THE FUTURE OF FOOD

Student Authors

Hailey Edmondson is a junior in plant genetics, breeding, and biotechnology at Purdue University. Her interest in agricultural biotechnology brought her from San Diego to Purdue University where she has researched sorghum protein digestibility in Dr. Cliff Weil's lab for the past two years.

Michael Busche is a senior in plant genetics, breeding, and biotechnology as well as biochemistry at Purdue University. He began researching sorghum protein digestibility in Dr. Cliff Weil's lab during his freshman year and has continued this work. Busche is currently assisting with Dr. Albrecht Melchinger's dihaploid corn breeding program in Stuttgart, Germany.

Mentor

Cliff Weil earned a BS in genetics from University of California, Davis in 1978, a PhD in genetics and development from Cornell, and completed postdoctoral scholarship at Ohio State. After his first faculty position at the University of Idaho, Weil moved to Purdue in 2001.

Moriah Massafaro earned her BS degree in biochemistry at Purdue University in 2014. Massafaro currently is working on her MS degree in agronomy with a focus on plant genetics and breeding. Her project involves researching increased protein digestibility in sorghum.

Abstract

The demand to produce high-yielding crops grows exponentially with the ever-expanding world population, and with this comes the obligation to examine the actual impact that these crops will have in areas with the most need. Sub-Saharan Africa, in particular, cannot support many of the otherwise staple agricultural crops due to climatic limitations. Sorghum, a heat-tolerant and drought-hearty grain source, is used as a primary cereal crop in regions of sub-Saharan Africa. However, sorghum grain is known to have a significantly lower protein digestibility rate compared to other common cereal grains. Cooked sorghum, in particular, has decreased levels of protein digestibility, rendering it less nutritionally effective, which is a serious concern in parts of the world that depend on sorghum grain as a primary source of protein and calories. The nutritional content that sorghum currently provides is inadequate, meaning that higher yields, though important, cannot be the sole aim of our agronomic research. Identifying and modifying the genes that influence sorghum protein digestibility would have a lasting impact on the millions of people who rely on sorghum nutritionally and economically. This study explored the genetic influences of sorghum proteins using a basic digestibility assay to screen thousands of sorghum mutants in search of those with high protein digestibility.

Edmondson, H., & Busche, M. (2015). The future of food: Identifying genetic mechanisms affecting sorghum protein Digestibility. *Journal of Purdue Undergraduate Research*, 5, 30–37. <http://dx.doi.org/10.5703/jpur.05.1.04>

Keywords

sorghum, genetics, protein digestibility

THE FUTURE OF FOOD:

Identifying Genetic Mechanisms Affecting Sorghum Protein Digestibility

Hailey Edmondson and Michael Busche, Agriculture

BACKGROUND

Sorghum is a cereal grain valued for its drought tolerance and heartiness in harsh environments, specifically in sub-Saharan Africa and parts of southern Asia. From 2003 to 2013, Africa was responsible for 40.7% of sorghum production worldwide (FAOSTAT, 2014). The African countries of Burkina Faso, Sudan, Nigeria, and Ethiopia collectively account for over 70% of the continent's sorghum production (Taylor, 2003). Due to sorghum's drought-stress and heat-stress tolerance, it is ideal for use in the unforgiving semiarid and subtropical climates of the African continent. Sorghum grain is often grown in regions of the continent too inhospitable and water-scarce to grow corn or rice, which are the more nutritious relatives of sorghum grain.

Many food-insecure populations of sub-Saharan Africa rely on sorghum as a primary source of calories and nutrients despite sorghum grain's nutritional shortcomings. Sorghum, like other cereal grains, is deficient in lysine, an essential amino acid; however, a more significant nutrient deficiency is protein digestibility, which is substantially less than that of other cereal grains and poses a nutrition concern. This characteristic creates a significant need for methods to increase protein digestibility of cooked sorghum grain. Various studies have confirmed this decreased protein digestibility, especially after cooking, in both in vivo and in vitro experiments (Duodu, Taylor, Belton, & Hamaker, 2003). In a study investigating cereal grain protein digestibility in vitro in maize, sorghum, barley, wheat, and rice, sorghum showed the lowest levels of pepsin digestibility in both cooked and uncooked grain (Hamaker, Kirleis, Butler, Axtell, & Mertz, 1987). A study of in vitro protein digestibility in sorghum showed a 24% digestibility decrease after cooking (Hamaker, Kirleis, Mertz, & Axtell, 1986).

While preparation and processing techniques can improve sorghum for human use, sorghum genetics is also being investigated as a means of finding lines with superior nutritional properties. In 1975, the mutant sorghum line P721Q, derived by mutagenesis of the existing P721N line, showed increased lysine content, as well as increased protein content (Mohan, 1975). This high-digestibility mutant is the subject of various research projects, with goals including the observation of protein body structure, comparison of in vitro digestibility with other lines, and the determination of the genetic mechanisms responsible for its unique digestibility characteristics. The research discussed in this paper seeks to identify genetic mechanisms causing this advantageous trait in the P721Q mutant using gene mapping techniques.

Several hypotheses exist regarding the factors that cause decreased protein digestibility levels in sorghum. Protein body structure has been hypothesized to affect digestibility levels in sorghum. These protein bodies are composed primarily of prolamins, a class of grain proteins that comprise the majority of protein content in cereals. Different prolamins exist in different grains; kafirins are the prolamin proteins specific to sorghum. Kafirin storage bodies "comprise 70% of whole grain flour protein" (Hamaker & Bugusu, 2003, p. 1). These kafirin protein bodies can be further divided into three subunits—alpha-, beta-, and gamma-kafirins—each with unique size, structure, and solubility qualities. Alpha-kafirins are the largest class, making up 80% of total kafirin, and are thus the "principal storage protein of sorghum" (Hamaker & Bugusu,

Figure 1 (Above). Sorghum is an important cereal grain crop in arid and semiarid climates like sub-Saharan Africa and southern Asia.

2003, p. 2). Gamma- and beta-kafirins compose 15% and 5% of kafirin content, respectively. The properties of these storage proteins and respective protein bodies are hypothesized to affect sorghum digestibility properties.

Observation of protein body structure in highly digestible lines reveals a unique folding pattern unlike other varieties. Using transmission electron microscopy, P851171—a high-digestibility sorghum line derived from the high-lysine mutant P721Q—appeared to have protein bodies with unique membrane folding and invaginations (Oria, Hamaker, Axtell, & Huang, 2000). This unique protein body microstructure was observed in both P721Q and a highly digestible line derived from P721Q, and had not been reported in any sorghum cultivars prior (see Figure 2 and Figure 3). These findings suggest a possible correlation between protein body characteristics and digestibility properties. Determining whether highly digestible F2 individuals in a mapping population all share the same

mutant protein structure will provide further evidence as to whether or not there is a correlation between these two sorghum seed traits.

P721Q High-Lysine Sorghum Line

P721Q is a high-lysine line of sorghum derived from seeds of the normal cultivar—P721N—that received ethyl-methanesulfonate (EMS) mutagenesis treatment. One particular mutant, P721Q, exhibited higher lysine content, as well as increased protein digestibility, compared to P721N. This increase in lysine content is due to “an increase in lysine-rich non-kafirins” and a “decrease in lysine-deficient kafirins” (Wu, Yuan, Guo, Holding, & Messing, 2013, p. 2). The kernel of this particular mutant has a floury, soft seed texture, which is not ideal for cooking applications and mold resistance. However, the novel nutritional properties of P721Q have inspired various research efforts aimed toward understanding and improving this superior line.

Figure 2. Micrograph images taken of P721N and P851171 illustrate the differences in protein bodies between the wild-type lines and the highly digestible mutant line derived from P721Q. Image A shows the protein bodies at low magnification within P851171. Images B and C contrast the protein folding structures of highly digestible P851171 and wild-type P721N, respectively (Oria et al., 2000). P851171 shows increases in folding and invaginations compared to P721N.

Figure 3. Micrograph image of the misfolded protein bodies in P721Q endosperm.

One breeding effort used P721Q and derived lines in order to develop high-lysine hybrids with higher yield and kernel weight (Monyo et al., 1988). In this particular case, P721Q was crossed with existing, high-yielding sorghum lines in order to find a combination expressing high-lysine, high yield, and a heartier endosperm. The best hybrid from this project resulted in a 35% lysine increase compared to the normal P721N cultivar and had a substantially higher yield than P721Q (Monyo, Ejeta, Nyquist, & Axtell, 1988).

The protein bodies within P721Q and derived lines have uniquely folded structures, hypothesized to allow enzymes greater access to the protein bodies, which results in increased digestion (Oria et al., 2000). In a recent study, the mutant kafirin gene from the P721Q mutant was transformed into maize in order to identify mutations affecting kafirin genes. The floury P721Q sorghum and the maize mutant *o2* share many phenotypic characteristics, including a softer seed texture and invaginated protein bodies (Wu et al., 2013). In maize, this phenotype is a result of mutations in the endoplasmic reticulum signal peptides responsible for indicating the proper folding and processing of proteins after synthesis. Using a simple sequence repeat (SSR) molecular marker to follow the prolamin mutation in P721Q and transgenic maize seeds with mutant kafirin genes, this study identified a mutant-specific A to G missense mutation in the kafirin gene, resulting in the substitution of the amino acid alanine for the amino acid threonine at the signal peptide cleavage site (Wu et al., 2013). The study concluded that the mutated protein bodies observed in the high-digestibility mutant were a result of this altered signal peptide, critical in the proper folding and processing of proteins in the endoplasmic reticulum. However, digestibility was not assessed and it is unclear whether or not mutant protein bodies are the sole contributing factor to increased digestibility in sorghum. The sorghum digestibility research described in this paper aims to identify the mutations contributing specifically to digestibility, which is a valuable trait with potential to improve sorghum nutrition for human use.

Methods for Genetic Mapping

Past research has successfully identified and characterized possible aspects of sorghum grain affecting digestibility properties, including protein storage structure and tannin content. The genetic mechanisms of these digestibility properties, however, are not yet known. The research described in this paper is currently working toward identifying loci responsible for high digestibility in mutant P721Q via genetic mapping.

Genetic mapping is the illustration of the position of genes in a genome and their location relative to other loci. Association mapping is a method of genetic mapping that

utilizes molecular markers, usually in the form of single nucleotide polymorphisms (SNPs), present at frequent intervals throughout a genome of interest. This technique takes advantage of the natural crossover that occurs during meiosis to determine the location of a gene of interest. The rate of crossover decreases the closer a molecular marker is to a gene. This is simply because the odds of a crossover instance occurring between two locations that are very close to one another on a chromosome are less likely than a crossover occurring between two locations farther apart on a chromosome.

As of now, research has been conducted to determine that changes in kafirin genes result in changed protein structure. By observing the effect of the mutant kafirin gene in maize, it has been determined that changes in the prolamin structural genes indeed affect protein body structure. However, it has yet to be determined that the change in protein structure accounts for the increased protein digestibility levels in high-digestibility lines like P721Q. In comparing Tx623—a previously sequenced line of sorghum with no known high-digestibility properties—to P721Q through this gene mapping project, possible causal mutations related to increased digestibility in P721Q may be identified, which will assist in the identification of genes related to this trait.

Variations of association mapping and the use of molecular markers are often applied to plant systems. The sorghum digestibility research described in this paper utilizes molecular markers and association mapping techniques in order to identify quantitative trait loci (QTL) responsible for levels of increased protein digestibility in mutant line P721Q.

In order to determine the causal mutations in the P721Q line contributing to increased digestibility, it is compared to an existing line with a previously sequenced genome. The sorghum line Tx623 was selected for use in this project because its sequence is known, which allows for easy comparison to P721Q. One aspect of the project is mapping using sequence data collected from the high-digestibility homozygotes in the F2 population resulting from Tx623 and P721Q parental line crosses.

Obtaining an F2 mapping population with high-digestibility homozygotes requires several steps. Initially, a cross is performed between the two parental lines—Tx623 and P721Q. The F1 population resulting from this cross will be heterozygous at regions controlling protein digestibility. Although the exact location and identity of these regions in the genome have yet to be characterized, it is assumed that the mutations controlling digestibility are recessive. The specific heritability characteristics can be confirmed through further testing. The F1 population, which is heterozygous for high protein digestibility, is

Figure 4. Diagram of bulked segregant sequencing. Blue indicates a mutation that is causal for the mutant phenotype. Crosses between wild-type and mutant lines lead to genetic crossover, which allows tracing of the causal mutations through the F₂ generation where individuals of interest can be selected for based on phenotype.

then self-pollinated. This will result in a 1:2:1 genotypic ratio. Therefore, 25% of the F₂ population should be homozygous for increased protein digestibility based on this segregation pattern. This method is known as bulked segregant analysis and utilizes gene mapping techniques (see Figure 4).

In order to reach the point of sequencing and mapping the high-digestibility F₂ individuals to find gene candidates responsible for increased digestibility characteristics, all F₂ individuals must be screened with a protein digestibility assay in order to identify individuals with high digestibility. If all three replicates of a given sample tested with the assay show high levels of digestibility, that individual is presumably homozygous at loci responsible for aspects of protein digestibility. After selecting for digestibility with the assay, individuals observed as having high digestibility will be sequenced and compared to each other in order to identify peaks of homozygosity within the mapping population.

MATERIALS AND METHODS

Before determining the mutations contributing to protein digestibility, high-digestibility individuals must be identified. Screening for this high-digestibility population is completed in the lab using a two-part assay (Aboubacar,

Axtell, Nduulu, & Hamaker, 2002). This digestibility assay is designed to mimic the preparation and digestion of cooked sorghum as it would occur in the human body. Initially, water and sorghum seed are ground into a paste and cooked in an oven, emulating culinary preparation practices commonly used in human populations. This aspect is crucial to digestibility levels because sorghum is observed to have a much lower rate of protein digestibility after being cooked. The cooked paste is then divided in half—one serves as the experimental sample while the other remains untreated and serves as the control. The experimental samples are treated with 290 μ L of a pepsin solution containing 0.9 grams of pepsin and 30 mL of 0.5M KH_2PO_4 and are then incubated. The addition of pepsin, a digestive enzyme that degrades proteins and is present in the human stomach, simulates the digestion that would occur in a human system. Samples are removed from incubation after one hour and the digestion reaction is stopped by the addition of 2N NaOH. After this digestion, the proteins remaining in each tube are resuspended through a series of centrifugation and vortex treatments after being washed with 0.1M KH_2PO_4 (pH 7.0) and ddH₂O. Following this series of washes, extraction buffer is added to each sample prior to an hour-long incubation period at 25°C in a shaking incubator at 250 rpm. The addition of an 11% TCA solution is used to precipitate the remaining proteins out of solution in order to be assessed

Figure 5. Scatter plot of the adjusted, average digestibility values by their rank from low to high. The average of three replicates from the digestibility assay is normalized by dividing by the high control for each plate. This allowed us to make a cut-off of 5% for the calling of the highs for bulk segregant analysis.

using spectrophotometry. The protein levels in each sample are then compared to one another and assessed using a turbidity test. Data collected from these turbidity tests is analyzed with a program to identify samples with digestibility levels significant enough to be considered highly digestible.

A standard microtiter plate spectrometer at a 562 nm absorbance level is used to perform the turbidity analysis, resulting in two absorbance values—experimental and control—from each sorghum sample. The program first normalizes the raw data values by subtracting absorbance values from the blank controls included in the original microtiter plate. The normalized absorbance value of the experimental sample is subtracted from that of the control sample. This difference is then divided by the normalized control sample's absorbance, resulting in the percent protein digestion for a given sample.

RESULTS AND DISCUSSION

The portion of this project that has been completed includes screening approximately 1,000 F2 individuals for digestibility levels using the two-part digestion assay and turbidity test. Samples within the highest 5% of

protein digestibility will be rescreened to confirm high-digestibility levels (see Figures 5 and 6). Once confirmed, the group of samples with the highest digestibility will serve as the mapping population for further investigation of gene candidates responsible for increased digestibility properties.

The assay is sensitive to changes in storage of samples, temperature, time between washes, and completeness of resuspension. Refining the protein digestibility assay and achieving consistent results was a major aspect of this project. Several iterations of the protocol were made throughout the screening process and are detailed here.

Storing samples overnight at a low temperature ensured that contaminating microbes did not continue to digest the protein present in the samples. This alone made no noticeable change in the digestibility values. Preparing the extraction buffer the day that it is used prevents the solution from losing potency. The odor associated with 2-mercaptoethanol seems to fade from the buffer after a period of time, indicating a change that may be affecting its mode of action. This change prevented the buffer itself from precipitating out of solution in the TCA/buffer mix used as a spectrophotometric blank.

Figure 6. The scatter plot of adjusted digestibility readings for F2 population individuals is a visual representation of the distribution of protein digestibility values and the top 5% of samples, which is illustrated by orange plot points.

Cold TCA has been used to precipitate biological molecules like protein and nucleic acids out of solution. The goal of this assay is to isolate only proteins remaining in the sample, so using TCA chilled to 4°C may be too aggressive. Using room temperature TCA throughout the experiment also caused a measureable decrease in the turbidity of the TCA/buffer solution when mixed. This produced reliable values for the blank control sample, allowing for a more accurate normalization of the raw data.

After spinning the samples at 15,000 rpm for 10 minutes, many samples retain a cloudy band of precipitate in the supernatant that does not pellet out. Traditionally, this was avoided and only the very top of the supernatant was drawn off for further use. It is possible that this precipitate contains the proteins in question, so homogenizing is essential for an accurate measurement of undigested protein. Extra caution is taken not to disturb the pellet during this process. This 100 µl of supernatant is spun down again in the case that any unwanted contaminants remain. This had no clear effect on the digestion values.

The adjustments in the TCA preparation and concentration proved to be the most influential changes to the protocol. The buffer continually precipitated out of

solution when the TCA was simply added and mixed with pipetting. When the solution was made with vigorous mixing, however, it retained homogeneity for much longer. Adjustments in the concentration caused significant changes in the average digestion value of the controls. It was found that the amount of TCA present in the microtiter plate negatively correlates with the percent digestion value of the sample. Controls that previously resulted in digestibility values of around 95% are now producing values closer to 20%.

Reading the plate at 15 minutes after mixing resulted in the highest measurements in digestibility. Plates read earlier than 15 minutes produce relatively lower absorbance values and, consequently, poor digestibility. These plates were not stored at 4°C during this time in order to prevent the TCA in the blank from precipitating out of the buffer.

The aforementioned changes to the procedure have resulted in more consistent readings from all samples tested; however, much of the sensitivity of the assay has been lost, with samples that formerly appeared to have high digestibility now producing low values very similar to the unmutagenized controls. Currently, the concentration of the digestive enzyme used, pepsin, is being

adjusted in an attempt to increase sensitivity. Adjusting this concentration or the time allowed for digestion should help in separating samples that are truly highly digestible from those that are not initially, but can be digested after heavy exposure to the digestive environment. Once reliable digestion values can be obtained, samples that were thought to be highly digestible will be rescreened for verification. Additionally, there will be further screening of the P721Q x Tx623 crosses, and mining the current database of sequence data for possible gene candidates.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Cliff Weil, Moriah Massafaro, and Phillip Rushton.

REFERENCES

- Aboubacar A., Axtell, J. D., Nduulu, L., & Hamaker, B. R. (2002). Turbidity assay for rapid and efficient identification of high protein digestibility sorghum lines. *Cereal Chemistry*, 80(1), 40–44. <http://dx.doi.org/10.1094/CCEM.2003.80.1.40>
- Duodu, K. G., Taylor, J. R. N., Belton, P. S., & Hamaker, B. R. (2003). Factors affecting sorghum protein digestibility. *Journal of Cereal Science*, 38(2), 117–131. [http://dx.doi.org/10.1016/S0733-5210\(03\)00016-X](http://dx.doi.org/10.1016/S0733-5210(03)00016-X)
- FAOSTAT. (2014). Production quantities by country, average 1993–2013. Retrieved from <http://faostat3.fao.org/browse/Q/QC/E>
- Hamaker, B. R., Kirleis, A. W., Mertz, E. T., & Axtell, J. D. (1986). Effect of cooking on the protein profiles and in vitro digestibility of sorghum and maize. *Journal of Agricultural Food Chemistry*, 34(4), 647–649. <http://dx.doi.org/10.1021/jf00070a014>
- Hamaker, B. R., Kirleis, A. W., Butler, L. G., Axtell, J. D., & Mertz, E. T. (1987). Improving the in vitro digestibility of sorghum with reducing agents. *Proceedings of the National Academy of Science*, 84(3), 626–628. <http://dx.doi.org/10.1073/pnas.84.3.626>
- Hamaker, B. R., & Bugusu, B. A. (2003). Overview: Sorghum proteins and food quality. In *Workshop on proteins of sorghum and millets: enhancing nutritional and functional properties for Africa [CD]* (Pretoria: South Africa).
- Mohan, D. P. (1975). *Mutants in Sorghum Bicolor (L. Moench)* (Doctoral dissertation). Purdue University, West Lafayette.
- Monyo, E. S., Ejeta, G., Nyquist, W. E., & Axtell, J. D. (1988). Combining ability of high lysine sorghum derived from P-721 opaque. *Crop Science*, 28(1), 70–75. <http://dx.doi.org/10.2135/cropsci1988.0011183X002800010017x>
- Oria, M. P., Hamaker, B. R., Axtell, J. D., & Huang, C. (2000). A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. *Proceedings of the National Academy of Sciences*, 97(10), 5065–5070. <http://dx.doi.org/10.1073/pnas.080076297>
- Taylor, J. R. N. (2003). Overview: Importance of sorghum in Africa. *Afripro: Workshop on the Proteins of Sorghum and Millets: Enhancing Nutritional and Functional Properties for Africa*, Pretoria, 2(4).
- Wu, Y., Yuan, L., Guo, X., Holding, D. R., & Messing, J. (2013). Mutation in the seed storage protein kafirin creates high-value food trait in sorghum. *Nature Communications*, 4, 1–7. <http://dx.doi.org/10.1038/ncomms3217>