xtxp11	69	2.1	0.144
Xtxp32	68	0.1	0.478
Xtxp88	65	1.5	0.656
Xtxp329	70	1.8	0.361
Xtxp43	64	1.2	0.521

A contrast analysis was calculated using the two markers segregating closest to the two QTLs—markers Xtxp11 and Xtxp329. In the analysis, the genotypic groups were labeled according to their alleles at Loci 1 and 2. For instance, "AB" indicates individuals in this genotypic group had the genotype of parental type A (the LD line Sureno) at Locus 1 and the genotype of parental type B (the HD line 9850029) at Locus 2. Genotypic groups included in the analysis were AA, AB, BA, and BB. The goal of the analysis was to determine whether the four genotypic groups were correlated with phenotypic value (turbidity average). The ANOVA (analysis of variance) in Table 4 indicates that there was a significant difference (α=0.05) in phenotypic values between at least two of the groups.

Table 4: Analysis of variance between the four phenotypic groups. There is a significant difference in phenotypic values between at least two groups.

ANOVA

Phenotypic Value

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.389	3	.130	2.809	.041
Within Groups	7.755	168	.046		
Total	8.144	171			

Next, two poc-hoc analyses, Fisher LSD and Tukey HSD (Table 5, 6), were used to calculate which of the phenotypes showed significant differences in phenotypic values. Phenotypic group BA showed significant differences with the other three groups, at a level of α=0.05.

Table 5: Fisher LSD analysis. Genotypic group BA is significantly different from the phenotypic values of the other three genotypic groups. "BA" represents individuals having parental type B (the HD line P850029) at Locus 1 and parental type A (the LD line Sureno) at Locus 2.

Multiple Comparisons

Dependent Variable: Phenotypic Value

LSD

(I) Genotype	(J) Genotype	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
AA	AB	.04679	.07804	.550	-.1073	.2009
	BA	-.19769*	.07468	.009	-.3451	-.0503
	BB	-.00434	.03743	.908	-.0782	.0696
AB	AA	-.04679	.07804	.550	-.2009	.1073
	BA	-.24448*	.09872	.014	-.4394	-.0496
	BB	-.05112	.07462	.494	-.1984	.0962
BA	AA	.19769*	.07468	.009	.0503	.3451
	AB	.24448*	.09872	.014	.0496	.4394
	BB	.19336*	.07110	.007	.0530	.3337
BB	AA	.00434	.03743	.908	-.0696	.0782
	AB	.05112	.07462	.494	-.0962	.1984
	BA	-.19336*	.07110	.007	-.3337	-.0530

*. The mean difference is significant at the .05 level.

Table 6: Tukey HSD analysis. Genotypic group BA is significantly different from the phenotypic values of the other three genotypic groups. "BA" represents individuals having parental type B (the HD line P850029) at Locus 1 and parental type A (the LD line Sureno) at Locus 2.

Multiple Comparisons

Dependent Variable: Phenotypic Value

Tukey HSD

(I) Genotype	(J) Genotype	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
AA	AB	.04679	.07804	.932	-.1557	.2493
	BA	-.19769*	.07468	.044	-.3915	-.0039
	BB	-.00434	.03743	.999	-.1015	.0928
AB	AA	-.04679	.07804	.932	-.2493	.1557
	BA	-.24448	.09872	.067	-.5006	.0117
	BB	-.05112	.07462	.903	-.2448	.1425
BA	AA	.19769*	.07468	.044	.0039	.3915
	AB	.24448	.09872	.067	-.0117	.5006
	BB	.19336*	.07110	.036	.0088	.3779
BB	AA	.00434	.03743	.999	-.0928	.1015
	AB	.05112	.07462	.903	-.1425	.2448
	BA	-.19336*	.07110	.036	-.3779	-.0088

*. The mean difference is significant at the .05 level.

The descriptive statistics of the genotypic groups and a visual representation of the average phenotypic values can be found below (Table 7, Fig. 10).

Table 7: Descriptive statistics of the genotypic groups. Genotypic group BA exhibits the highest turbidity average (lowest protein digestibility), while group AB exhibits the lowest turbidity average. "BA" represents individuals having parental type B (the HD line P850029) at Locus 1 and parental type A (the LD line Sureno) at Locus 2. Phenotypic values are represented by turbidity averages.

Descriptives

Phenotypic Value

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
AA	48	.5785	.19814	.02860	.5209	.6360	.17	.96
AB	9	.5317	.20453	.06818	.3745	.6889	.30	1.00
BA	10	.7762	.27981	.08848	.5760	.9763	.37	1.25
BB	105	.5828	.21640	.02112	.5409	.6247	.16	1.20
Total	172	.5902	.21823	.01664	.5573	.6230	.16	1.25

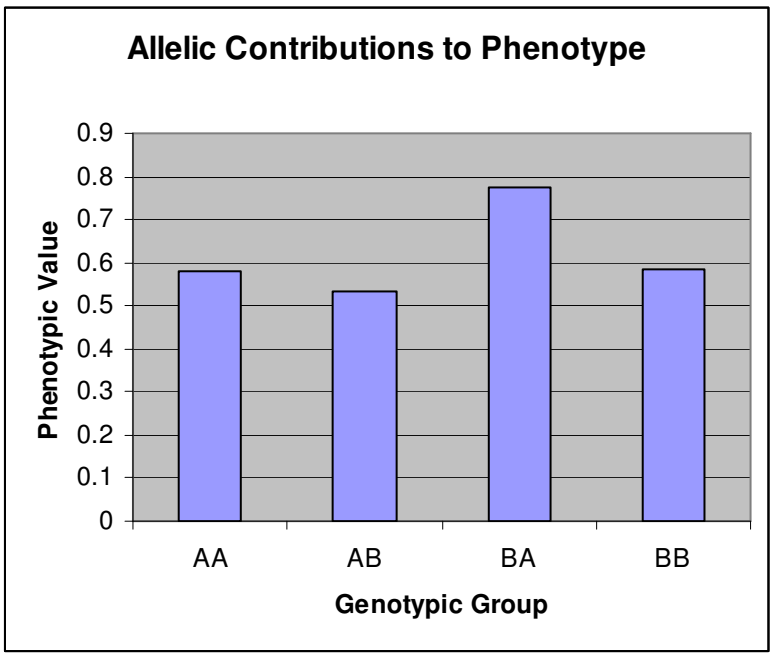


Fig. 10: Phenotypic averages of the genotypic groups. Genotypic group "BA" represents individuals having parental type B (the HD line P850029) at Locus 1 and parental type A (the LD line Sureno) at Locus 2. Phenotypic values are represented by turbidity averages.

The results of these analyses indicate that the highest protein digestibility is found in the AB genotypic group (that is, individuals with the parental type A allele at Locus 1 and parental type B allele at Locus 2), while the lowest is in the BA group. The favorable alleles at the two loci contributing to protein digestibility are segregating in repulsion in the parental lines. Furthermore, this explains why the two parental phenotypic values are not as different from each other as expected; each possesses favorable alleles at one locus and unfavorable alleles at the other. The two favorable alleles in repulsion also explains the transgressive segregation shown in the phenotypic values of the whole population. When recombination occurs to break the repulsion, phenotypic values can be expected to surpass either parent's value. A slight gain in digestibility may be obtained by breaking the linkage group in the parental lines to acquire favorable alleles segregating together. Whether the gain in digestibility will confer a biologically significant increase in digestibility must be evaluated within the context of a breeding program.

CONCLUSIONS

The turbidity assay data (Fig. 2) indicates that the phenotypic value of protein digestibility does not occur in a step-wise fashion, but rather as a gradient. When the broad distribution for phenotypes is viewed in light of the QTL analysis, it becomes evident that there are complex genetic interactions controlling protein digestibility, possibly with many genetic and molecular modifiers controlling phenotype. Furthermore, because the two significant QTLs are essentially working against each other in terms of overall effect on protein digestibility (as analyzed by the turbidity assay), it is likely that each locus has a greater influence on phenotypic variation than was calculated. Since the QTLs are opposed and closely linked, QTL Cartographer may have underestimated each of the QTLs' significances. Still, summing the two QTLs contributes to 47% of total phenotypic variation, which is a very reasonable percentage. According to Tanksley (1993), the average percent of phenotypic variation explained by QTLs in experimental studies is 30-40%.

It should be noted that, as with any genetic linkage map, the results are only applicable for the population and environment studied. It is possible that the QTLs identified are unique to the P850029 (HD) x Sureno (LD) crossed population. More research is necessary to determine if the QTLs identified in this study hold true for other sorghum lines.

Although protein digestibility was originally thought to be a simply-inherited, recessive trait controlled by a single gene, this is not likely the case—the gradient of

phenotypic values is probably caused by various alleles at multiple loci working favorably or unfavorably to influence protein digestibility. A hypothesis emerges that protein digestibility is controlled by two distinct, side-by-side regions on Chromosome 1—Locus 1 (an inhibitor of digestibility) and Locus 2 (a promoter of digestibility). Further research is necessary to determine how these two gene segments interact. (Note: the regions of DNA identified to influence protein digestibility shall be referred to simply as "gene segments" because it is unclear how many genes are interacting at these loci.) There are several possibilities about the types of products the segments encode. For instance, it is possible that the inhibitory gene region possesses an allele which alters the kafirins in a way that γ- and β-kafirins remain tightly bound to the periphery of the α-kafirins or to each other, allosterically blocking enzymatic digestion. Another theory is that the positively-acting locus could have an allele to modify the kafirin structures, preventing γ- and β-kafirin binding. Furthermore, the two distinct gene segments on Chromosome 1 could encode transcription factors to affect a third, unidentified gene segment, or they may code for a gene product in a pathway leading to increased/decreased protein digestibility.

There are clearly many possibilities as to how the gene segments affect protein digestibility (for which further research is needed), but the QTLs found in this study may in the meantime prove beneficial to breeders using MAS. Turbidity assays are inherently highly variable due to the nature of the procedure and machine. While a spectrophotometer is an excellent tool to use when measuring light absorbance, light scattering—which is the trait evaluated in a turbidity assay—is much more prone to

fluctuations in readings. Replacing the turbidity assay with marker-assisted selection will eliminate at least some of the error in the selection process, as molecular markers are a more reliable tool to correctly identify favorable alleles.

REFERENCES

Aboubacar, A., Axtell, John D., Nduulu, Lexington, and Hamaker, Bruce R. 2003. Turbidity Assay for Rapid and Efficient Identification of High Protein Digestibility Sorghum Lines. Cereal Chemistry 80(1): 40-44.

Axtell, J. D., Kirleis, A., Hassen, M.M., D'Croz-Mason, N., Mertz, E.T., Munck, L. 1981. Digestibility of sorghum proteins. Proceedings of the National Academy of Sciences of the United States of America 78: 1333-1335.

Bhattramakki, D., Dong, J., Chhabra, K., and Hart, G.E. 2000. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43: 988-1002.

Dellaporta, S. L., Wood, J., Hicks, J. B. 1983. Maize DNA minipreps. Maize Genetics Cooperation News Letter 57: 26-29.

Doggett, H. 1988. Sorghum. Longman Scientific and Technical. Wiley, New York.

Dowling, L. F., Arndt, C., and Hamaker, B.R. 2002. Economic viability of high digestibility sorghum as feed for market broilers. Agronomy Journal 94: 1050-1058.

Duodu, K. G., Taylor, J.R.N., Belton, P.S., and Hamaker, B.R. 2003. Factors affecting sorghum protein digestibility. Journal of Cereal Science 38: 117-131.

Hamaker, B. R., Mohamed, A.A., Habben, J.E., Huang, C.P., and Larkins, B.A. 1995. Efficient procedure for extracting maize and sorghum kernel proteins reveals higher prolamin contents than the conventional method. Cereal Chemistry 76: 583-588.

International Crops Research Institute of the Semi-Arid Tropics, 1996. The World Sorghum and Millet Economies: Facts, Trends and Outlook, ICRISAT/FAO. Patancheru, India/Rome, Italy: 1-2.

Klopfenstein, C. F., Hoseney, R.C. 1995. Nutritional properties of sorghum and the millets. Sorghum and Millets: Chemistry and Technology. D. A. V. Dendy, ed. St. Paul, MN, American Association of Cereal Chemists: 125-168.

Kosambi, D. D. 1944. The estimation of map distance from recombination values. Annals of Eugenics 12(3): 172-175.

MacLean, W. C., Lopez de Romana, G., Placko, R.P. and Graham, G.G. 1981. Protein quality and digestibility of sorghum in preschool children: Balance studies and plasma free amino acids. Journal of Nutrition 111: 1928-1936.

Michelmore, R. W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences 88: 9828-9832.

Mohan, D. P. 1975. Chemically induced high-lysine mutants in *Sorghum bicolor* (L.) Moench. Dissertation. W. Lafayette, IN, Purdue Univ.

Morgante, M. and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3(1): 175-182.

Murty, D. S., Kumar, K.A. 1995. Traditional uses of sorghum and millets. Sorghum and Millets: Chemistry and Technology. D. A. V. Dendy, ed. St. Paul, MN, American Association of Cereal Chemists: 185-221.

Oria, M. P., Hamaker, B.R., and Shull, J.M. 1995. Resistance of sorghum α-, β- and γ-kafirins to pepsin digestion. Journal of Agriculture and Food Chemistry 43: 2148-2152.

Oria, M. P., Hamaker, B.R., Axtell, J.D., and Huang, C.P. 2000. A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proceedings of the National Academy of Sciences 97: 5065-5070.

Shantharam, S. 1995. The importance of biosafety in the deployment of private transgenic sorghums in the environment. African Crop Science Journal 3: 131-133.

Shull, J. M., Watterson, J.J., and Kirleis, A.W. 1991. Proposed nomenclature for the alcohol soluble proteins (kafirins) of *Sorghum bicolor* (L.) Moench based on molecular weight, solubility, and structure. Journal of Agriculture and Food Chemistry 39: 83-87.

Shull, J. M., Watterson, J.J., and Kirleis, A.W. 1992. Purification and immunocytochemical localization of kafirins in *Sorghum bicolor* (L.) Moench endosperm. Protoplasma 171: 64-74.

Tanksley, S. D. 1993. Mapping polygenes. Annual Review of Genetics 27: 205-233.

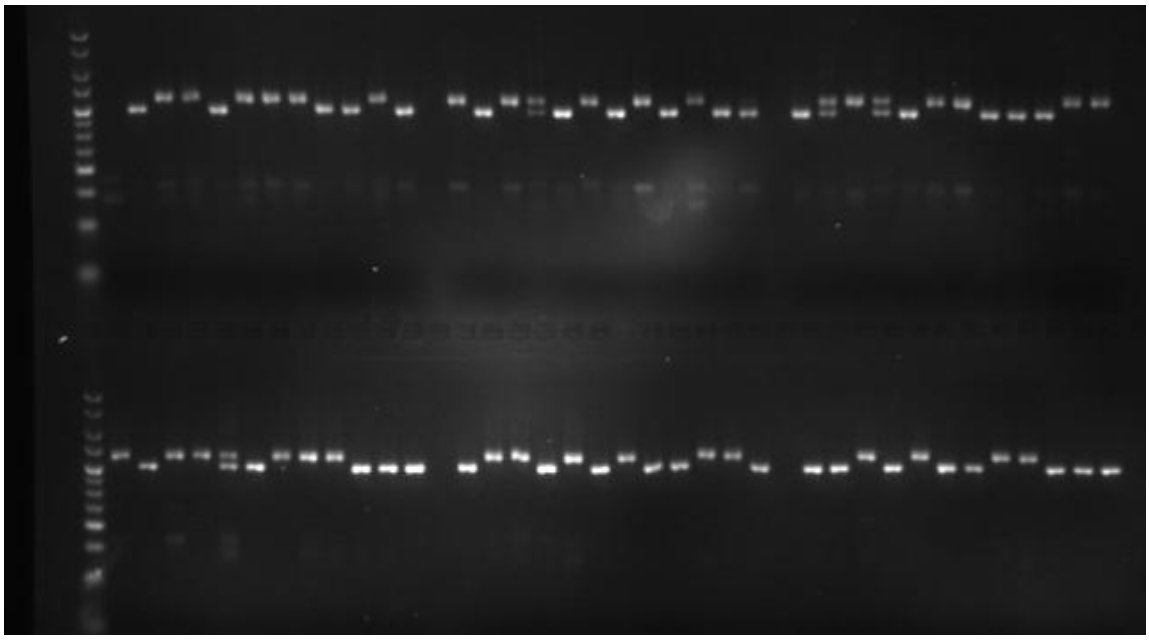
Taylor, J. R. N., Dewar, J., Taylor, J., Von Asheraden, R.F. 1997. Factors affecting the porridge-making quality of South African sorghums. Journal of the Science of Food and Agriculture(73): 464-470.

Weaver, C. A., Hamaker, B.R., and Axtell, J.D. 1998. Discovery of grain sorghum germ plasm with high uncooked and cooked in vitro protein digestibilities. Cereal Chemistry 75(5): 665-670.

Wu, Y. and Huang, Y. 2007. An SSR genetic map of *Sorghum bicolor* (L.) Moench and its comparison to a published genetic map. Genome 50: 84-89.

APPENDIX A

An example of an agarose gel used to determine genotypes of individual lines using SSR primers. Individuals were scored as being of the LD parental genotype, HD parental genotype, or heterozygous.



APPENDIX B

Primer information for the eight primers showing polymorphism between the parents and between the bulked segregant DNA groups.

Name	Oligo Sequence	Annealing Temperature (°C)
Xtxp88	GTT TTC CCA GTC ACG CGT GAA TCA GCG AGT GTT GG	53
Xtxp32	GTT TTC CCA GTC ACG AGA AAT TCA CCA TGC TGC AG	53
Xtxp43	GTT TTC CCA GTC ACG AGT CAC AGC ACA CTG CTT GTC	53
Xtxp11	GTT TTC CCA GTC ACG ACA AGC TCG AGA AAT TCA ACA TGC TG	55
Xtxp297	GTT TTC CCA GTC ACG ACA AGC GAC CCA TAT GTG GTT TAG TCG CAA AG	55
Xtxp329	GTT TTC CCA GTC ACG ACA AGC ACT AGG AAG GTG TTT AGT TTA AGG G	55
Xtxp335	GTT TTC CCA GTC ACG ACA AGC TAT TTC CTC TTG AAA GAA TCA GGG	55
Xtxp484	GTT TTC CCA GTC ACG ACA AGC CAA CCA CCC CCA ACT CTC T	55

VITA

Name:	Jennifer Ann Winn
Address:	TAMU 2474 College Station, TX 77843-2474
Email Address:	jenniewinn@hotmail.com
Education:	B.S., Biology, Texas A&M University, 2006. M.S., Plant Breeding, Texas A&M University, 2007.