Advances in the Study of Endosperm Vitreousness in Silage Maize and Identification of Quantitative Trait Loci Using Novel PhenotypingTechniques

By

José Ignacio Varela

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Plant Breeding and Plant Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 06/22/2022

The dissertation is approved by the following members of the Final Oral Committee:

Natalia de Leon, Professor, Agronomy
Shawn Kaeppler, Professor, Agronomy
William Tracy, Professor, Agronomy
Edgar Spalding, Professor, Botany
Jeff Endelman, Associate Professor, Horticulture

Dedication

This work is dedicated to my father Juan Pablo and my mother Maria del Carmen. Since I was young they told me the very few important things in life: hard work and respect for people around you. Both of them represent an example of extraordinary persons and I wish to pass their values as legacy to my progeny.

Acknowledgements

First and foremost, to my best friend and life partner Valentina. Her company and support at every moment made this journey possible and enjoyable.

Thanks to all members of my committe: Bill Tracy, Jeff Endelman, Edgar Spalding and Shawn Kaeppler for their support and guidance.

Thank you Natalia de León for your mentorship all these years. Your spirit, attitude towards life, and generosity made my period as student and mentee an enriching and joyfull experience. Your inconditional support made this possible. I will be eternaly gratefull for the guidance you gave me every time that I needed.

Abstract

Maize (*Zea mays* L.) is a crop of significant economic and agronomic importance that has many uses as food and livestock feed. These applications rely heavily on high levels of productivity and optimal quality of the mature kernel. Endosperm composition and texture dictates the quality grading of most maize varieties targeted to the food and feed markets. Endosperm vitreousness is particularly important in the production of maize silage as it influences starch degradability and therefore directly affects the nutritive value of the feed. Much of the knowledge of the genetics and biochemistry of endosperm vitreousness has resulted from research conducted on classic endosperm mutants. Significantly less emphasis has been placed on the characterization of natural variation available for this trait at a population level. One of the limitations has been the lack of reliable high throughput methods to measure this trait for large experiments.

This dissertation contains a literature review and three research projects designed to advance the current knowledge of maize endosperm vitreousness and related traits focused in three areas: phenotyping, effects of ensiling, and genetic dissection. First, we present a novel, non-destructive high throughput method to estimate endosperm vitreousness using Near Infrared Spectroscopy in a hyperspectral flatbed scanner. Results show that applying hyperspectral imaging of intact kernels can predict endosperm vitreouesness, total kernel protein, total kernel density and mass in a high throughput manner without the need for sample destruction nor extensive sample preparation. The system allows also to apply computer vision algorithms to extract physical characteristics of samples such as kernel length and width. Second, we performed an evaluation of the influence of different levels of endosperm vitreousness and α-zeins in the nutritional value, fermentation profile and starch digestibility of whole plant maize silage during ensiling. We found that α-zeins degrades over time during ensiling and starch digestibility increases, probably due to the

degradation of the α-zeins that are part of the starch-protein matrix. Third, we performed genetic mapping of endosperm vitreousness applying genome wide association studies to a diversity panel and also linkage mapping on a six parent multiparent population. We identified several genomic regions associated with endosperm vitreousness, protein, density and mass. Overall, results from these projects provide advances for important maize traits characterization and insights into the genetic control of key maize endosperm traits as well as practical considerations of the impact of endosperm vitreousness in maize silage.

Table of Contents

List of Figures

List of Tables

List of Supplemental Figures

List of Supplemental Tables

Chapter 1 Review of Kernel Vitrousness in Maize

Introduction

Maize silage is a fundamental component of United States agriculture as it is the primary source of feed for dairy cattle with more than 6,4 million acres harvested and more than 130 million tons produced in 2021 (USDA National Statistics Service, 2022). Maize prevalence over other feed crops is due to two main factors: it produces a high energy density feed and yields more Dry Matter (DM) per hectare than any other forage in a single harvest (Cueva et al., 2021). Maize silage includes grain and stover (non-grain portion) of the plant. Whole-plant yields of silage maize have increased dramatically since the early part of the twentieth century. During this period, whole-plant yield in the U.S has increased at a linear rate of approximately 0.09 Mg/ha/yr on a DM basis (Coors and Lauer, 2001). Advances in fertilization, weed management, mechanization and planting densities strategies are some of the most impactful practices that are in continuous improvement to exploit the genetic potential of each new hybrid releases.

As a fermented feed source, maize also provides the three main characteristics of plant materials necessary to ensure good quality silage: (1) adequate level of fermentable substrate, (2) a relatively low buffering capacity, and (3) a dry matter concentration greater than 200 g/kg (McDonald et al., 1991). These characteristics in combination with anaerobic storage conditions promote effective fermentation. Compared with other forages, maize silage has relatively high fiber concentration, as measured by neutral and acid detergent fiber (NDF and ADF, respectively), but low lignin concentrations, leading to low lignin/NDF or lignin/ADF ratios (Harlan et al., 1991). In contrast to most other forages, maize has consistent nutritive value over a longer harvest period. Part of that consistency results from the highly digestible grain, which

compensates for the decline in quality of the stover fraction as the plant develops and matures (Coors and Lauer, 2001).

The characteristic high Harvest Index (HI) of maize, defined as kilograms of grain divided by the total kilograms of above ground biomass (stover plus grain) is a key characteristic of silage as the kernel portion is the main source of digestible energy. The endosperm, which accounts for approximately 75%-80% of the maize kernel by weight, is the primary source of digestible energy and is composed of starch (50%-70%), proteins (6-15%), and small amounts of fat (such as phospholipids) and ash.

The relatively high proportion of starch contained in the endosperm of kernels on a silage mix, does not ensure by itself an effective energy delivery to the animal, as starch digestibility inside the rumen (largest stomach compartment in ruminants) also depends on other endosperm property including hardness, which is commonly referred to as vitreousness by plant breeders and animal scientists. Vitreousness is an endosperm quality descriptor whose name arises because of the optical properties associated with different endosperm types. Hard endosperm type maize has a dense, glass-type appearance that is translucent to light, whereas soft endosperm generally has low density and a floury/chalky appearance, opaque (not-translucent) to visible light. Several studies have evaluated the relationship between endosperm vitreousness and *in situ* starch or DM degradability (Philippeau and Michalet-Doreau, 1998; Correa et al., 2002; Ngonyamo-Majee et al., 2008). Such studies have observed a strong negative relationship between endosperm vitreousness and *in situ* or DM degradability, meaning that as the endosperm vitreousness increases *in situ* starch or DM degradability decreases.

Vitreous endosperm is also important for resistance to insect and fungal damage, resilience during harvest and storage, and numerous other end use characteristics. The purpose of this

review is to outline the current knowledge of maize endosperm vitreousness biochemistry and genetics, to investigate the relationship of this trait with ruminal nutrition, and to present and discuss the approaches currently available to measure this trait.

Biochemistry and genetics of the vitreous endosperm

Endosperm is a product of double fertilization in the female gametophyte (embryo sac), and it functions as a nutritive organ to support embryogenesis and seedling development (Olsen and Becraft, 2013; Olsen, 2004). In maize, the endosperm initially differentiates into four main cell types, which are termed starchy endosperm (SE), aleurone (AL), embryo-surrounding region (ESR), and basal endosperm transfer layer (BETL) (Becraft and Gutierrez-Marcos, 2012). Each cell type has unique morphological and functional properties. The BETL mediates transport of nutrients into the kernel, AL is the epidermal cell layer that synthesizes hydrolases to mobilize the starch and storage proteins to support seedling establishment during germination, and SE is the central and largest portion of the endosperm which accumulates starch and storage proteins (Becraft and Gutierrez-Marcos, 2012).

Storage proteins play a central role in the formation of the vitreous endosperm. Granular protein inclusions in the maize aleurone and peripheral endosperm cells were observed as early as 1885 (Harz, 1885) and later by Weatherwax (1930) who noted the appearance and growth of protein granules in the inner endosperm cells, though to a lesser extent than in the peripheral cells. In his 1955 manuscript, Donald Duvick presented a detailed account of the appearance and growth of endosperm cytoplasmic inclusions. He described the presence of "starch grains" and "protein granules" (Duvick, 1955) and characterized the gradient of these two from the aleurone layer to the central endosperm. Early observations suggested that starch grains were larger and more

numerous in the central endosperm cells and protein granules were larger and more numerous in the peripheral layers (Duvick, 1955) as depicted in Figure 1-1. These factors were later proposed to be central to the formation of the "vitreous" outer endosperm and the soft, floury central endosperm (Duvick, 1961). Later observations of the outer starchy endosperm cell layers showed that there is an increase in the size of protein bodies from the first to fifth subaleurone cell layer that reflects the developmental gradient of these cells (Lending and Larkins, 1989). The "protein granules" and "protein bodies" (PBs) are interchangeable; the former was preferred by early to mid 20th authors and the latter was coined later and it's the one used currently.

Duvick (1961) suggests that the proportion of protein to starch granules affects the viscous cytoplasm which when desiccated will give rise to a rigid matrix. This ratio is the primary factor responsible for the abrupt partition between vitreous and floury endosperm. Moving towards the center of the kernel and out of this range, with progressively smaller protein granules and larger starch granules, the volume occupied by the cytoplasm is insufficient to prevent shattering of the matrix during kernel desiccation, giving rise to the floury appearance.

Early studies provided evidence that PBs in the maize kernel are largely made up of prolamins proteins (Duvick, 1961) and vitreous endosperm can have more than double the percentage of prolamins compared to floury endosperm (Hamilton et al., 1951). These and other early observations of the structure and composition of the maize kernel laid the groundwork for subsequent studies that have characterized its nutritional and physical properties. PBs appear to be central to the generation of hard kernels (Holding and Larkins, 2006)(Figure 1-2).

There are two predominant types of storage proteins in maize kernel: prolamins and globulins. The maize embryo contains about 10% proteins which are mostly globulins, and the endosperm, accumulates primarily prolamin proteins (Coleman and Larkins, 1999). Prolamins are endosperm

storage proteins that are high in proline (amino acid) and are found in the seeds of all cereal grains. Proline is a highly hydrophobic amino acid capable of forming complex protein structures and thus proteins with high proline contents develop tertiary structures that are intensely hydrophobic and are soluble in aqueous alcohol solutions (Lasztity, 1984; Momany et al., 2006), specifically all prolamins are soluble in 70% ethanol. Prolamins for each cereal grain have specific and historically determined names: gliadin (wheat), hordein (barley), secalin (rye), zein (maize), kafirin (sorghum) and avenin (oats). Within a maize context, the word "prolamin" and "zein" are used intercheangeable in the literature. Throughout the manuscrip these type of proteins will be referred as zeins.

There are differences in aqueous solubility and ability to form disulfide interactions, and the current nomenclature separates zeins in α-, β-, γ-, δ- types based on these properties (Esen, 1987; Coleman and Larkins, 1999).

In the maize kernel, α-zeins resolve at 19-kDa and 22-kDa on protein separation by electrophoresis using SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and constitute 60 to 70% of the total zeins, β-zeins resolves at 15-kDa and constitute 5 to 10% of the total zeins, γ-zeins resolve at 16, 27 and 50-kDa and constitutes 20-25% of the total zeins and δ-zeins resolve at 10-kDa and 18-kDa and make up less than 5% of the total zeins (Wu et al., 2010).

The formation of PBs structures is a complex process where specific interactions between the different types of zeins influence the formation of discrete PBs accretions.

Inmunolocalization with specific zein antibodies determined that young meristematic cells closest to the aleurone layer contain small PBs that are composed entirely of β- and γ-zein (Lending and Larkins, 1989; Lopes and Larkins, 1991). More mature cells farther away from the

aleurone layer contains larger PBs with central locules composed primarily of α-zeins and small amounts of δ-zeins (Figure 1-3). This suggests that these proteins penetrate the matrix of cysteine-rich and cross-linked β- and γ-zeins (Holding and Larkins 2006). In the largest PBs (1-2 μm), these locules fuse into a solid core. In these PBs, a thin layer of β- and γ-zein completely surrounds the protein body, although these proteins can also be found internally as small central inclusions or stands (Figure 1-3) (Holding and Larkins, 2006).

Genetic control of zeins

Regions of genome found to encode α-, β-, γ-, and δ- zeins are found on six of the ten maize chromosomes (Xu and Messing, 2008). The gene duplication that gave rise to the 22-kD and 19-kD α-zeins occurred in the progenitor of maize and sorghum (Larkins et al., 2017). Following this duplication, there were numerous chromosomal translocations and gene duplications and deletions, as well as point mutations, that created the mosaic of α-zein loci that exist among different maize genetic backgrounds (Larkins et al., 2017). There are (variably) 41-48 α-zein genes organized in complex multigene families on chromosome 1,4, and 7 (Miclaus et al., 2011). These genes are encoded by four different subfamilies: Z1A, Z1B, Z1C, and Z1D. There is also copy number and expression variation across different maize backgrounds that have been found to affect the production of these zeins. Some of the 19-kD α-zeins controlling genes (twelve genes in B73 for example) are on the short arm of chromosome 4, belonging to the Z1A1 group (nine genes) and the smaller Z1A2 group (three genes). Additional 19-kD α-zeins genes are in three clusters: Z1B (nine genes on the short arm of chromosome 7) and Z1D (five genes) on the short arm of chromosome 1 (Song and Messing, 2002). The short arm of chromosome 4 contains a cluster of

22-kD α-zein genes, 14 of them classified in the Z1C1 subfamily and one to the Z1C2 subfamily in B73 (Song and Messing, 2003). In contrast to the highly duplicated nature of α-zein genes, genomic regions encoding the β-, γ-, and δ-zeins are typically single copy. The 50-kD and 27 kD-γ zein genes are in close proximity on the short arm of chromosome 7 (Yuan et al., 2014), whereas the 16-kD y-zein is on the long arm of chromosome 2. The 15-kD β-zein is on the short arm of chromosome 6, and the 10-kD and 18-kD δ are on chromosome 9 and 6, respectively (Larkins et al., 2017).

Zein genes are temporally and spatially regulated through the activity of shared cis-acting nucleotides sequences in their promoter and transcription factors (TFs) that recognize them (Larkins et al., 2017). None of the different types of zein genes contain introns; hence, post-transcriptional processing of mRNAs does not appear to play a major role in gene expression (Larkins et al, 2017). Three different types of zeins TFs have been described: (1) bZIP proteins, encoded by O2 and Opaque2 heterodimerizing protein 1 and 2 (OHP1/2) (Vicente-Carboja et al., 1997),(2) the Prolamin-Box Binding Factor (PBF1, an endosperm-specific zinc finger protein called ZmDOF3 (Zhang et al., 2015), and (3) a MADS box protein called ZmMADS47 (Qiao et al., 2016). PBF binds the P Box (TGTAAAG) which is a cis-acting regulatory sequence found in the promoters of prolamins genes in most cereals (Wu and Messing, 2012), while O2 and OHP1/2 recognize the O2 box (TT/CCACGT). ZmMADS47 binds a CATGT motif that flanks the O2 site in α-zein and 50-kD γ-zein promoters.

Known mutations in prolamin genes that alter endosperm vitreousness

Over the past century, a variety of maize mutants with an opaque or floury endosperm phenotype have been identified and are termed “opaque” [e.g., *opaque 1-17* (*o1, o2, o5-o7, o9-o11, o13-o17*),

recessive], "floury" [e.g., *floury1-4* (*fl1-fl4*), semidominant], or other assorted names [e.g., defective endosperm *De-B30* and *Mucronate* (*Mc*), *mutator-tagged opaque 140* (*mto 140*), recessive] (Gibbon and Larkins, 2005; Larkins et al., 2017). The diversity of these mutations raises the question as to what forms the basis of the opaque endosperm phenotype. This phenotype is commonly associated with altered PBs and/or starch granules, that is, mutants with an opaque phenotype display defective PBs structure or functions to varying degrees depending on the causal gene (Zhang et al., 2018).

The large number of endosperm mutants discovered and studied provide insights into the multiple molecular mechanism associated with the opaque endosperm phenotype. The nature of the underlying genes generally falls into three main categories: (1) zein coding sequences (e.g., *FL2*, *FL4*, *De-B30*, and *Mc*) and genes encoding non-zein proteins associated with PBs (e.g., *O1*, *O10*, and *FL1*); (2) genes encoding enzymes involved in endosperm metabolic processes (e.g., *O5*, *O6*, *O7* and *MTO140*); and (3) transcriptional regulatory genes (e.g., *FL3*, *O2*, and *O11*)(Zhang et al., 2018).

The most well-known of the maize opaque mutants is *opaque2 (o2)* which has been widely studied because of the increased lysine and tryptophan accumulation (Mertz et al., 1964) resulting from its reduced accumulation of α-zeins. The *O2* gene encodes a basic domain/leucine zipper (b-ZIP) transcriptional factor that binds the promotor of a subset of α -zein genes as well as a gene encoding a ribosome-inactivating protein (Habben et al., 1993). Transcription of these genes is dramatically reduced in *opaque-2* mutants, and there is also significant reduction in the transcription of genes encoding β- γ- δ- zeins. Although the soft kernels and yield penalty of *o2* prevented its commercial success when first released, subsequent breeding projects, including those in Mexico (Vasal et al., 1980) and South Africa (Geevers and Lake, 1992), led to the development of vitreous endosperm

o2 varieties called quality protein maize (QPM). QPM kernels maintain the low levels of α-zein and thus, retain high levels of lysine and tryptophan while providing improved agronomics. Despite the success of this project, the genetic basis of *o2* endosperm is complicated and poorly understood (Holding, 2014). QPM endosperm accumulates larger numbers of small, γ-zein rich protein bodies which are proposed to allow the formation of a rigid glassy matrix similar in texture compared to mature wild type endosperm. γ-Zein is known to be essential for endosperm modification in QPM (Wu et al., 2010), although the extent to which this particular zein alone is sufficient to provide the observed phenotype is unknown.

Although the functional redundancy resulting from the multimember α-zeins gene families has prevented the identification of recessive mutants, several dominant opaque mutants have also been characterized. The phenotypes in these mutants result from the accumulation of defective zeins that interfere with normal zein deposition and cause ER stress response (Coleman et al., 1997; Kim et al., 2004; Kim et al., 2006; Wu et al., 2013). Classic examples of this type of mutation is *fl2*, *De-B30*, and *Mc*. These three mutations show pleiotropic effects and result in a general reduction of overall zeins (causing increases in lysine-containing protein similar to *o2*) and lobed protein bodies (Lending and Larkins, 1992). In the case of *fl2* and *De-B30,* this result from signal peptides that remain attached in the 22-kD α-zein and 19-kD α-zein, respectively, resulting in aggregation of these proteins at the ER membrane (Gillikin et al., 1997; Kim et al., 2004). The *Mc* mutant (Soave and Salamini, 1984) results from a 38-bp deletion that leads to a frame-shift mutation in the 16-kD γ-zein (Kim et al., 2006). The abnormal zeins produced in these mutants result in ER stress and cause a constitutive unfolded protein response (UPR), as evidenced by the dramatic up-regulation of a number of UPR-associated genes in these mutants (Hunter et al., 2002).

The characterization of opaque mutants has shown that vitreous endosperm formation depends on the correct expression and processing of zeins themselves but also on factors that may have indirect roles in prolamin protein bodies such as *fl1* and *o1*(Holding et al., 2007; Wang et al., 2012). These two mutations are associated with defects in myosin or myosin-related proteins and may be associated with cytoskeleton surrounding the RER (Clore et al., 1996). Mutant *o1* affects ER streaming and could influence zein translocation into the RER, while *fl1* is associated with the PBs membrane and results in a subtle change in zein organization within PBs (Larkins et al., 2017).

Not all opaque mutations directly or indirectly affect zein synthesis. For example, the *o5* mutation corresponds to a defective monogalactosyldiacylglycerol synthase (MGD1) and affects the phospholipid composition of amyloplast membranes (Myers *et al*., 2011). Table 1-1 presents a summary of the maize traditional opaque endosperm mutants mapped.

Impact of maize endosperm vitreousness in ruminant starch digestibility

Ruminal starch digestibility efficiency is a subject of great interest for dairy scientist whom are looking to optimize the energy conversion from the forage to the animal. Greater endosperm starch digestibility results in increased energy available for dairy cows and thereby greater milk production, feed efficiency, or both (Firkins et al., 2001; Ferraretto et al., 2013). This complex process may be affected by several factors. First the starchy endosperm is protected by the pericarp, which, if intact, is highly resistant to microbial attachment (McAllister et al., 1994). Mechanical processing of the kernels breaks up the pericarp leaving the starchy endosperm exposed for ruminal bacterial attack. However, even exposed endosperms are not fully digested, as starch granules are tightly packed and encapsulated by proteins that forms, what is known as, the starch-protein matrix mostly composed of zeins.

Several studies have observed a strong negative relationship between endosperm vitreousness and in situ starch or DM degradability, meaning as endosperm vitreousness increases *in situ* starch DM degradability decreases. Ngonyamo-Majee et al., 2008 evaluated *in situ* DM degradability of 31 maize inbred lines differing in endosperm vitreousness levels. This study was done using only the grain fraction which was ground and passed through a 6-mm screen, placed in dacron (permeable polyester) bags and incubated for 14 hours in cannulated steers. They found a negative relationship between endosperm vitreousness and *in situ* Ruminal Dry Matter Degradability (RDMD) with a coefficient of determination R^2=0.72 ($p<0.001$) and Total Dry Matter Degradability (TDMD) R^2=0.34 ($p<0.001$). Similar results were observed *in vivo* by Allen et al., (2008) whom fed eight ruminally and duodenally cannulated lactating dairy cows with maize that had a range of 25 and 66% of vitreous endosperm. Feeding cows with maize with 66% of vitreous endosperm reduced ruminal and total tract starch digestion by 19.1 and 7.1 percentage units, respectively. Correa et al., 2002 also studied the relationship of endosperm vitreousness with ruminal *in situ* degradation of starch using five vitreous Brazilian hybrids kernel and 14 U.S. dent hybrids and also compared the correlation between endosperm vitreousness and kernel density. In this study, they observed that the correlation between kernel vitreousness or density and ruminal starch availability were r=-0.93 and r=-0.87, respectively, suggesting that kernel vitreousness and density may be a useful parameter to select maize hybrids for ruminal starch availability and suggested density as a more practical measurement than vitreousness for screening large maize data sets (Correa et al., 2002).

It is hypothesized that the zeins of the starch-protein matrix acts as limiting factor to starch hydrolysis in the rumen. Zeins are not soluble in water nor soluble in solvents typically found in the innate rumen environment (Lawton, 2002). Potentially, starch digestion requires rumen bacteria to first degrade prolamin via proteolysis before amylolytic activity in the rumen can

actively hydrolyze starch to glucose (Cotta, 1998). Because glucose uptake by rumen bacteria is momentary (Franklund and Glass, 1987) and the rumen has extensive amylolytic capacity to hydrolyze starch to glucose, proteolysis of hydrophobic zeins in the rumen should therefore be a rate limiting step associated with starch digestion.

McAllister et el., (1993) investigated the potential influence of starch protein matrix on starch digestion in a classical study where they observed that when maize was treated with protease (proteolytic enzymes), *in vitro* starch digestion increased approximately two-fold and therefore they concluded that the protein matrix in maize was a major factor in ruminal starch digestion.

Ensiling is a method for preserving forages used by farmers for several thousands of years. Silage making is an especially attractive option for preserving forages in northern temperatures regions with restricted growing season. Maize has several practical attributes that make a suitable crop for ensiling. Maize has a high dry matter yield that is obtained in a single harvest, and harvest can be completed prior to significant leaf loss (Coors and Lauer, 2001). Second, maize silage is characterized by stable fermentation due to dry matter concentration greater than 200 g/Kg, high nonstructural carbohydrate concentration, and low buffering capacity (Fisher and Burns, 1987; McDonald et al., 1991)

The ensiling period is a dynamic process that start with harvesting and chopping all the above ground plant material, at a specific moisture (typically 30% of DM for maize). Soon after chopped, the forage must be compacted and tightly sealed to remove as much oxygen as possible. This is where the “aerobic phase” starts and continues until oxygen is depleted, shortly after ensiling. During this stage plant sugars are broken down to carbon dioxide, water and heat through respiration. During the aerobic phase the degradation of plant proteins to nonprotein nitrogen

(NPN), peptides, amino acids and ammonia by plant cell proteases occurs. The extent of proteolysis will depend on the rate of pH decline, temperature and moisture content of the silo. Once the oxygen has been depleted, the second phase known as the "anaerobic fermentation phase" begins. During this phase, a succession of different populations of anaerobic bacteria, dominated by lactic acid bacteria ferment sugars and convert them into mostly lactic acid, but also acetic acid, ethanol, carbon dioxide and a few other minor products. The production of acid lowers the pH of the ensiled maize which inhibits the growth of other microbes. The bacteria remain active until the silage reaches a stable pH of 4 or below, or until the fermentation sugars are depleted. This marks the start of the "storage phase" where the ensiled material remains relatively stable with minimal microbial and enzymatic activity if the system is kept anaerobic. This stage prevails until the silo is opened in the last phase called "feedout phase".

An important question to ask is what happens with endosperm content, and particularly with starch and the protein matrix, during a prolonged ensiling period which involves active microbiological activity. Philippeau and Michalet-Doreau (1998), observed that ensiling grains increased ruminal starch degradability and hypothesized that ensiling increases accessibility of starch granules to rumen microorganisms, because hydrophobic zein proteins are degraded by the process. Der Bedrosian et al., 2012 compared two commercial hybrids, one of them carrying the brown midrib (BMR) mutation which reduces lignin level and improves fiber digestibility. The hybrids were harvested at two different maturities and ensiled for various lengths of time (0 to 360 days). Measurements included nutrients, fermentation end products and *in vitro* digestion of starch among other things. They found that concentrations of soluble Nitrogen (N) and ammonia-N increased with length of storage, indicating that proteolytic mechanisms were active beyond two to three months of storage for both hybrids at both maturities. The *in vitro* digestion of starch

generally increased with length of storage, probably as a result of proteolysis. They also found a significant correlation between Soluble Protein (SP) and starch digestibility which further strengthens that proteolysis is probably the reason for increased starch digestibility (Der Bedrosian et al., 2012). Likewise, Jurjanz and Monteils (2005), observed ruminal degradability of starch to be greater for maize kernels after ensiling (92.3%) compared with those that were not ensiled (70.2%).

The increasing evidence supporting higher starch digestibility after ensiling provides a new perspective regarding which is the optimum maize endosperm vitreousness level for prolonged storage conditions. Based on these observations and contrary to prior suggestions, it is reasonable to hypothesize that harder endosperms types might be more efficient for increased animal productivity when proteolytic activity inherent to the ensiling environment is able to mitigate the negative effects of the starch protein matrix.

Phenotypes challenges of maize endosperm traits

Research studies that involve comparing physical features of samples requires direct measurements, or at least an estimation, of the feature of interest. When designing experiments, researchers need to consider several characteristics of the possible alternatives to measure such as cost of deploying and operation, sample preparation, speed, accuracy, and difficulty to implement. In the context of plant breeding, cost and speed are particularly relevant as oftentimes hundreds or thousands of samples need to be measured to characterize populations or to perform genetic mapping.

With the advent of digital imaging and computer vision algorithms, many alternatives have become available to the research community to measure ear, cob and kernel traits to help elucidate

important questions related to yield components. For example, a method for evaluating kernel size and shape on the ear was described in a patent by Pioneer Hi-Bred, Int. in 2007 (Hausmann et al., 2009) using digital cameras and computer vision algorithms. Publicly developed methods for analyzing ear images were published by Miller et al., (2017) and Warman et al., (2021). These methods have many advantages such as high speed, low cost of operation and the possibility of reanalyzing the images in the future. In the case of compositional traits such kernel starch, protein and oils, spectrophotometry techniques such as Near Infrared Spectrophotometry (NIR) have been widely adopted as a high-throughput and easier method than wet lab measurements. This spectroscopy-based method relies on collecting transmitted or reflected light from a sample between 780 and 2500nm and applying chemometric methods that exploit the inherent properties of the C-H, O-H, N-H and S-H organic bonds to absorb NIR light through overtone vibrations (Hacisalihoglu et al., 2016).

Endosperm vitreousness is an important trait for maize breeders and geneticist due to the multiple implications this trait has on potential end uses of the kernels. Along with scientists' interest to understand the biochemistry and genetics behind this trait, a need for better methods to measure it in a high-throughput fashion has risen, particularly fueled by breeding efforts towards superior starch digestibility maize silage hybrids.

The direct method to measure this trait is obtained by mechanically separating the floury soft portion from the hard section of the endosperm to calculate a mass ratio. This technique is labor intensive as it requires the removal of the pericarp and the embryo and entails dexterity to appropriately separate the endosperm in two fractions (vitreous and floury). Another method commonly used to asses this trait, takes advantage of the differential light transmissivity of the different endosperm types and consists of simply visually scoring the translucence over a light

box. This method might work for classifying extremely divergent phenotypes, but it is not conducive for fine classification or a quantitative assessment of vitreousness. The main advantages of this method it is simplicity and affordability.

An alternative to the manual dissection is to measure the mechanical resistance of the kernel to grinding. The Stenvert hardness test was first used to assess wheat hardness (Stenvert, 1974) and then successfully used by Pomeranz (Pomeranz et al., 1985) to assess variation in maize hardness. This test timed the period for a 17 mL tube, attached to a grinder, to fill with ground material, thereby relating hardness to resistance to grinding. A shorter period would indicates softer kernels (Pomeranz et al., 1985). This method requires sample destruction and only provides information on total hardness of a sample but it is not able to discern variation within kernels. The main advantages of this method are that it provides an objective measure and is less laborious than manual dissection. Gustafson and de León (2010) used the Stenvert hardness method with the Mo17 x B73 Recombinant Inbred Line (RIL) population to perform linkage mapping.

NIR has been also applied to predict endosperm hardness and vitreousness. Ngonyamo-Majee et al. (2008) developed NIR prediction curves for maize inbreds lines using ground kernels and proposed this method as a screening tool in large-scale breeding trials to develop corn hybrids of desired endosperm properties for improved ruminal degradabilities (Ngonyamo-Majee et al., 2008). Although good prediction accuracies can be achieved when calibration equations are appropriately developed, sample grinding is time consuming and requires relatively large sample sizes.

Hyperspectral imaging (HI) technologies combines the advantages of spectroscopy and imaging technologies (Feng et al., 2019). HI allows the collection of spectral information in a pixel basis which allows to infer physical or chemical properties of heterogenous samples and their spatial

distribution. This technology has successfully been implemented to classify maize seeds varieties (Zhao et al, 2018). This technology combined with chemometrics was tested as a new alternative to predict vitreousness in a high-throughput and non-destructive fashion in chapter 2.

Figures

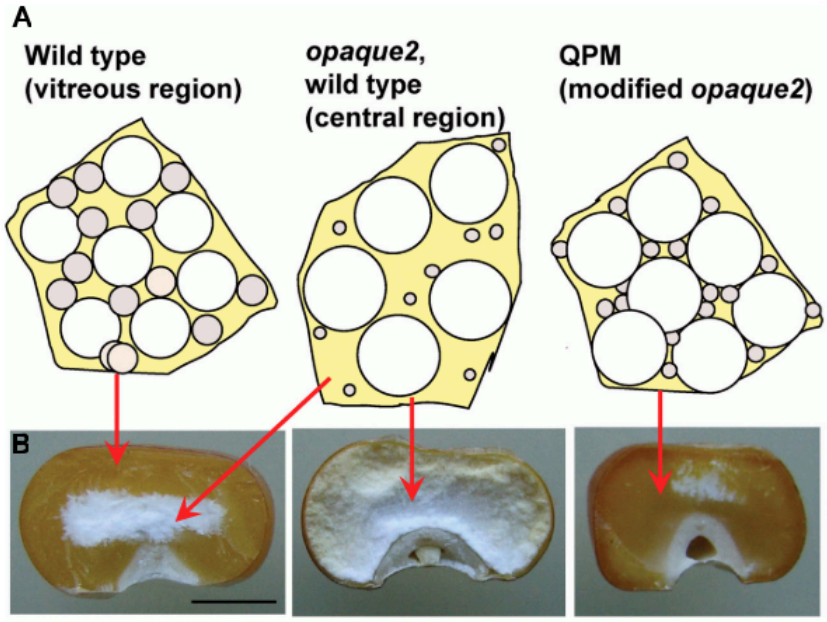


Figure 1-1 Vitreous endosperm formation in maize kernel

(A) Individual cells of developing endosperm are represented with the relative size and abundance of starch grains (white spheres) and zein protein bodies (gray spheres) that are thought to result in vitreous or opaque endosperm in normal as well as in opaque-2 and modified opaque2 (QPM) kernels. (B) Mature kernels of wild type, opaque-2 and QPM cracked in half to reveal extent of vitreous endosperm. From Holding, 2014.

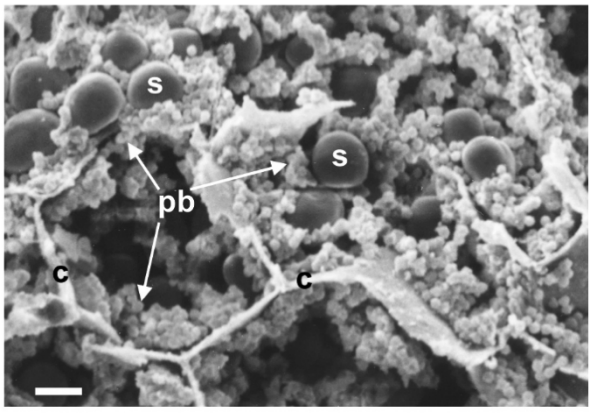


Figure 1-2 Scanning electron micrograph of endosperm cells in a developing maize kernel

The kernel was frozen and manually fractured prior to imaging and so some cellular contents may have been lost. Starch grains appear as grey spheres and adhering protein bodies are seen as smaller white spheres. Representative starch grains, protein bodies and cells walls are marked with s, pb and c, respectively. Scale bar in bottom left represent 10 μm. From Holding and Larkins 2006.

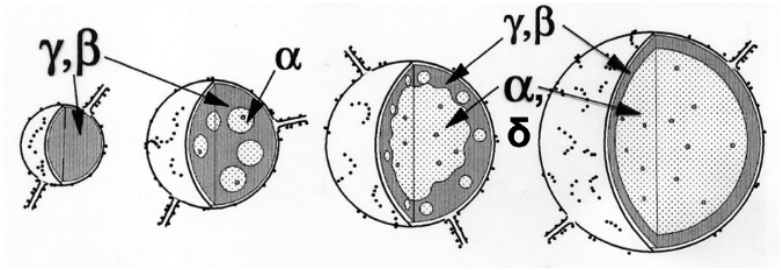


Figure 1-3 Diagrammatic representation of the protein body development

Diagrammatic representation of the process of protein body development showing localization of α-, β-, γ- and δ zein proteins. From Holding and Larkins 2006.

Tables

Table 1-1 Summary of opaqe endosperm mutants mapped

Mutant	Locus	Inheritance	Gene name	Protein type	Reference
o1	Zm00001d052110	recessive	*opaque-1*	Myosin XI motor	Wang et al., 2012
o2	Zm00001d018971	recessive	*opaque-2*	bZIP Transcription Factor	Schmidt et al., 1987 and 1990
o5	Zm00001d020537	recessive	*opaque-5*	Monogalactosyldiacylglycerol synthase	Myers et al., 2011
o6	Zm00001d010056	recessive	*PRO1*	D1-pyrroline-5-carboxylate synthetase	Wang et al., 2014
o7	Zm00001d026649	recessive	*AAE3*	Acyl-CoA synthetase	Miclaus et al., 2011 and Wang et al 2011
o10	Zm00001d033654	recessive	*opaque-10*	Cereal-specific non-zein PB protein	Yao et al., 2016
o11	Zm00001d003677	recessive	*opaque-11*	bHLH TF	Feng et al., 2018
fl1	Zm00001d003398	semidominant	*floury-1*	Non-zein PB protein	Holding et al., 2007
fl2	Zm00001d049243	semidominant	*floury-2*	22-kD α-zein	Coleman et al., 1997 and Gillikin et al 1997
fl3	Zm00001d009292	semidominant	*floury-3*	PLATZ TF	Li et al., 2017
fl4	Zm00001d048851	semidominant	*floury-4*	19-kD α-zeins	Wang et al., 2014
De-B30	Zm00001d019158	dominant	*defective endosperm B30*	19-kD α-zeins	Kim et al., 2004
Mc	Zm00001d005793	dominant	*Mucronate*	16-kD γ-zein	Kim et al., 2006
mto140	Zm00001d014734	recessive	*AroDH-1*	Arogenate dehydrogenase	Holding et al., 2010

Summary of maize traditional opaque endosperm mutants that has been mapped. Table modified from Zhang et al., 2018.

References

Allen, M.S., Longuski, R.A., Ying, Y. (2008). Endosperm type of dry ground corn grain affects ruminal and total tract digestion of starch in lactating dairy cows. *J. Dairy Sci*. 91: 520.

Becraft P.W., Gutierrez-Marcos J. (2012). Endosperm development: dynamic processes and cellular innovations underlying sibling altruism Wiley interdisciplinary reviews. *Dev Biol*. 1: 579–593.

Clore, A.M., Dannenhoffer, J.M. and Larkins, B.A. (1996). EF-1α is associated with a cytoskeletal network surrounding protein bodies in maize endosperm cells. *Plant Cell* 11: 2003–2014.

Coleman, C. E., Clore, A. M., Ranch, J. P., Higgins, R., Lopes, M. A., Larkins, B. A. (1997). Expression of a mutant alpha-zein creates the floury2 pheno- type in transgenic maize. *Proc. Natl. Acad. Sci*. 94: 7094–7097. doi: 10.1073/pnas.94.13.7094.

Coleman, C.E., Larkins, B.A. (1999). The Prolamins of Maize. pp. 109-139. In: P.Shewry, R.Casey (Eds.), Seed Proteins. Kluwer Academic Publishers, Dordrecht.

Coors, J.G., Lauer, J.G. (2001). Silage corn. In A.R, Hallauer, (Eds.). Specialty Corns (pp.347-392). CRC Press, Boca Raton, FL.

Correa, C.E.S., Shaver, R.D., Pereira, M.N., Lauer, J.G., Kohn, K. (2002). Relationship between corn vitreousness and ruminal in situ starch degradability. *J. Dairy Sci*. *85*: 3008-3012.

Cotta, M. (1988). Amylolytic of selected species of ruminal bacteria. *Appl. and Envir. Microb*. 54: 772-776.

Cueva, S.F., Stefenoni, H., Melgar, A., Räisänen, S.E., Lage, C.F.A., Wasson, D.E., Fetter, M.E., Pelaez, A.M., Roth, G.W., Hristov, A.N. (2021). Lactational performance, rumen fermentation, and enteric methane emission of dairy cows fed an amylase-enabled corn silage. *J. Dairy Sci.* 104: 9827-9841. Doi: https://doi.org/10.3168/jds.2021-20251

Das, O.P., Ward, K., Ray, S. Messing, J. (1991). Sequence variation between alleles reveals two types of copy correction at the 27-kDa zein locus of maize. *Genomics* 11: 849–856.

Der Bedrosian, M.C., Kung Jr., L, Nestor Jr., K. E. (2012). The effects of hybrid, maturity and length of storage on the composition and nutritive value of corn silage. *J. Dairy Sci.* 95: 5115–5126.

Dombrink-Kurtzman, M.A., Bietz, J.A. (1993). Zein composition in hard and soft endosperm of maize. *Cereal Chem.* 70: 105- 108.

Duvick, D.N. (1955). Cytoplasmic inclusions of the developing and mature maize endosperm. *Amer. J. Bot.* 42: 717-725.

Duvick, D.N. (1961). Protein granules of maize endosperm cells. *Cereal Chem*. 38: 374-385.

Esen, A. (1987). A proposed nomenclature for the Alcohol-Soluble Proteins (Zeins) of Maize (Zea-Mays-L). *J. Cereal Sci.* 5: 117-128.

Feng, F., Qi, W., Lv, Y., Yan, S., Xu, L., Yang, W., Yuan, Y., Chen, Y., Zhao, H., Song, R. (2018). *OPAQUE11* is a central hub of the regulatory network for maize endosperm development and nutrient metabolism. *Plant Cell.* 30: 375–396. doi: 10.1105/tpc.17.00616.

Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods.* 15: 1-25. Doi: https://doi.org/10.1186/s13007-019-0476-y.

Ferraretto, L. F., Crump, P. M., Shaver, R. D. (2013). Effect of cereal grain type and corn grain harvesting and processing methods on intake, digestion, and milk production by dairy cows through a meta-analysis. *J. Dairy Sci*. 96: 533–550.

Firkins, J.L., Eastridge, M.L., St-Pierre, N.R., Noftsger, S.M. (2001). Effects of grain variability and processing on starch utilization by lactating dairy cattle. *J. Anim. Sci*. 79: E218– E238.

Fisher, D. S., Burns, J.C. (1987). Quality analysis of summer-annual forages, II. Effects of forages carbohydrates constituents on silage fermentation. *Agron. J.* 79: 242.

Franklund, C.V., Glass, T.L. (1987). Glucose uptake by the cellulolytic ruminal anaerobe Bacteroides succinogenes. *J. Bact.*169: 500-506.

Geevers, H. O., Lake, J. K. (1992). Development of modified *opaque-2* maize in South Africa, in Quality Protein Maize, ed. E. T. Mertz. St. Paul, Minnesota: American Society of Cereal Chemists, 49–78.

Gibbon, B.C., Larkins, B.A. (2005). Molecular genetic approaches to developing quality protein maize. *Trends in Gen*. 21: 227–233. doi: 10.1016/j.tig.2005.02.009.

Gillikin, J. W., Zhang, F., Coleman, C. E., Bass, H.W., Larkins, B. A., Boston, R. S. (1997). A defective signal peptide tethers the *floury-2* zein to the endoplasmic reticulum membrane. *Plant Physiol*. 114: 345–352. doi: 10.1104/pp.114.1.345.

Gustafson, T., de Leon, N. (2010). Genetic analysis of maize (*Zea mays* L.) endosperm vitreousness and related hardness traits in the Intermated B73 x Mo17 recombinant inbred line population. *Crop Sci* 50: 2318-2327.

Habben, J.E., Kirleis, A.W., Larkins, B.A. (1993). The origin of lysine-containing proteins in *opaque-2* maize endosperm. *Plant Mol Biol.* 23: 825-838. doi: 10.1007/BF00021537.

Hacisalihoglu, G., Gustin, J.L., Louisma, J., Armstrong, P., Peter, G.F., Walker, A.R., Settles, A.M. (2016). Enhanced single seed trait prediction in soybean (*Glycine max*) and robust calibration model transfer with Near-Infrared Reflectance Spectroscopy. *J. of Agr. and Food Chem*. 64: 1079-1086. doi: https://doi.org/10.1021/acs.jafc.5b05508

Hamilton, T.S., Hamilton, B.C., Johnson, B.C., Mitchell, H.H. (1951). The dependence of the physical and chemical composition of the corn kernel on soil fertility and cropping system. *Cereal Chem.* 28: 163-176.

Harlan, D.W., Holter, J.B., Hayes, H.H. (1991). Detergent fiber traits to predict productive energy of forages fed free choices to nonlactating dairy cattle, *J. Dairy Sci.* 74, 1337.

Harz, C.O. (1885). Landwirtschaftliche Samenkunde. Paul Perey, Berlin.

Hausmann, N.J., Abadie, T.E., Cooper, M., Lafitte, H.R., Schussler, J.R. (2009). Method and system for digital image analysis of ear traits. https://patents.google.com/patent/US8073235B2/en (accessed 4 June 2022).

Holding, D.R. (2014). Recent advances in the study of prolamin storage protein organization and function. *Front. in Pl. Sci.* 5-276. doi: 10.3389/fpls.2014.00276.

Holding, D.R., Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica* 51: 243-254.

Holding, D.R., Meeley, R.B., Hazebroek, J., Selinger, D., Gruis, F., Jung, R., Larkins, B.A. (2010). Identification and characterization of the maize *arogenate dehydrogenase* gene family. *J Exp Bot.* 61: 3663–3673. doi: 10.1093/jxb/erq179.

Holding, D.R., Otegui, M.S., Li, B.L., Meeley, R.B., Dam, T., Hunter, B.G., Jung, R., Larkins, B.A. (2007). The maize *floury1* gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *Plant Cell* 19: 2569–2582. doi: 10.1105/tpc.107/053538

Hunter, B.G., Beatty, M.K., Singletary, G.W., Hamaker, B.R., Dilkes, B.P., Larkins, B.A., Jung, R. (2002). Maize opaque endosperm mutations create extensive changes in patterns of gene expression. *Plant Cell* 14: 2591–2612. doi: 10.1105/tpc. 003905.

Jurjanz, S., Monteils, V. (2005). Ruminal degradability of corn forages depending on the processing method employed. *An. Res.* 3: 15-23.

Kim, C.S., Gibbon, B.C., Gillikin, J.W., Larkins, B.A., Boston, R.S., Jung, R. (2006). The maize *Mucronate* mutation is a deletion in the 16-kDa gamma- zein gene that induces the unfolded protein response. *Plant J.* 48: 440–451. doi: 10.1111/j.1365-313X.2006.02884.x.

Kim, C.S., Hunter, B.G., Kraft, J., Boston, R.S., Yans, S., Jung, R., Larkins, B.A. (2004). A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an

opaque endosperm phenotype in the maize De∗-B30 mutant. *Plant Physiol.* 134: 380–387. doi: 10.1104/pp.103.031310.

Kim, C.S., Woo, Y.M., Clore, A.M., Burnett, R.J., Carneiro, N.P., Larkins, B.A. (2002). Zein protein interactions, rather than the asymmetric distribution of zein mRNAs on endoplasmic reticulum mem branes, influence protein body formation in maize endosperm. *Plant Cell* 14: 655–672.

Kriz, A.L. (1999). 7S globulins of cereals. pp. 477-498. In: P. Shewry, R. Casey (Eds.), Seed Proteins. Kluwer Academic Publishers, Dordrecht.

Larkins, B.A., Wu, Y., Song, R., Messing, J. (2017). Maize seed storage proteins. In B.A. Larkins (Eds), Maize Kernel Development. 175-189. CABI.

Lasztity, R. (1984). The Chemistry of Cereal Proteins. CRC Press, Inc. Boca Raton, FL.

Lawton, J.W. (2002) Zein: A history of processing and use. *Cereal Chem.* 79: 1-18.

Lending, C.R., Larkins, B.A. (1989). Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell.* 1: 1011-1023.

Lending, C.R., Larkins, B.A. (1992). Effect of the *floury-2* locus on protein body formation during maize endosperm development. *Protoplasma.* 171: 123–133. doi: 10.1007/BF01403727.

Li, Q., Wang, J., Ye, J., Zheng, X., Xiang, X, Li, C., Fu, M., Wang, Q., Zhang, Z., Wu, Y. (2017). The maize imprinted gene *Floury3* Encodes a PLATZ protein required for tRNA and 5S rRNA transcription through interaction with RNA polymerase III. *Plant Cell*. 29: 2661–2675. doi: 10.1105/tpc.17.00576.

Lopes M.A., Larkins, B.A. (1991). Gamma-zein content is related to endosperm modification in Quality Protein Maize. *Crop Sci.* 31: 1655-1662.

McAllister, T.A., Bae, H.D., Jones, G.A., Cheng, K.J. (1994). Microbial attachment and feed digestion in the rumen. *J. Anim. Sci*. 72: 3004–3018.

McAllister, T.A., Phillippe, R.C., Rode, L.M., Cheng, K.J. (1993). Effect of the protein matrix on the digestion of cereal grains by ruminal microorganisms. *J. Anim. Sci*. 71: 205-212.

McDonald, P., Henderson, A.R., Heron, S.J.E. (1991). The biochemistry of silage, Cambridge University Press, Marlow, England.

Mertz, E.T., Nelson, O.E., Bates, L.S. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science.* 145: 279–280. doi: 10.1126/science.145.3629.279.

Miclaus M., Wu Y., Xu J.-H, Dooner H.K., Messing J. (2011b). The maize high-lysine mutant *opaque7* is defective in an acyl-CoA synthetase-like protein. *Genetics.* 189: 1271–1280. doi: 10.1534/genetics.111.133918.

Miclaus, M., Xu, J.-H., Messing, J. (2011a). Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. *PLOS Gen.* 7: e1002131. https://doi.org/10.1371/journal.pgen.1002131.

Miller, N.D., Haase, N.J., Lee, J., Kaeppler, S.M., de Leon, N., Spalding, E.P. (2017). A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. *Plant J.* 89: 169–178. doi: 10.1111/tpj.13320.

Momany, F.A., Sessa, D.J., Lawton, J.W., Selling, G.W., Hamaker, S.A., Willet, J.L. (2006). Structural characterization of alpha-zein. *J. Agric. Food Chem.* 54: 543-547.

Mu-Forster, C., Wasserman, B.P. (1998). Surface localization of zein storage proteins in starch granules from maize endosperm: Proteolytic removal by thermolysin and in vitro cross-linking of granule-associated polypeptides. *Plant Physiol.* 116: 1563-1571.

Myers, A.M., James, M.G., Lin, Q., Yi, G., Stinard, P.S., Hennen-Bierwagen, T.A. Becraft, P.W. (2011). Maize *opaque5* encodes monogalactosyldiacylglycerol synthase and specifically affects C18:3/C18:2 galactolipids necessary for amyloplast and chloroplast function. *Plant Cell.* 23: 2331–2347.

Ngonyamo-Majee, D., Shaver, R.D., Coors, J.G., Sapienza, D., Lauer, J.G. (2008). Relationship between kernel vitreousness and dry matter degradability for diverse corn germplasm. II. Ruminal and post-ruminal degradabilities. *Anim. Feed Sci. Technol.* 42: 259-274.

Olsen, O.-A. (2004). Nuclear endosperm development in cereals and Arabidopsis thaliana. *Plant Cell.* 16: S214–S227. doi: 10.1105/tpc.017111.

Olsen, O.-A., Becraft, P.W. (2013). Endosperm development. In: Becraft PW, editor. Seed genomics. Hoboken: Wiley. pp. 43–62.

Philippeau, C., Michalet-Doreau, B. (1998). Influence of genotype and ensiling of corn grain on in situ degradation of starch in the rumen. *J. Dairy Sci*. 81: 2178-2184.

Pomeranz, Y., Czuchajowska, Z., Martin, C.R., Lai, F.S. (1985). Determination of corn hardness by the Stenvert hardness tester. *Cereal Chem.* 62: 108-112.

Qiao, Z., Qi, W., Wang, Q., Feng, Y., Yang, Q., et al. (2016). *ZmMADS47* regulates zein gene transcription through interaction with Opaque2. *PLOS Genetics* 12: e1005991.

Schmidt, R.J., Burr, F.A., Aukerman, M.J., Burr, B. (1990). Maize regulatory gene *Opaque-2* encodes a protein with a leucine-zipper motif that binds to zein DNA. *Proc. Natl. Acad. Sci*. 87: 46–50. doi: 10.1073/pnas.87.1.46.

Schmidt, R.J., Burr, F.A., Burr, B. (1987). Transposon tagging and molecular analysis of the maize regulatory locus *opaque-2*. *Science.* 238: 960–963. doi: 10.1126/science.2823388.

Soave, C., Salamini, F. (1984). Organization and regulation of zein genes in maize endosperm. *Philos. Trans. R. Soc. Lond. Series B Biol. Sci.* 304: 341–343. doi: 10.1098/rstb.1984.0029.

Song, R., Messing, J. (2002). Contiguous genomic DNA sequence comprising the 19-kDa-zein gene family from Zea mays. *Plant Phys.* 130: 1626–1635.

Song, R., Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc. Natl. Acad. Sci.* 100: 9055–9060.

Stenvert, N. L. J. (1974). Grinding resistance, a simple measure of wheat hardness. *J. of Flour and Animal Feed Milling*. 156: 24-25.

Tsai, C.Y., Larkins, B.A, Glover, D.V. (1978). Interaction of the *opaque-2* gene with starch-forming mutant genes on the synthesis of zein in maize endosperm. *Biochem. Genet*. 16: 883-896.

USDA National Statistics Service. (2022). Crop Production 2021 Summary. United States Department of Agriculture. Accessed January 15, 2022. https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan22.pdf

Vasal, S.K., Villegas, E., Bjarnason, M., Gela, W., Goertz, P. (1980). Genetic modifiers and breeding strategies in developing hard endosperm *opaque2* materials, in Quality Traits of Maize for Grain and Silage Use, eds W. G. Pollmer and R. H. Phillips. London: Martinus Nijhoff, 37–73.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. Schmidt, R.J. (1997). A maize zinc-finger protein binds the prolamin box in zein promoters and interacts with the basic leucine zipper transcriptional activator *Opaque2*. *Proc. Natl. Acad. Sci*. 94: 7685–7690.

Wang, G., Qi, W., Wu, Q., Yao, D., Zhang, J., Zhu, J., Wang, G., Wang, G., Tang, Y., Song, R. (2014). Identification and characterization of maize *floury4* as a novel semidominant opaque mutant that disrupts protein body assembly. *Plant Physiol.* 165: 582–594. doi: 10.1104/pp.114.238030.

Wang, G., Wang, F., Wang, G., Wang, F., Zhang, X., Zhong, M., Zhang, J., Lin, D., Tang, Y., Xu, Z., Song, R. (2012). *Opaque1* encodes a myosin XI motor protein that is required for endoplasmic reticulum motility and protein body formation in maize endosperm. *Plant Cell.* 24: 3447–3462. doi: 10.1105/tpc.112.101360.

Wang, G., Zhang, J., Wang, G., Fan, X., Sun, X., Qin, H., Xu, N., Zhong, M., Qiao, Z., Tang, Y., and Song, R. (2014). *Proline responding1* plays a critical role in regulating general protein synthesis and the cell cycle in maize. *Plant Cell.* 26: 2582–2600. doi: 10.1105/tpc.114.125559.

Wang, G.F., Wang, F., Wang, G., Wang, F., Zhang, X. W., Zhong, M. Y., Zhang, J., Lin, D., Tang, Y., Xu, Z., Song, R. (2012). *Opaque1* encodes a myosin XI motor protein that is required for endoplasmic reticulum motility and protein body formation in maize endosperm. *Plant Cell* 24: 3447–3462. doi: 10.1105/tpc.112.101360.

Warman, C., Sullivan, C.M., Preece, J., Buchanan, M.E., Vejlupkova, Z., Jaiswal, P., Fowler, J.E. (2021). A cost-effective maize ear phenotyping platform enables rapid categorization and quantification of kernels. *Plant J.* 106: 566–579. doi: https://doi.org/10.1111/tpj.15166.

Weatherwax, P. (1930). Endosperm of *Zea* and *Coix*. *Am. J. of Bot.* 17: 371-380.

Woo, Y.M., Hu, D.W.N., Larkins, B.A., Jung, R. (2001). Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell* 13: 2297–2317. doi: 10.1105/tpc.13.10.2297.

Wu, Y., Holding, D.R., Messing, J. (2010). Gamma-zein is essential for endosperm modification in quality protein maize. *Proc. Natl. Acad. Sci.* 107: 12810–12815. doi: 10.1073/pnas.1004721107.

Wu, Y., Wang, W. Messing, J. (2012). Balancing of sulfur in maize seed. *BMC Plant Biology* 12: 77. Doi:10.1186/1471-2229-12-77.

Wu, Y., Yuan, L., Guo, X., Holding, D. R., Messing, J. (2013). Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. *Nat. Commun*. 4: 2217. Doi: 10.1038/ncomms3217.

Xu, J.-H., Messing, J. (2008). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl. Acad. Sci.* 105: 14330–14335.

Yao, D., Qi, W., Li, X., Yang, Q., Yan, S., Ling, H., Wang, G., Wang, G., and Song, R. (2016). Maize *opaque10* Encodes a Cereal-Specific Protein That Is Essential for the Proper Distribution of Zeins in Endosperm Protein Bodies. *PLoS Gen.* 12 (8): e 1006270.

Yuan, L., Dou, Y., Kianian, S.F., Zhang, C. Holding, D.R. (2014). Deletion mutagenesis identifies a haplo insufficient role for γ-zein in *opaque2* endosperm modification. *Plant Physiol.* 164: 119–130.

Zhang, S., Zhan, J., Yadegari, R. (2018). Maize *opaque* mutants are no longer so opaque. *Plant Reproduction.* 31: 319-326.

Zhang, Z., Yang, J. Wu, Y. (2015). Transcriptional regulation of zein gene expression in maize through the additive and synergistic action of *opaque2*, prolamine-box binding factor and *O2* heterodimerizing proteins. *Plant Cell* 27: 1162–1172.

Zhao, Y., Zhu, S., Zhang, C., Feng, X., Feng, L., He, Y. (2018). Application of hyperspectral imaging and chemometrics for variety classification of maize seeds. *RSC Advances*. *8*: 1337–1345. doi: https://doi.org/10.1039/C7RA05954J.

Chapter 2 A Novel High-Throughput Hyperspectral Scanner and Analytical Methods for Predicting Maize Kernel Composition and Physical Traits

Author Contributions

This chapter was published as:

Jose I. Varela[1], Nathan D. Miller[2], Valentina Infante[1,3], Shawn M. Kaeppler[1,4], Natalia de Leon[1], Edgar P. Spalding[2]* (2022) A Novel High-Throughput Hyperspectral Scanner and Analytical Methods for Predicting Maize Kernel Composition and Physical Traits.

Food Chem 391:133264. doi: 10.1016/j.foodchem.2022.133264

* Corresponding author.

This was a highly collaborative project:

Jose Varela: Formal analysis, Methodology, Validation, Software, Writing-Original Draft. Nathan D. Miller: Software. Valentina Infante: Investigation. Shawn M. Kaeppler: Supervision, Writing-Reviewing and Editing. Natalia de León: Conceptualization, Funding acquisition, Supervision, Writing-Reviewing and Editing. Edgar Spalding: Conceptualization, Resources, Writing-Reviewing and Editing.

Abstract

Large-scale investigations of maize kernel traits important to researchers, breeders, and processors require high throughput methods, which are presently lacking. To address this bottleneck, we developed a novel flatbed platform that automatically acquires and analyzes multiwavelength near-infrared (NIR hyperspectral) images of maize kernels precisely enough to support robust predictions of protein content, density, and endosperm vitreousness. The upward facing-camera design and the automated ability to analyze the embryo or abgerminal sides of each individual kernel in a sample with the appropriate side-specific model helped to produce a superior combination of throughput and prediction accuracy compared to other single-kernel platforms. Protein was predicted to within 0.85% (root mean square error of prediction), density to within 0.038g/cm^3, and endosperm vitreousness percentage to within 6.3%. Kernel length and width were also accurately measured so that each kernel in a rapidly scanned sample was comprehensively characterized.

Keywords

Hyperspectral imaging; maize composition; vitreousness; NIR; PLSR; PLS-DA

Introduction

The chemical composition and structural characteristics of a maize (*Zea mays*) kernel determine how well suited it is for its various industrial uses. For example, wet millers prefer kernels with softer endosperm because they require less steeping time and allow better starch-protein separation while dry millers desire hard endosperms (Wu & Bergquist, 1991). The differences between hard and soft endosperms are mostly due to differences in how densely the starch granules are embedded within a complex protein matrix (Gustin et al., 2013) and to differences in the physicochemical properties of the starch itself (Xu et al., 2019). Hard endosperms have high vitreousness, referring to their glass-like optical properties while soft endosperms generally scatter light more and therefore appear opaque. Supplemental Figure 2-1 displays the visible differences between hard (vitreous) and soft (opaque) endosperms.

In the case of livestock feed, the bioavailability of starch in the endosperm is highly dependent on endosperm hardness (Philippeau & Michalet-Doreau, 1997; Dias Junior et al., 2016) because the protein matrix may affect how microorganisms in the rumen gain access to the starch granules (McAllister et al., 1994). Furthermore, endosperm vitreousness has been shown to affect resilience during harvest, storage, resistance to insects and fungi, and other practical characteristics (Holding & Larkins, 2006). Unfortunately, this important trait is difficult to measure directly.

Measuring vitreousness percentage typically requires manually removing the pericarp and embryo and then laboriously dissecting the floury soft endosperm from the vitreous hard portion to calculate a mass ratio. Alternative methods range from visually ranking light transmission of samples on a light box to quantifying resistance to grinding as a proxy for this trait (Gustafson & de León 2010). Endosperm vitreousness is highly correlated with total kernel density (Correa et

al., 2002), which can be measured by determining the volume of gas or liquid a known mass of kernels can displace, or by determining the percentage of kernels that float on a salt solution having a known specific gravity (Bergquist & Thompson, 1992). Both methods are low throughput, which limits the feasibility of evaluating the number of samples needed for large-scale studies. Reliable, automated, non-invasive measurement of vitreousness percentage would enable larger-scale investigations of endosperm quality.

Near-infrared (NIR) spectroscopy has been widely used to infer chemical composition and physical properties of maize kernels. The methodology relies on collecting transmitted or reflected light at wavelengths between 780 nm and 2500 nm and applying chemometric methods that exploit the inherent property of the C-H, O-H, N-H and S-H organic bonds to absorb NIR light through overtone vibrations (Seisler et al., 2008). NIR spectra from biological material have multiple overlapping absorbance patterns due to the complex mixture of organic compounds therefore multivariate statistical approaches are required to build the equations that can predict the trait given the spectrum (Spielbauer et al., 2009). To study cereal grains, spectra are typically measured from fine ground powders or bulk whole grain samples (Orman & Schuman 1991). NIR spectroscopy and properly calibrated equations can produce useful predictions of major seed constituents like starch, oil and proteins quickly and inexpensively (Fox & Manley, 2014). Ngonyamo-Majee et al. (2008) used NIR spectroscopy to develop vitreousness prediction equations for maize kernels by scanning powder from ground kernels. Although accurate predictions were attained, sample preparation is time-consuming and destructive.

Commercially available NIR analyzers acquire spectra from bulk grain samples. These units capture a spectrum representing an average of many intact kernels in unknown orientations within the sample. The spectra contain features that can predict kernel composition traits. At

least one study of intact kernels identified two spectral features that correlated well with endosperm hardness (Robutti, 1995). Alternatively, custom-built NIR reflectance spectrometers can capture a spectrum as a kernel tumbles down a tube, displaying different positions to the sensor (Spielbauer et al., 2009). These units cannot produce information about how composition varies between grain tissues such as the starchy endosperm and oil-rich embryo, which is relatively large in maize. Studies have shown that the germinal or abgerminal side of the kernel reflect different NIR spectra (Orman and Schumann 1992; Weinstock et al., 2006). Hyperspectral imaging combines NIR spectroscopy with pixel-based imaging (Feng et al., 2019). Hyperspectral imaging has been used to study maize and wheat kernels (Williams and Kucheryavskiy, 2016; Caporaso et al., 2018; Zhao et al., 2018) and to associate spectral signatures with specific tissues (Miao et al., 2020). The hyperspectral imaging work presented here is based on previous work demonstrating that maize kernels with categorically distinct degrees of endosperm hardness could be separated by analyzing NIR hyperspectral images (Williams et al., 2009; McGoverin et al., 2012). The novel flatbed imaging platform we describe acquired multiwavelength NIR images of large numbers of maize kernels that properly trained algorithms could process to predict endosperm vitreousness, kernel density, and kernel protein content while simultaneously measuring morphometric features such as kernel length and width.

Materials And Methods

Seed population and genetic relationships

WiDiv-942 is a panel of 942 maize inbred lines that produce physiologically mature kernels in the Midwest region of the United States (Mazaheri et al., 2019). Using a subset of 501 lines from the full WiDiv population, Renk et al. (2021) demonstrated that this population contains a

significant amount of variation in kernel composition phenotypes such as carbohydrates, oil, and protein. Thus, this population is appropriate for building models that can predict composition traits from NIR spectral information.

All kernels were produced in 2018 in a field at the University of Wisconsin, West Madison Agricultural Research Station (WMARS) using one-row plots of 3.7 m length and 0.76 m spacing and arranged according to a randomized complete block design. Plants were self-pollinated, to avoid any potential xenia effect of foreign pollen. The uppermost ear was harvested from five plants per row after plants reached physiological maturity, and the ears were then dried with forced air until approximately 12% moisture, and then stored indoors. The genetic relationships between inbred members of a population can be visualized by performing principal components analysis (PCA) on sets of DNA sequence features called single nucleotide polymorphisms (SNPs). To show the relationships between members of WiDiv-942 used here, PCA was performed on a set of 5000 SNPs evenly sampled from the 899,784 SNPs Mazaheri et al. (2019) originally identified. The principal components scores were computed using the package 'PCAtools' version 1.2.0 (Blighe, 2019) in R version 3.6.3 (R Core Team, 2020). Supplemental Figure 2-2 shows the distribution of the samples selected to build each prediction equation within the full WiDiv-942 space.

Three set of genotypes were used to build and test prediction models for endosperm vitreousness, kernel protein, and kernel density. The three sets were formed with members of the WiDiv-942 population because of the diversity it displays for the traits under study and because of future intentions to use it in genetic studies of the compositional phenotypes. A preliminary investigation explored variation in vitreousness levels enough to indicate which genotypes may represent an even sampling of the range of values across the population. This information was

used to create Set 1, comprising 1428 kernels from 149 genotypes used to build a model that predicts vitreousness. To select lines for building the protein prediction equations, the full WiDiV-942 population was scanned as powder using a commercial NIR analyzer (NIRS DS 2500, FOSS, Hilleroed, Denmark) and a model provided by the manufacturer. These results were used to select Set 2, comprising 479 kernels from 150 lines used to create a protein prediction model. The results of the preliminary vitreousness survey were used to guide genotype selection for density because density correlates highly with vitreousness. Set 3 refers to the 448 kernels extracted from 150 lines (Table 2-1) to create a density prediction model. Set 3 was also used to build weight and volume prediction models as those traits were directly measured for density calculations. Ten kernels per genotype were used for vitreousness and three kernels for protein and density. The total number of kernels shown in Table 2-1 is not exactly the product of genotypes and kernels per genotype because a few samples were discarded during laboratory analysis.

Ground truth measurements of traits

Endosperm vitreousness

The 10 kernels from each of the lines in Set 1 that were scanned and then dissected for ground truth measurement of vitreousness following the methods described in Correa et al. (2002) and Ngonyamo-Majee et al. (2008) were not selected randomly. Instead, for each line, 100 kernels were randomly selected and sorted into groups of 10 kernels that were visually most like each other. One kernel was randomly chosen from each of the 10 groups to make the final sample. This process produced a representative sample set that reduced bias toward a particular size or shape. After scanning the germinal and abgerminal sides in the grid configuration, each kernel

was soaked in distilled water for three minutes, then the pericarp and embryo were removed with a scalpel. The complete endosperm thus isolated was weighed on an electronic analytical balance (Ohaus AX224/E, Parsippany, New Jersey, USA). The floury endosperm component was manually removed using an electric rotary tool equipped with a 1/16 inch round engraving accessory with the aid of magnifying glasses to enhance accuracy. After all the floury endosperm was carefully removed, the weight of the remaining vitreous endosperm was recorded to calculate vitreousness as a percentage of the total endosperm weight.

Density

The density of 3 kernels randomly selected from each of the lines in Set 3 was determined by measuring buoyant force with an analytical balance according to the Archimedes principle. Supplemental Figure 2-3 shows the apparatus used. A 50 mL beaker containing 30 mL of distilled water at 22ºC was placed on a microbalance (Ohaus AX224/E, Parsippany, NJ, USA). A kernel attached to a needle was submerged using a drill press stand to control the motion. The balance recorded an increase in mass after the kernel and a marked section of the needle were submerged. The known volume of the submerged section of the needle was subtracted from the total displaced volume. The measured weight of the fluid that the kernel displaces was converted to a volume (specific density of water is 0.998 g mL^{-1} at 22°C). Dividing the mass of the kernel, measured separately, by the measured volume gives kernel density. The accuracy of the method was determined to be greater than 99% by measuring kernel-sized pieces of pure minerals of known densities (pyrite, quartz, fluorite, aventurite and hematite).

Protein

Three kernels from each of the lines of Set 2 were randomly selected for total C and N analysis, from which total protein was calculated. Each kernel was ground using a mortar and pestle and transferred to a 2 mL microcentrifuge tube. The powder was dried at 60^{o} C for 48 h and stored in borosilicate glass desiccators. A microbalance (Mettler-Toledo XP6, Columbus, Ohio, USA) was used to prepare tin foil capsules containing 10 mg of powder, which were combusted in an elemental analyzer (CE Elantech EA1112, Lakewood, New Jersey, USA). BBOT (2,5-Bis (5-tert-butyl-benzoxazol-2-yl) thiophene) and Atropine were used as calibration standards as recommended by the manufacturer. Total protein was calculated as N x 6.25 expressed on a dry weight basis.

Dimensions

The length and width of 480 kernels from Set 2 were measured with a precision digital caliper (Mitutoyo 500, Aurora, Illinois) to determine the accuracy of the automated image processing-based measurements. Kernel length was measured from the kernel tip to the center of the cap. Kernel width was defined as the largest distance perpendicular to kernel length axis.

Hyperspectral imaging device

The data in this study were acquired with a hyperspectral imaging device developed in collaboration with Middleton Spectral Vision (Middleton, Wisconsin, USA). The device uses a 12-bit NIR line-scan camera (Specim model FX17e, Oulu, Finland) to collect images of samples placed above it on a horizontal 16 x 107 cm glass plate. A bank of eight broadband quartz halogen lamps provided full spectrum illumination (Figure 2-1). When the camera is fitted with a

33-mm focal length lens, its line of 640 pixels covers 102 mm, resulting in a spatial resolution of 0.16 mm. The camera and lamps are mounted beneath the rectangular glass sample bed, facing upward. A motor translates the upward-facing camera and lamps along the long axis of the sample bed at 16.5 mm s^{-1} as the camera acquires lines (frames) at a rate of 100 s^{-1}. Each pixel in the line registers the energy in 224 wavelength bands between 950 nm and 1700 nm, corresponding to a spectral resolution of 3.3 nm.

The scanner and software were enabled to automatically scan a dark and white reference image to normalize each pixel value when processing the data. A piece of porous white polytetrafluoroethylene was added to the beginning of scanning region to collect a "white reference". A baseline "dark" reference was acquired by closing the camera shutter for 0.6 seconds. The spectra at each pixel in subsequent scans of biological material were corrected according to Equation 1.

$$I = \frac{I_0 - I_d}{I_w - I_d} \qquad Eq. 1$$

where I is the corrected image, I_0 is the raw image, I_d is the dark reference image, and I_w is the white reference image.

The maize kernels to be scanned were either scattered randomly on the glass sample bed or placed in a 5x24 grid fixture with one kernel per cell (Supplemental Figure 2-4). In both sample configurations, individual kernels were the analyzed unit. The grid arrangement allows a single indexed kernel to be retrieved for a posterior use such as destructive ground-truth measurement. The scattered arrangement is much faster, but it does not allow a particular kernel to be chosen for a future use. When the grid was used, each kernel was scanned twice - once with the embryo (germinal side) facing the camera and then the kernel was flipped to capture the abgerminal side. A model was trained to distinguish between the two sides of a kernel (section 2.6). When kernels

were scattered on the device, this model was used to determine which side of each kernel faced the camera. The two types of sample presentation served different purposes. The grid was used to associate image data with the same kernel used for destructive ground-truth measurements, which is necessary for building predictive models. The scattered kernel method was used in a high-throughput manner to make model-based inferences from many genotypes.

Computational methods and feature extraction pipeline

The raw images the device produces are multichannel images, which means they are *m* x *n* x *z* matrices where *m* and *n* are the width and length of the image in pixels and *z* is the number of wavelength bins (224 in this work) at which the photon fluence has been measured. The following image processing steps were coded in the MATLAB version R2019b computer language to extract information from these hyperspectral images of kernels in either the gridded or scattered configurations.

i. Raw images that were spectrally corrected with white and dark reference scans using Eq.1 were converted to absorbance values using Eq. 2.

$$Absorbance = \log_{10}\left(\frac{1}{reflectance}\right) \quad \text{Eq. 2}$$

ii. Sweeping a line-scan camera does not necessarily produce square pixels. Absorbance images were resized using a bicubic interpolation method (*imresize* in MATLAB) and the inner distance of the first cell of the grid as a reference.

iii. Each pixel having an absorbance value at 1090 nm less than 1.4 was set to zero. This threshold value, determined by inspection, accurately segmented the kernels from the background to create a binary mask.

iv. Small objects in the binary image, those containing fewer than 700 pixels, were dust or debris and therefore removed from the binary mask.

v. The average absorbance of a 3x3 pixels square region centered at the kernel's center of mass was calculated for each wavelength in the absorbance array. These data were used to determine if a kernel in a scattered sample was germinal-facing or abgerminal-facing as explained in section 2.6.

vi. For each object in a size-filtered binary mask (i.e., kernel), the absorbance at each wavelength (absorption spectra) were averaged across pixels to create an average absorption spectrum. This average spectrum was associated with a unique center-of-mass coordinate pair.

vii. Kernel length, width and area were extracted based on the size-filtered binary mask using the methods Miller et al. (2017) previously developed. Briefly, analysis of contour curvature initiates a process that identifies the tip of the kernel, which marks one end of the major axis (kernel length) and the longest orthogonal segment is kernel width.

viii. For each scanned image containing many kernels, the averaged absorbances were exported to the R programing environment were the orientation and PLSR predicted traits (vitreousness, protein, density, weight and volume) were calculated with methods described in 2.5

Prediction model construction, calibration, and validation.

Models to predict vitreousness, protein, density, weight, and volume, were built and tested using the 'pls' package version 2.8-0 (Liland et al., 2021) in R. Separate models were created for germinal-facing images and abgerminal facing images. The dependent variables were spectral data extracted from hyperspectral images obtained in the grid configuration and the independent variables were trait measurements (section 2.2) made on the same kernels. For each trait, two-

thirds of the kernels (n = 320 for protein, n= 302 for density, and n = 949 for vitreousness) were used for building the partial least squares regression (PLSR) model and one third (n = 159 protein, n=146 for density, and n = 479 for vitreousness) was held out to validate the model. The Kennard-Stone algorithm (Kennard & Stone, 1969) was used to partition the raw spectral data into these training and validation sets using the 'prospectr' package version 0.2.1 (Stevens and Ramirez-Lopez 2020) in R. Samples from the same genotype were assigned to one group or the other but not both to reduce associations between the calibration and validation data sets that could inflate prediction accuracy.

After partitioning the raw spectral data into training and validation sets, each spectrum was subjected to one of 13 filters included in the 'waves' package (Hershberger et al., 2020). This step determines the spectral filtering and smoothing pretreatment that enables a PLSR model to produce the most accurate predictions. To guard against overfitting, cross-validation with the "one-sigma heuristic" strategy included in the 'selectNcomp' function of the 'pls' package was used to determine the lowest number of latent factors the model could use before the error of prediction (RMSEP) increased more than one standard error from the best result achievable using any number of factors. The selected model was used to perform an external validation using the remaining one third of the kernels. Performance statistics were calculated for the cross-validation of the calibration set and the external validation set (Table 2-2). For each trait-kernel orientation (germinal/abgerminal) combination, the pretreatment method with the lowest Root Mean Square Error of Prediction (RMSEP) was chosen.

Kernel side classification.

Because composition differs between grain tissues, separate models were created for predicting traits from germinal-side and abgerminal-side images. To apply the correct model to each kernel in a scattered kernel image, the orientation of each kernel must be analytically determined. Partial least squares discriminant analysis (PLS-DA), a type of supervised learning (Barker and Rayens, 2003; Lee et al., 2018; Ruiz-Perez et al., 2020), was used to classify each kernel in an image as germinal or abgerminal based on spectral information obtained from the 3x3 pixel square section at the center of mass. Images from kernel Set 3 (protein) were used to train and test the PLS-DA model using the function 'plsda' from the package 'mdatools' version 0.11.3 (Kucheryavskiy, 2020) using the "SIMPLS" algorithm in R. 480 individualized kernels were scanned using the grid. The kernels were scanned in both orientations generating 8 images with 120 kernels each were the spectral data was extracted from each kernel. From a total of 960 manually classified kernels (480 germinal side and 480 abgerminal side), half were randomly selected to train the algorithm. The optimal number of latent variables was determined using the 'leave one out cross-validation' (LOOCV) scheme. The lowest RMSECV value determined the optimal number of latent variables (LV). Finally, the computed model was used to predict samples in the held-out validation set. To test the effect of spectral pretreatments on model performance, the same 13 spectral pretreatments evaluated during construction of the PLSR models were applied and subsequently submitted to PLS-DA calibration and cross-validation. The pretreatment with the highest classification accuracy defined as the number of correct predictions divided by the total number of predictions was selected and used. Classification accuracy with the validation set was calculated to generate a confusion matrix, along with specificity and sensitivity rates (Supplemental Table 2-1).

Complete pipeline

The processes, measurements, and analyses described in Sections 2.3-2.6 were combined to produce a pipeline shown in Figure 1B-E.

Results and Discussion

Variability of maize kernel traits in ground-truth sets

Directly measured traits ranged widely across the kernel samples (Table 2-1). Kernel volume displayed the largest range (5.6-fold). Kernel weight was second at 5-fold, followed by vitreousness (2.7-fold), protein (2.4-fold) and density (1.4-fold). The protein range of 8.02%-19.45% agreed reasonably well with a previous single-kernel study (Baye et al., 2006), and a density range of 1.0-1.35 g cm^{-3} was in accord with the single-kernel findings of Gustin et al. (2013). The ranges were also consistent with values from commercial hybrids (Correa et al., 2002). The range of kernel weight (0.08-0.42 g) and volume (0.07-0.39 cm^3) were likewise as expected for a diverse collection of inbred lines (Gustin et al., 2013). The large phenotypic diversity found in WiDiv-942 produced a large range of endosperm vitreousness (35%-95%), comparable to previous studies of this trait (Correa et al., 2002, Ngonyamo-Majee et al., 2008). This amount of variation represented in a sample of more than 1400 kernels endowed the training data with enough vitreousness variation to be generally useful in studies of maize.

Single kernel predictions based on NIR absorbance.

The first derivative of the Standard Normal Variate (SNV) scatter correction method was found to be the best pretreatment for protein prediction. With it, we obtained a RMSEP of 0.85% and an r^2 of 0.84 when predicting protein from the abgerminal-side spectra. The effects of different

spectral pretreatments on the model's performance are compared in Supplemental Figure 2-5. The highest accuracy obtained from abgerminal side agrees with the results reported by Jiang et al., 2007 who also scanned maize kernels from both orientations to predict protein using NIR. Our protein results closely agreed with Gustin et al., 2013 who obtained an r^2 of 0.86 and SEP = 0.89 using a single kernel NIR based prediction models. The protein results reported here also performed very similar to the ones shown by Spielbauer et al., 2009 who obtained SEP=0.81 and r^2=0.91 when scanning maize kernels in a custom made NIR machine.

Only moderate r^2 (0.43) but very low RMSEP (0.038 g/cm^3) was achieved when total kernel density was predicted given abgerminal-side spectral pretreated with the Savitzky-Golay smoothing filter (Table 2-2). Gustin et al. (2013) reported SEP for single seed NIR prediction of total kernel density of 0.067 g cm^{-3} using microcomputed tomography as the reference method. Our low r^2 value may be due to the narrow distribution of density values in the subset of the diversity panel selected. By contrast, NIR studies of maize kernels oftentimes include endosperm mutants or genotypes that have undergone selection for extreme endosperm characteristics, which may extend the range in a way that raises r^2, a statistic that is highly dependent on the range of the validation set (Davies & Fearn, 2006).

Moderately high r^2 (0.56) was achieved for endosperm vitreousness (Figure 2-2) using the abgerminal side and using the Savitzky-Golay smoothing and first order derivative with a windows size of 11. An RMSEP of 6.3% shows that this system compares favorably in terms of throughput and accuracy with previous reports. Ngonyamo-Majee et al. (2008) reported a prediction accuracy of 6.04 % for this trait, but used pre-processed ground kernels as the sample. An early application of NIR hyperspectral imaging to infer endosperm quality in intact kernels (Williams et al., 2009) proved the principle by showing that nondestructive analyses could

distinguish different categories of hardness. The results in Figure 2-2 show how the present platform predicted continuous values of endosperm vitreousness from spectral images of intact kernels to produce results that could be used in quantitative genetics and gene mapping studies. Vitreousness is defined as a mass ratio of two completely separable solid phases, which sometimes is difficult to achieve, at least with the currently available mechanical separation methods. Oftentimes there is as a very thin transition zone of hard floury to vitreous area rather than a completely distinguishable boundary between a floury and vitreous part. Despite this potential source of noise in the ground truth data, the RMSEP (~ 6%) indicates this non-destructive method that takes kernel orientation into account will be effective and broadly applicable in studies of vitreousness.

Figure 2-2 shows that our models for predicting total kernel volume and weight showed moderate-high r^2 (0.69 both) and low RMSEP values (0.022 cm^3 and 0.027 g, respectively,) that were very similar to previously reported predictions based on single-kernel spectra (Spielbauer et al., 2009**).**

Correlation between the genotypic means of vitreousness, kernel protein and kernel density were calculated for the genotypes that are shared in the three Sets (Supplemental Figure 2-6). A positive significant Pearson's correlation coefficient of 0.67 (pval<0.001) was found between endosperm vitreousness and kernel density. This result agrees with the value reported by Correa et al., 2002 who found a positive significant correlation between these two traits (r = 0.87) and suggested that density may be a reliable tool for screening large maize data sets for vitreousness. Total protein was positively correlated (r = 0.39, pval<0.01) with endosperm vitreousness, in close agreement with the r = 0.41 found by Ngonyamo-Majee et al. (2007).

Kernel position detection based on embryonic reflectance for high throughput scanning in bulk configuration.

When scattering the kernels over the scanning region, they will randomly land with either the germinal side (embryo) or abgerminal side facing the camera. In order to use the most accurate NIR model developed in section 3.2 without manually orienting each kernel, we explored using PLS-DA to classify the orientation of each kernel in the image based on NIR absorbance information. We hypothesized that a successful orientation classifier could be based on a reliable difference in some regions of the NIR spectra between the embryo and the endosperm (Orman & Schumann, 1992). To explore this, 80 kernels from Set 3 were randomly chosen and the average absorbance for both positions was plotted. The region in the vicinity of 1200 and 1720 nm has distinguishable higher absorbances for the embryo compared to endosperm as shown in Figure 3. Osborne et al., (1993) has described that the major absorption band in oil is due to a long chain fatty acid moiety that gives rise to CH_2 second overtone at 1200 nm and the band near 1180 nm has been assigned as the second overtone of the fundamental C-H absorption of pure fatty acids containing *cis* double bonds such as oleic acid (Sato et al., 1991). Cho and Iwamoto (1989) correlated the absorption bands at 1710 and 1725 nm to linoleic and oleic acids, respectively. The larger absorbances observed in the aforementioned regions in our study could, therefore, be correlated to the two dominant fatty acids presents in corn, which are mostly presented in the embryonic region (Barrera-Arellano et al., 2019).

Instead of using the average absorbance of the whole kernel exposed to the camera as input to train the classification algorithm, a 3x3 pixel square at the center of mass was sampled. This region invariably overlaid the embryo when it faced the camera, and the endosperm when the abgerminal side faced down. Absorbance data from this centrally-located group of pixels

provided a substantially clearer signal than averaging the whole region. Absorbance data of Set 3 were used to calibrate and test the PLS-DA algorithm (section 2.6). Half of the manually classified samples were randomly selected for calibration while the other half was used for validation. Almost perfect classification accuracy was achieved with over 99.5% of correct position identification in both the calibration and validation dataset (Supplemental Table 1). The previously developed PLS-DA model (section 2.6) was included in the high throughput NIR processing pipeline to classify each kernel (Figure 2-1E) so its orientation can determine which trait-specific NIR prediction model is used (Figure 2-4).

Single kernel morphometric feature extraction based on image analysis.

The binary mask used to extract NIR absorbances from individual kernels was also used to measure kernel length and width (Figure 2-1D). Measuring the kernel length (major axis) and width (minor axis) depends on correctly identifying the kernel tip. The algorithm developed by Miller et al. (2017) was slightly modified to measure kernel length and width from the images this novel flatbed NIR scanner produces. The correlation between hand and image-based measurements of kernel length was 0.95 (Figure 2-2F) and 0.9 for kernel width. In general, this correlation may be limited by inaccuracy of hand measurements or inability of the image processing algorithm to correctly identify the tip of each kernel, particularly in kernels having imperfectly flat faces and rounded edges.

Conclusions

This upward-focused flatbed hyperspectral imaging scanner generated data that custom models used to predict important maize kernel traits with high throughput (75 kernels nondestructively prepared and measured in approximately 60 s). Endosperm vitreousness of a single kernel was

predicted to a useful degree across a wide range of kernel types, indicating that this automated alternative to a laborious manual method would be generally useful rather than population dependent. Accuracy of kernel protein and density predictions, which are relevant to multiple grain markets, were similar to or greater than previous reports. The imaging capabilities of the platform allowed germinal or abgerminal side-specific models to be applied, which improved accuracy for some traits, and it enabled kernel size dimensions to be measured at the same time. These features could make the platform described here useful to food processors, livestock feed producers, maize researchers, and breeders. The platform may prove useful in the study of other seed crops.

Figures

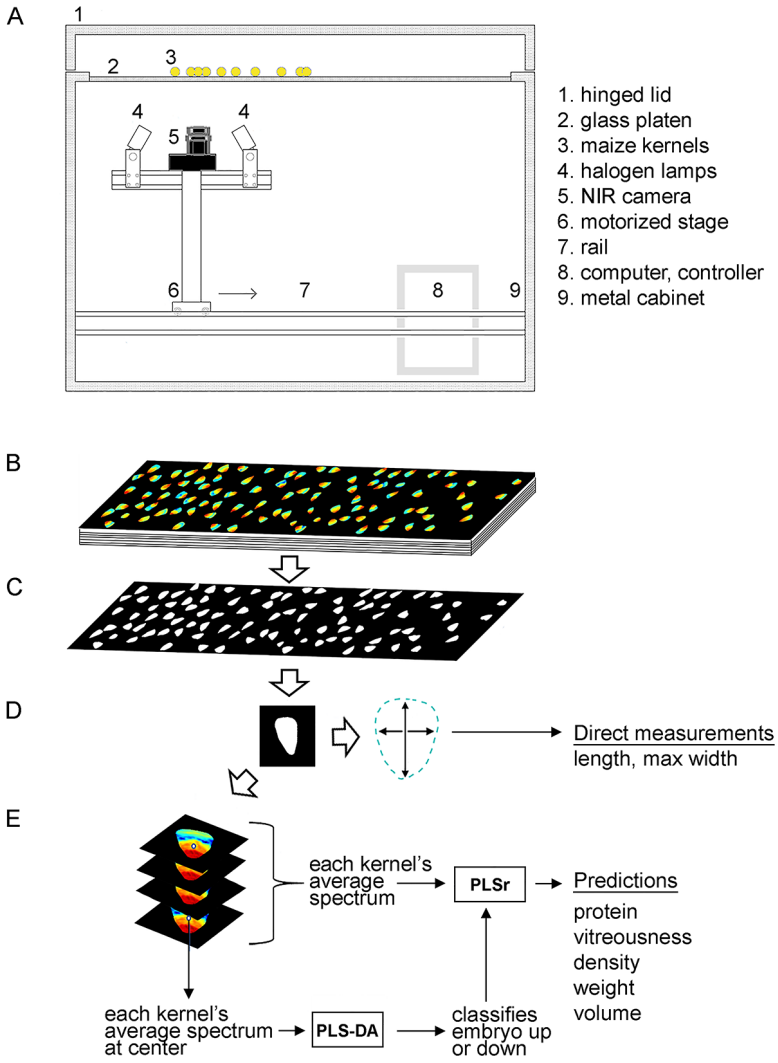


Figure 2-1 Acquisition and processing of hyperspectral images of maize kernels

A.Diagram of the flatbed hyperspectral scanner. B. The scanner produces a stack of 224 images, the pixels in each one registering the amount of light in a different narrow band of wavelengths between 950 nm and 1700 nm. C. A binarization process produces a mask with kernel pixels set to 1 and background pixels set to 0. D. The length and maximum width of each separate kernel is directly measured from the binarized image. E. The spectrum of each pixel in each kernel object is obtained from the stack depicted in B. The spectra from each pixel in the kernel object are averaged. To determine if the averaged spectra are from the germinal side of the kernel, or its abgerminal side, the average spectrum of pixels at the center of the kernel is used as an input to a PLS-DA classifier that is trained to distinguish between the two sides. The up or down label and the kernel's average spectrum are the inputs to a PLSR model that is trained to predict the indicated kernel traits.

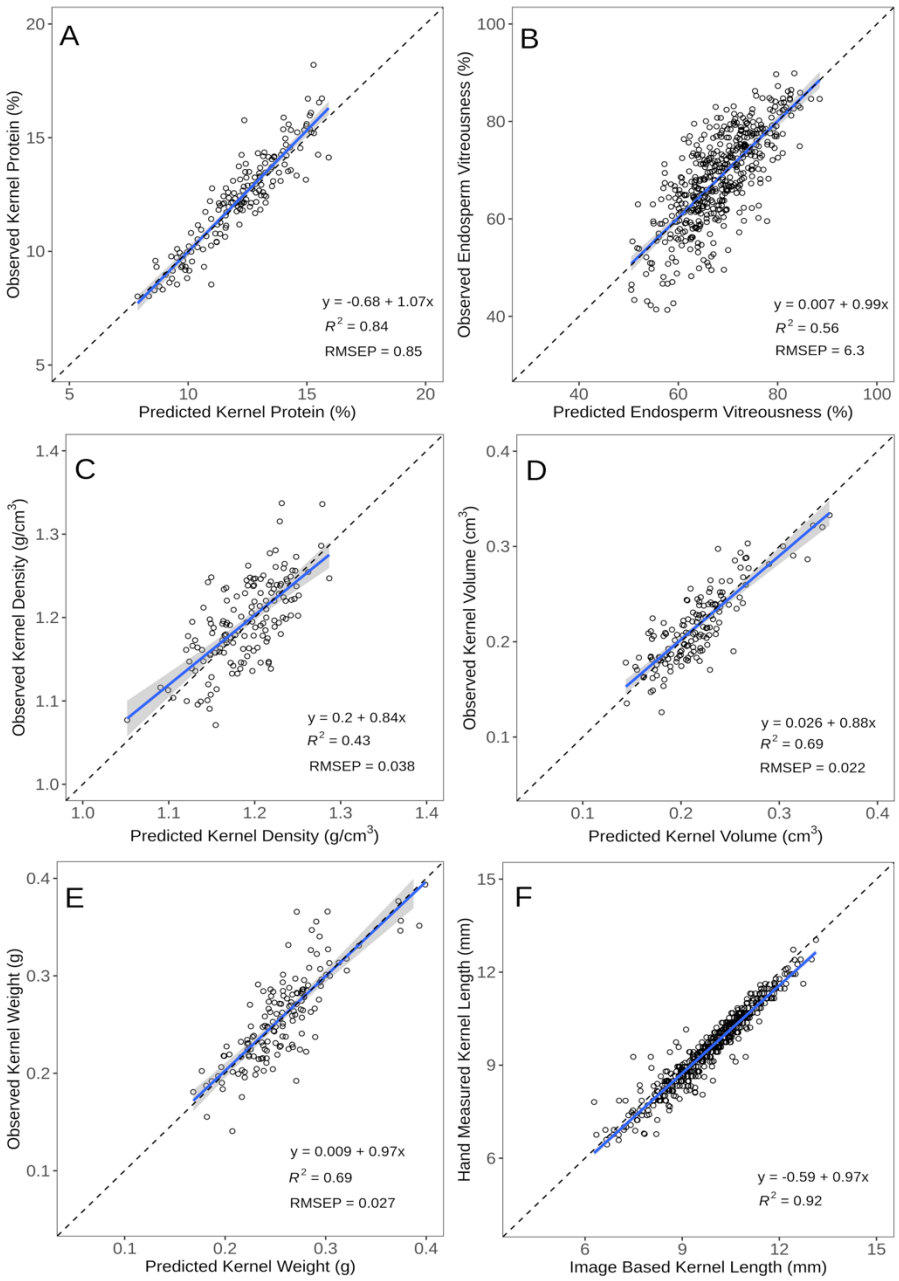


Figure 2-2 Scatter plots of observed (ground truth) measurement versus spectral-based model predictions or direct measurements of dimensions

(A) kernel protein, (B) endosperm vitreousness, (C) kernel density, (D) kernel volume, (E) kernel weight. A-E, NIR-predicted and lab measured kernel and endosperm traits. (F) image-based kernel length versus hand measured kernel length. Each panel displays values for the external validation. The blue line represents the linear regression line based on the equation on the upper left. The dotted line shows the perfect agreement. r^2, Coefficient of determination; RMSEP, root mean square error of prediction.

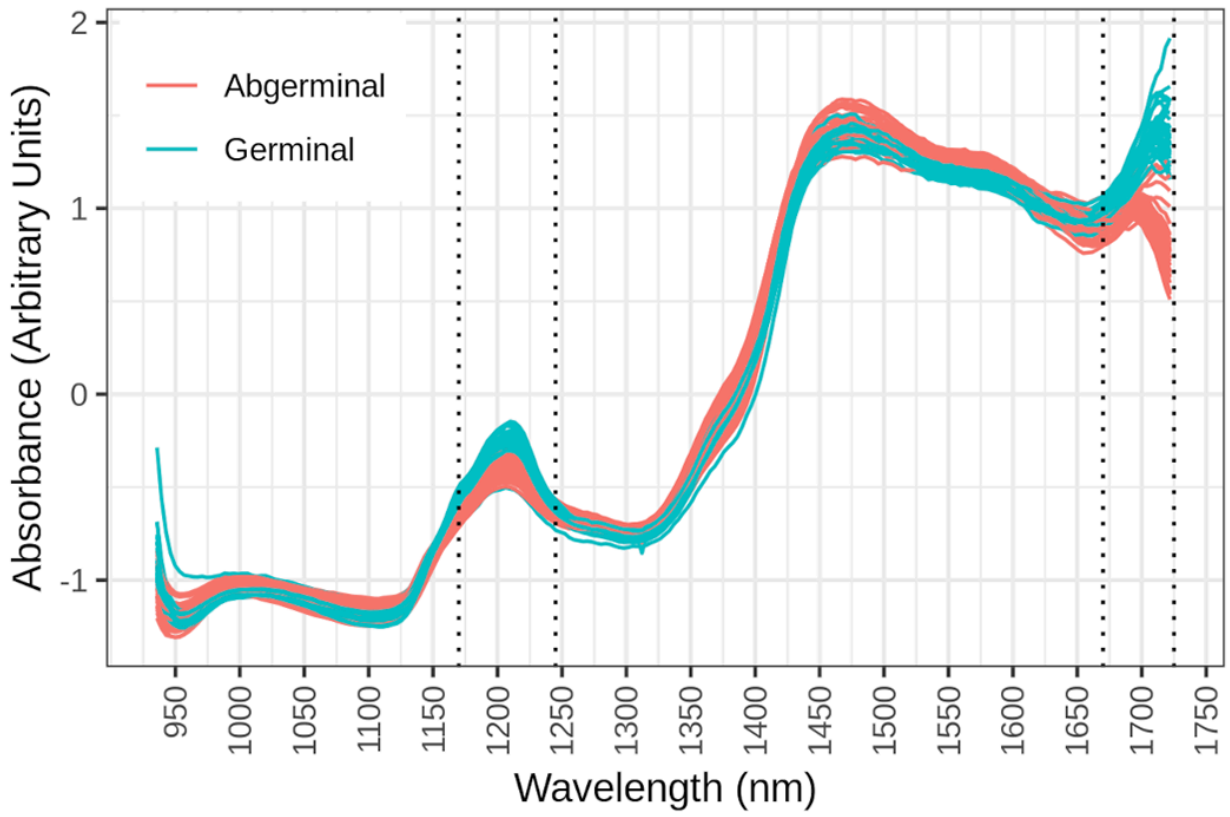


Figure 2-3 Reflectance spectra in the NIR region for 80 maize kernel samples

Half of the samples were manually oriented with the germinal (embryo) side facing the camera (red lines, n=40) and the other half with the abgerminal side facing the camera (light blue lines, n=40). The vertical dotted lines indicate spectral regions that differed most between the two sides of the kernel.

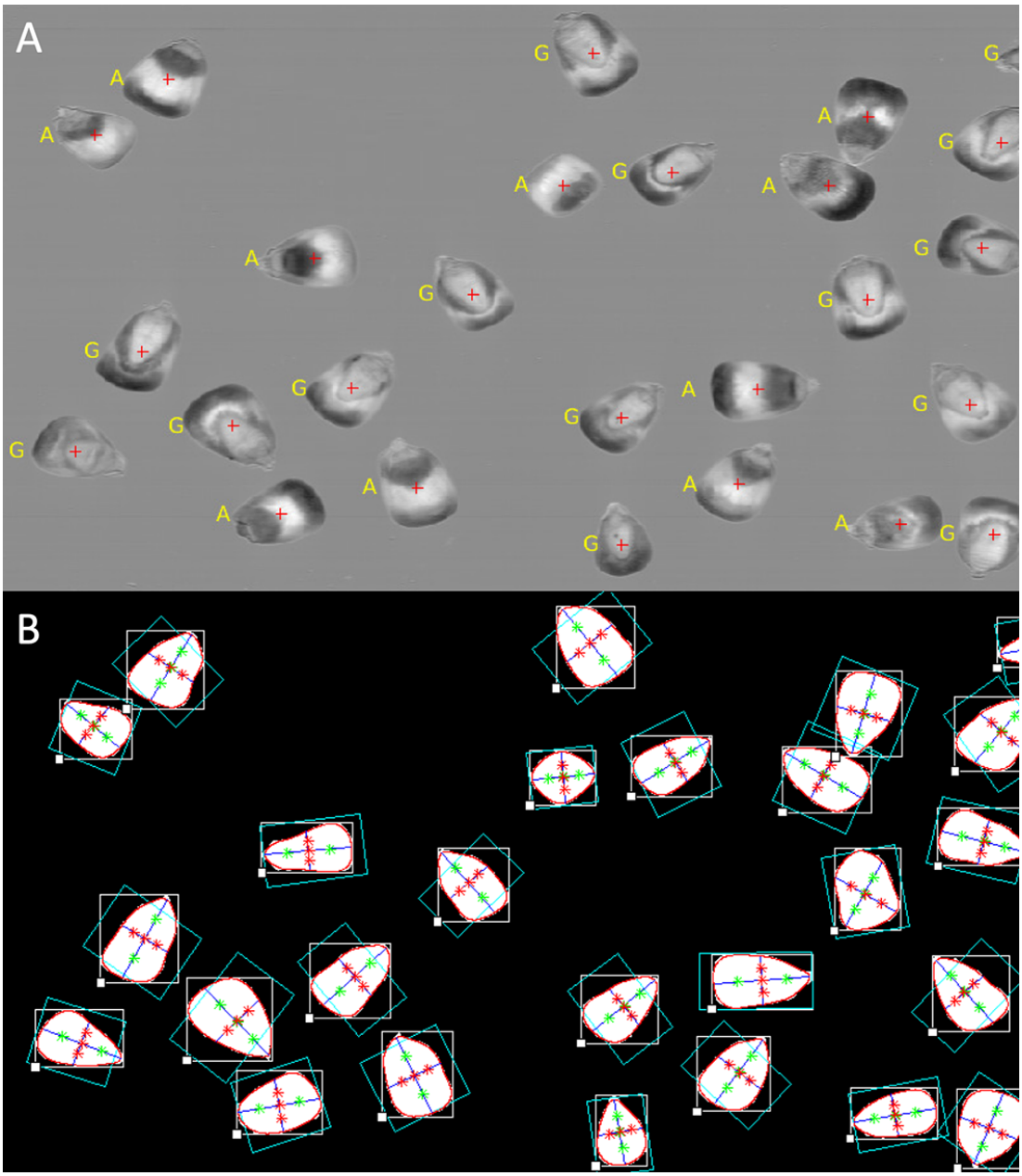


Figure 2-4 Hyperspectral image of kernels and binary morphology measurements

(A) Third Principal Component score image of raw NIR absorbances showing a section of the bulk scanning system surface. The red crosses represent the center of mass of each kernel. In yellow the PLS-DA kernel orientation classification output, G = Germinal and A = Abgerminal side of the kernel facing the camera. (B) Binary mask of a section of the bulk scanning system surface generated from the NIR image collected in the flatbed scanner. Kernel contours are depicted in red, kernel major axis and minor axis are depicted as a blue line.

Tables

Table 2-1 Composition statistics of maize kernels used for calibration and validation

Cross-validation								**External validation**					
Trait	**Set**	**Genotypes (n)**	**Kernels (n)**	**Mean**	**SD**	**CV**	**Range**	**Genotypes (n)**	**Kernels (n)**	**Mean**	**SD**	**CV**	**Range**
Vitreousness (%)	1	99	949	66.97	12.09	0.18	35.08-95.8	50	479	68.36	9.48	0.14	41.33-89.87
Protein (%)	2	107	320	12.51	2.13	0.17	8.17-19.45	53	159	12.32	2.1	0.17	8.02-18.2
Density (g/cm^3)	3	104	302	1.18	0.06	0.05	1.0-1.35	52	146	1.19	0.05	0.04	1.07-1.34
Volume (cm^3)	3	104	302	0.21	0.05	0.23	0.07-0.39	52	146	0.22	0.04	0.18	0.13-0.33
Weight (g)	3	103	299	0.25	0.06	0.23	0.08-0.42	53	149	0.27	0.05	0.18	0.14-0.41

SD = standard deviation, CV = coefficient of variation

Table 2-2 Performance of the PLS model for five kernel traits and two orientations using the optimal spectral pretreatment

Kernel trait	Set	Kernel Orientation	Spectra pretreatment	PLS factors	Cross-validation			External validation	
					RMSE	PRESS	R^2	RMSE	R^2
Vitreousness (%)	1	Abgerminal	SG.D1W11	13	7.42	52361.7	0.62	6.3	0.56
		Germinal	SG	14	7.1	48611.9	6.5	6.6	0.51
Protein (%)	2	Abgerminal	SNV1D	11	1.06	360.9	0.75	0.85	0.84
		Germinal	SNVSG	17	1.06	358.3	0.76	0.92	0.79
Density (g/cm^3)	3	Abgerminal	SG	7	0.036	0.41	0.67	0.038	0.43
		Germinal	SNV	8	0.042	0.53	0.52	0.04	0.52
Volume (cm^3)	3	Abgerminal	Raw	17	0.029	0.26	0.63	0.027	0.54
		Germinal	Raw	12	0.029	0.24	0.67	0.022	0.69
Weight (g)	3	Abgerminal	Raw	17	0.032	0.31	0.68	0.032	0.57
		Germinal	SG.D1W11	9	0.033	0.34	0.66	0.027	0.69

RMSE, Root mean square error of prediction; R^2, Coefficient of determination; PRESS, Predicted residual sum of squares.
Spectral pretreatments: SG.D1W11, Savitzky-Golay + 1st derivative using windows size of 11; SG, Savitzky-Golay; SNV1D, Standard Normal Variate + 1st derivative; SNVSG, Standard Normal Variate + Savitzky-Golay; SNV, Standard Normal Variate.

Supplemental Materials

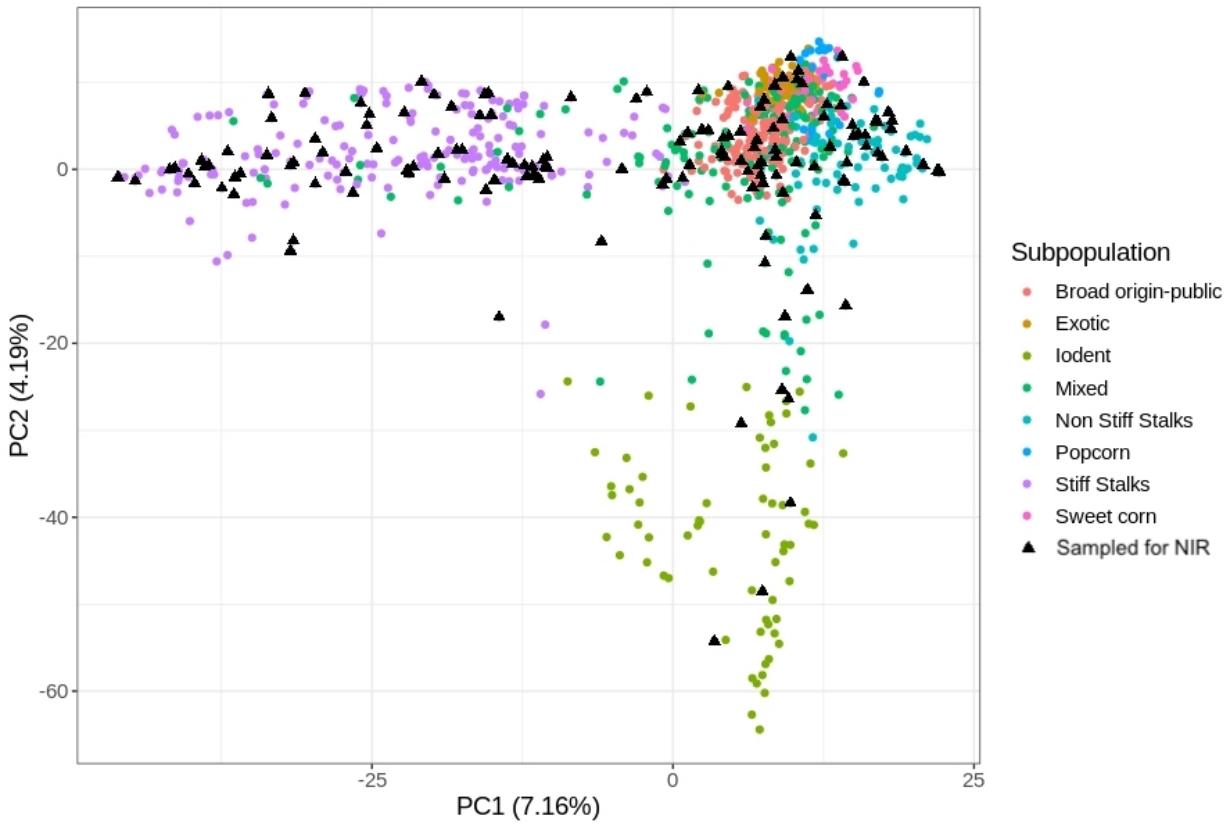
Supplemental Figure 2-1 Example of four inbred lines with contrasting endosperm vitreousness levels

Within each column the same kernel is displayed. Upper row: Image with regular settings (illumination from above). Second row: kernel displayed over a light box (illumination from below) to visualize endosperm' translucency. Third row: Longitudinally bisected kernel

Supplemental Table 2-1. Confusion matrix and performance statistics for the discrimation of maize kernel orientation.

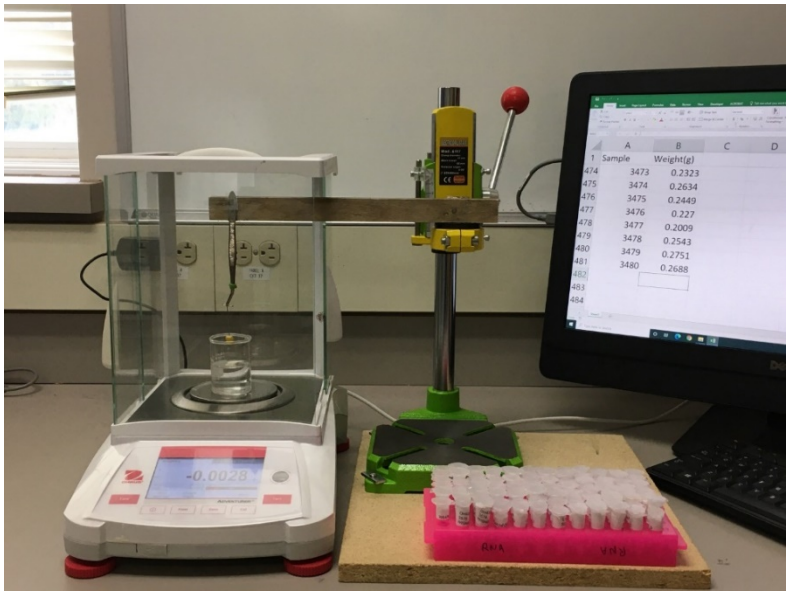
	Calibration		Prediction	
Observed/Predicted	**Germinal Side**	**Abgerminal Side**	**Germinal Side**	**Abgerminal Side**
Germinal Side	228	1	250	1
Abgerminal Side	0	251	1	228
Specificity (TNR)	1	0.996	0.996	0.996
Sensitivity (TPR)	0.996	1	0.996	0.996
Accuracy	0.998		0.996	

Specificity = [(True Negative)/(True Negative + False Positive)], Sensitivity = [(True Positive)/(False Negative + True Positive)]. Accuracy = [(Correct predictions/ Total number of predictions)].

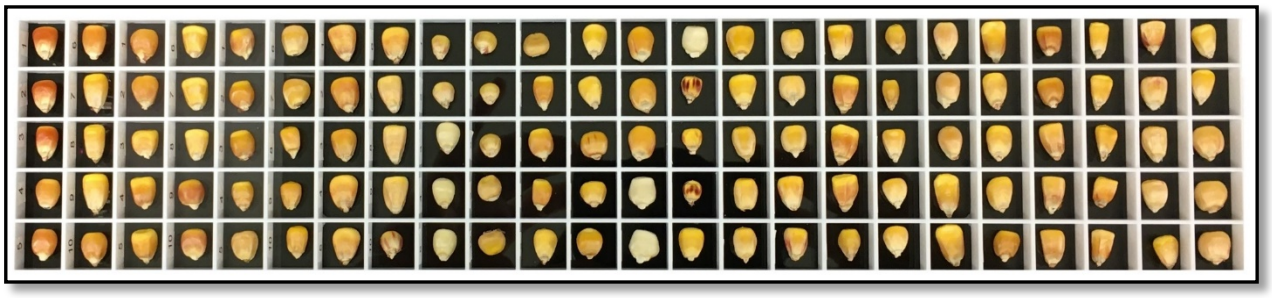


Supplemental Figure 2-2 Scatterplot of the first two principal component's scores

Principas componentes were calculated based on 5000 bi-allelic Single Nucleotide Polymorphisms (SNP) of the full Wisconsin Diversity Panel set (942 genotypes). Each color denotes a subpopulation where the genotype is classified to belong to according to Mazaheri et al., (2019) Genome-wide association analysis of stalk biomass and anatomical traits in maize. BMC Plant Biology.2019,19:45. https://doi.org/10.1186/s12870-019-1653-x. Black triangles are the 149 genotypes selected to build the vitreousness calibration curves.

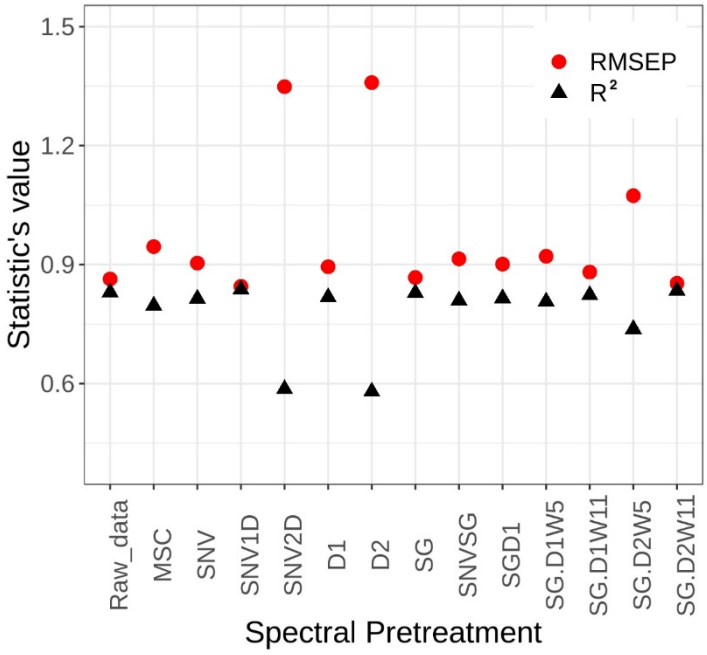


Supplemental Figure 2-3 Apparatus used to measure maize kernel volume and subsequently calculate density



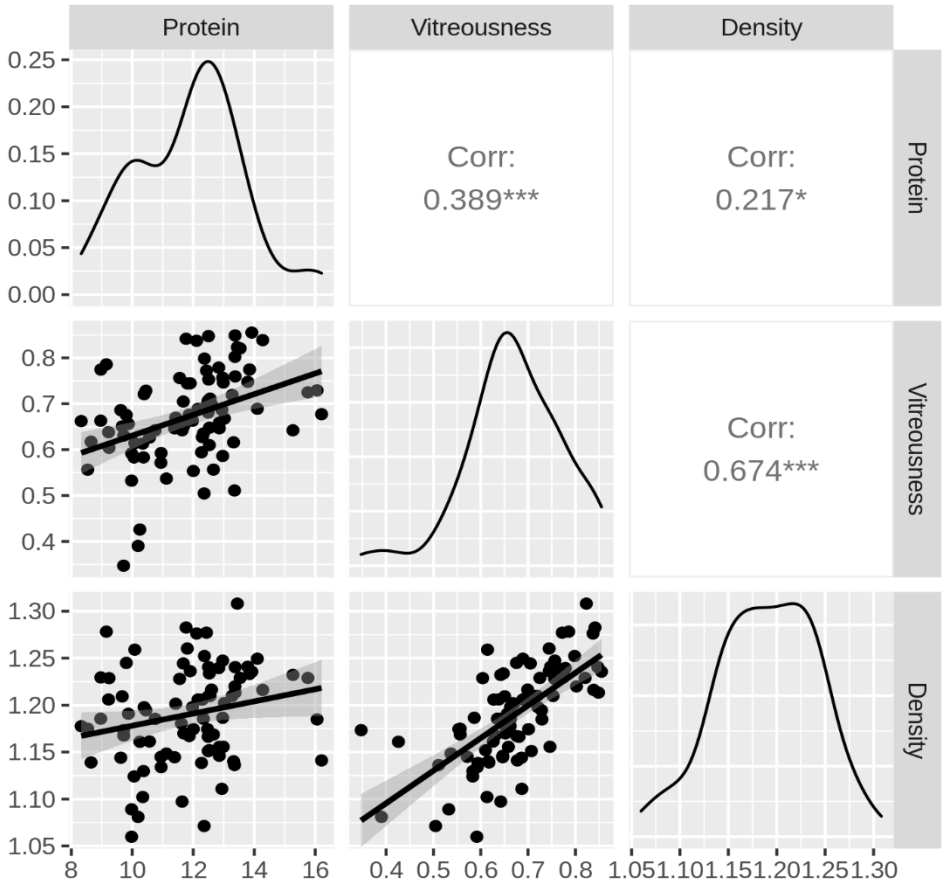
Supplemental Figure 2-4 Grid sample fixture

120 kernels per scan were arranged in 24 columns by 5 rows grid. Each cell is a square of 150mm by 150 mm of internal length. Three kernels per genotype were used for all calibrations.



Supplemental Figure 2-5 Scatterplot of distribution of two performance statistic

Scatterplot of distribution of two performance statistic (R^2 = coefficient of determination and RMSEP = Root mean square error of prediction) for each spectral pretreatment. This example shows the coefficient obtained for total kernel protein using the abgerminal side. MSC = Multiplicative scatter correction, SNV = Standard normal variate, SNVXD = Standard normal variate over the derivative where X (1^{st} or 2^{nd}) denotes the order. DX = Derivative where X (1^{st} or 2^{nd}) denotes the order. SG = Savitzky-Golay, SNVSG = Standard normal variate + Savitzky-Golay. SGD1 = Savitzky-Golay + 1^{st} derivative. SG.DXWY = Savitzky-Golay + derivative where X is the order and Y is the windows size.



Supplemental Figure 2-6 Correlation plot matrix between kernel traits

Pearson's coefficient of correlation (r) was calculated between the genotypic means of shared lines in the 3 sets. *** pval<0.001, ** pval<0.01 and * pval<0.05.

Acknowledgements

This work is based on the research supported in part by United States Department of Agriculture (USDA) grant WIS03049 to NDL and National Science Foundation grant 1940115 to EPS. The authors wish to thank Dr. Gabor Kemeny, Chris Draves, Jack Heese and Stuart Smith from Middleton Spectral Vision for their technical support with the NIR hyperspectral system. The authors are grateful to Maggie Phillips for her technical support with the C/N analyzer.

References

Barker, M., Rayens, W. (2003). Partial Least Squares for Discrimination. *Journal of Chemometrics. 17*: 166-173. Doi: https://doi.org/10.1002/cem.785.

Barrera-Arellano, D., Badan-Ribeiro, A.P., Serna-Saldivar, S.O. (2019). 'Corn Oil: Composition, Processing, and Utilization'. In Serna-Saldivar, S.O. (Eds.). *Corn: Chemistry and Technology* (pp. 539- 613). Elsevier.

Baye, T.M., Pearson, T.C., Settles, A.M. (2006). Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *Journal of Cereal Science*. *43:* 236-243.

Bergquist, R., Thompson, D. (1992). Corn grain density characterized by two specific gravity techniques. *Crop Sci*. *32*: 1287−1290.

Blighe, K. (2019). PCAtools: Everything Principal Component Analysis. R package version 1.2.0. https://github.com/kevinblighe/PCAtools

Caporaso, N., Whitworth, M.B., Fisk, I.D. (2018). Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry*. 240: 32-42. https://doi.org/10.1016/j.foodchem.2017.07.048

Cho., R.K., Iwamoto., M. (1989). The purity identification of sesame oil by near infrared reflectance spectroscopy. Proceedings of the 2nd International NIRS conference, Tsukuba, Japan.

Correa, C.E.S., Shaver, R.D., Pereira, M.N., Lauer, J.G., Kohn, K. (2002). Relationship between corn vitreousness and ruminal in situ starch degradability. *Journal of Dairy Science*. *85*: 3008-3012.

Davies, A.M.C., Fearn, T. (2006). Back to basics: calibration statistics. *SpectroscopyEurope*. *18*: 31-32.

Dias Junior, G.S., Ferraretto, L.F., Salvati, G.G.S., de Resende, L.C., Hoffman, P.C., Pereira, M.N., Shaver, R.D. (2016). Relationship between processing score and kernel-fraction particle size in whole-plant corn silage. *Journal of Dairy Science*. *99*: 2719-2729. http://dx.doi.org/10.3168/jds.2015-10411

Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods.15,*1-25. https://doi.org/10.1186/s13007-019-0476-y

Fox, G., Manley, M. (2014). Applications of single kernel conventional and hyperspectral imaging near infrared spectroscopy in cereals. *Journal of the Science of Food and Agriculture* 94: 174-179.

Gustafson, T. J., de Leon. N. (2010). Genetic analysis of Maize (Zea mays L.) endosperm vitreousness and related hardness traits in the intermated B73 x Mo17 recombinant inbred line population. *Crop Science*. *50*: 2318-2327.

Gustin, J.L., Jackson, S., Williams, C., Patel, A., Armstrong, P., Peter, G.F., Settles, A.M. (2013). Analysis of maize *(Zea mays)* kernel density and volume using microcomputed tomography and single-kernel Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry.61:* 10872-10880. http://dx.doi.org/10.1021/jf403790v

Hacisalihoglu, G., Gustin, J.L., Louisma, J., Armstrong, P., Peter, G.F., Walker, A.R., Settles, A.M. (2016). Enhanced single seed trait prediction in soybean (*Glycine max*) and robust calibration model transfer with Near-Infrared Reflectance Spectroscopy. *Journal of Agricultural and Food Chemistry*.*64*: 1079-1086. https://doi.org/10.1021/acs.jafc.5b05508

Hershberger, J., Gore, M.A. (2020). waves: Vis-NIR Spectral Analysis Wrapper. R package version 0.1.0.

Holding, D.R., Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica*. *51*: 243-254.

Jiang, H.Y., Zhu, Y.J., Wei, L.M., Dai, J.R., Song, T.M., Yan, Y.L., Chen., S.J. (2007). Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (Zea mays L.). *Plant Breeding*. 126: 492-497. doi: 10.1111/j.1439-0523.2007.01338.x.

Kennard, R.W., Stone, L.A. (1969). Computer aided design of experiments. *Technometrics*. *11:* 137-148.

Kucheryavskiy, S (2020). Mdatools – R package for chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 198: 103937. https://doi.org/10.1016/j.chemolab.2020.103937

Lee, L.C., Liong, C-Y., Jemain, A.A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*. *143*: 3526-3539.

Liland, K.H., Mevik, B-H., Wehrens, R. (2021). pls: Partial Least Squares and Principal Component Regression. R package version 2.8-0. https://CRAN.R-project.org/package=pls

Mazaheri, M., Heckwolf, M., Vaillancourt, B., Gage, J.L., Burdo, B., Heckwolf, S., Barry, K., Lipzen, A., Ribeiro, C.B., Kono, T.J.Y., Kaeppler, H.F., Spalding, E.P., Hirsch, C.N., Buell, C.R., de Leon, N., Kaeppler, S.M. (2019). Genome-wide association analysis of stalk biomass and anatomical traits in maize. *BMC Plant Biology.19*: 45. https://doi.org/10.1186/s12870-019-1653-x

McAllister, T.A., Phillippe, R.C., Rode, L.M., Cheng, K.J. (1994). Effect of the protein matrix on the digestion of cereal grains by ruminal microorganisms. *Journal of Animal Science*. *71*: 205-212.

McGoverin CM, Manley M. Classification of Maize Kernel Hardness Using near Infrared Hyperspectral Imaging. *Journal of Near Infrared Spectroscopy*. 20: 529-535. doi:10.1255/jnirs.

Miao, C., Pages, A., Xu, Z., Rodene, E., Yang, J., Schnable, J.C. (2020). Semantic segmentation of sorghum using hyperspectral data identifies genetic associations. *Plant Phenomics*. *2020*:4216373. Doi: https://doi.org/10.34133/2020/4216373.

Miller, N.D., Haase, N.J., Lee, J., Kaeppler, S.M., De León, N., Spalding, E.P. (2017). A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. *ThePlant Journal*. *89*: 169-178. Doi: https://doi.org/10.1111/tpj.13320.

Ngonyamo-Majee, D., Shaver, R.D., Coors, J.G., Sapienza, D., Correa, C.E.S., Lauer, J.G., Berzaghi, P. (2008). Relationship between kernel vitreousness and dry matter degradability for diverse corn germplasm I. Development of near-infrared reflectance spectroscopy calibrations. *Animal Feed Science Technology*.*142*: 247-258. https://doi.org/10.1016/j.anifeedsci.2007.09.023
Orman, B.A., Schumann, R. A. (1991) Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. *Journal of the American Oil Chemists' Society*. 69:1036-1038.

Orman, B.A., Schumann, R. A. (1992). Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. *Journal of the American Oil Chemists' Society*. 69: 1036-1038.

Osborne, B.G., Fearn, T., Hindle., P.H. (1993). *Practical NIR spectroscopy: with applications in food and beverage analysis*. Longman Scientific and Technical.

Philippeau, C., Michalet-Doreau, C. (1997). Influence of genotype and stage of maturity of maize on rate of ruminal starch degradation. *Animal Feed Science Technology.68:* 25-35

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Renk, J.S., Gilbert, A.M., Hattery, T.J., O'Connor, C.H., Monnahan, P.J., Anderson, N., Waters, A.J., Eickholt, D.P., Flint-Garcia, S.A., Yandeau-Nelson, M.D., Hirsch, C. N. (2021). Genetic control of kernel compositional variation in a maize diversity panel. *The Plant Genome*. e200115. Doi: https://doi.org/10.1002/tpg2.20115.

Robutti, J. L. (1995). Maize kernel hardness estimation in breeding by near infrared transmission analysis. *Cereal Chem*. 72: 632-636.

Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., & Narasimhan, G. (2020). So, you think you can PLS-DA? *BMC Bioinformatics*. *21*, 2. https://doi.org/10.1186/s12859-019-3310-7.

Sato, T., Kawano, S., Iwamoto, M. (1991). Near infrared spectral patterns of fatty acid analysis from fats and oils. *Journal of the American Oil Chemistry Society*. *68*: 827–833.

Siesler, H. W.; Ozaki, Y.; Kawata, S.; Heise, H. M. (2008). Near-infrared spectroscopy: Principles, instruments, applications; John Wiley and Sons, Inc.: New York.

Spielbauer, G., Armstrong, P., Baier, J.W., Allen, W.B., Richardson, K., Shen, B., Settles, A.M. (2009). High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chemistry*. *86:* 556−564.

Stevens, A., Ramirez-Lopez, L. (2020). An introduction to the 'prospectr' package. R package Vignette R package version 0.2.1.

Weinstock, B. A., Janni, J., Hagen, L., Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multi- variate analysis. *Applied Spectroscopy*. 60: 9-16.

Williams, P., Geladi, P., Fox, G. and Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyper-spectral imaging and multivariate data analysis, *Anal. Chim. Acta* 653: 121. doi: 10.1016/j.aca.2009.09.005.

Williams, P.J., Kucheryavskiy; S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*. 209:131–8. Doi: https://doi.org/10.1016/j.foodchem.2016.04.044

Wu, Y.V., Bergquist, R. (1991). Relation of corn grain density to yields of dry milling products. *Cereal Chemistry*. *68*: 542−544.

Xu, A., Lin, L., Guo, K., Liu, T., Yin, Z., Wei, C. (2019). Physicochemical properties of starches from vitreous and floury endosperms from the same maize kernels. *Food Chemistry.* 291: 149-156. Doi: https://doi.org/10.1016/j.foodchem.2019.04.024.

Zhao, Y., Zhu, S., Zhang, C., Feng, X., Feng, L., He, Y. (2018). Application of hyperspectral imaging and chemometrics for variety classification of maize seeds. *RSC Advances*. *8*: 1337–1345. Doi: https://doi.org/10.1039/C7RA05954J.

Conflict Of Interest

The authors declare no conflict of interest.

Chapter 3 Effect of Endosperm Type and Ensiling Time to Whole Plant Corn Silage With Different Dosages of Endosperm Modifiers on Nitrogen Fraction, Fermentation Products, Zein Profile and Starch Digestibility

Author Contributions

José Varela, Luiz Ferraretto and Natalia de León

This chapter has been formatted and will be submitted to the Journal of Dairy Sciences.

This was a highly collaborative project:

Jose Varela: Conceptualization, Formal analysis, Methodology, Validation, Writing-Original Draft. Luiz Ferraretto: Conceptualization, Writing-Reviewing and Editing. Natalia de León: Conceptualization, Funding acquisition, Supervision, Writing-Reviewing and Editing.

Abstract

Zeins are commercially important proteins found in maize endosperms. The objective of this study was to evaluate the effect of altering zeins levels in maize inbreds lines carrying endosperm mutations with differential allelic dosage and analyze its impact on the composition, nutritive value, and starch digestibility of whole plant corn silage (WPCS) at five ensiling time points. Three inbred lines carrying three different endosperm modifiers (*opaque-2*, *floury-2* and *soft endosperm-1*) were pollinated with two pollen sources to form pairs of nears isogenic lines with either two or three doses of the mutant allele for each triploid endosperm modifier. The experiment was designed as a split-plot design with three replications. Pollinated genotype was the main plot factor and ensiling time was the sub-plot level factor. Agronomic precautions were taken to mimic hybrid WPCS to the extent possible. Samples were collected at approximately 30% DM on a forage harvester and ensiled in heat-sealed plastic bags for 0, 30, 60, 120 and 240 d. Thus, the experiment consisted of 30 treatments (six genotypes x five ensiling times points) and 90 mini-silos (three replications per treatment). Measurements included nutrient analysis including crude protein (CP), soluble CP, amylase-treated neutral detergent fiber (aNDF), acid detergent fiber (ADF), lignin, starch, fermentation end products, zein concentration, and *in vitro* Starch Digestibility (ivSD).

The nutritional profile of the inbred-based silage samples was similar to hybrids values reported in literature. Both total zeins and α-zeins shows a strong negative correlation with ivSD which agrees with the general hypothesis that the degradation of hydrophobic zeins increases ruminal starch degradability. Significant differences were found in fresh (unfermented) sample kernels for endosperm vitreousness and zein profiles between and within isogenic pairs. The *o2* homozygous (three doses of mutant allele) had the highest reduction in vitreousness level (74.5 to 38%) and zein concentration (6.2 to 4.7 % of DM) compared to the heterozygous counterpart (two doses of

mutant allele). All genotypes showed significant reduction of total zeins and α-zeins during progressive ensiling times. ivSD increased with ensiling time for all genotypes with significant effect of genotype, ensiling time but not for genotype by ensiling time interaction which suggest that the ensiling period did not attenuate the difference in ivSD between near isogenic pairs caused by zeins in WPCS. Homozygous *o2* was the only mutant with significantly higher ivSD (30 and 60 days of ensiling) compared with the heterozygous version which suggest that, under all other conditions equal, substantial reductions in α-zeins are required to significantly improve ivSD in the silo.

Introduction

Greater starch digestibility on the grain fraction of Whole Plant Corn Silage (WPCS) results in increased energy availability for dairy cows and thereby greater milk production, feed efficiency, or both (Firkins et al., 2001; Ferraretto et al., 2013). The starchy endosperm contained in the kernel is protected by the pericarp, which, if intact, is highly resistant to microbial attachment (McAllister et al., 1994). Mechanical processing of the kernels breaks up the pericarp which leaves the starchy endosperm exposed for ruminal bacterial attack. However, even the exposed endosperm is not fully digested as starch granules are tightly packed and encapsulated by proteins that form what is known as the starch-protein matrix. Prolamins, also referred as zeins in maize, are the most abundant class of proteins in maize kernels with four subclasses (α, β, γ, δ) defined according to their primary structure and different solubilities (Yao et al., 2016). As zein proteins develop and distend with advancing maturity, β- and γ-zeins cross-link and α- and δ-zeins penetrate their network, thereby encapsulating starch into a hydrophobic starch-protein matrix (Mu-Forster and Wasserman 1998). Greater cross-linking of zein yields vitreous endosperm, which has been extensively studied due to the negative effect of this type of endosperm on starch digestibility. Kernels with higher vitreousness percentage have reduced starch degradability *in vitro* or *in situ* (Philippeau and Michalet-Doreau, 1998; Ngogyamo-Majee et al., 2009; Lopes et al., 2009) and *in vivo* (Philippeau and Michalet-Doreau, 1998; Taylor and Allen, 2005) compared with floury endosperm types.

Comparisons of WPCS produced from maize hybrids with different vitreousness levels provide valuable information about the overall effect of the kernel over feed digestibility and kinetics but the effect of specific endosperm type are still susceptible to be confounded with the effect of the genotype as a whole (forage + grain). Different genotypes will likely have differences in plant

biomass distribution (*e.g.,* plant reproductive to vegetative ratio *, i.e.,* harvest index (HI)), pericarp thickness, intrinsic proteolytic machinery (plant enzymes or bacterial proteases), nutritional profile or other traits that could directly alter or at least interact with starch digestibility and/or the fermentation kinetics in case of long-term ensiling experiments. This probable inter relationships hampers the ability to establish the effect of specific endosperm type on starch digestibility of WPCS.

Near isogenic lines with endosperm modifiers have been used to assess the effect of endosperm types on digestibility for isolated grain portion for *in situ* and *in vitro* measurements (Ngonyamo-Majee et al., 2009) or as isolated dry rolled kernels of hybrids as part of an experimental diet *in situ* and *in vitro* (Lopes., et al 2009). The first objective of this work is to study the effect of endosperm vitreousness over the nitrogen fraction, prolamins content and in *in vitro* starch digestibility (ivSD) using WPCS made of near isogenic inbred lines with differential endosperm allelic dosage of the modifiers to trigger changes in endosperm prolamins and texture. Using controlled pollinations, we aim to alter specific endosperm traits such as vitreousness and zeins levels while keeping the non-grain portion of the WPCS constant.

During ensiling, the hydrophobic starch-protein matrices that surrounds starch granules are broken down, making starch more accessible for degradation by ruminal microorganisms (Kotarski et al., 1992). This response is intensified with prolonged storage (Kung et al., 2018). Several mechanisms are responsible for this proteolytic activity in the silo, including kernel proteases, microorganisms, and solubilization by fermentation acids (Simpson, 2001; Junges et al., 2017). Although it is well known that extended ensiling improves ruminal ivSD, the effect of ensiling time on ivSD and zeins concentration on WPCS produced from maize genotypes with contrasting endosperm vitreousness is not well understood.

The second objective of this work is to assess fermentation profile, soluble Crude Protein (% of CP), ammonia-N (% of N), ivSD, and zein degradation kinetics in WPCS of near-isogenic lines with dissimilar vitreousness and zeins content over five ensiling times.

Materials And Methods

Silage production

WPCS material was generated using inbred lines and endosperm modifiers. Each inbred carried one endosperm mutation. The lines selected were W64A*o2o2* (*Opaque 2*), Oh43*fl2fl2* (*Floury 2*) and Oh43*h1h*1 (*Soft endosperm*).

Each of the inbred lines was treated with two different pollination regimes which resulted in six different silage sources.

Plots were differentially pollinated with the objective of generating pairs of silage sources with different endosperm characteristics maintaining the same non-grain component. To achieve that, half of the plots from each original inbred line were self-pollinated to attain maximum allelic dosage of the endosperm modifier with homozygosity, and the other half was cross pollinated with pollen from the isogenic wild-type version of each corresponding line (W64A and Oh43, respectively) which results in two copies of the maternal endosperm mutant allele and one paternal copy of the wild type allele. After pollination, each resulting silage source is considered a unique "silage genotype" that is identified according to the final endosperm allelic dosage of the modifier (Self-pollinated: W64A*o2o2o2*, Oh43*fl2fl2fl2*, Oh43*h1h1h1*. Cross pollinated: W64A*o2o2wt*, Oh43*fl2fl2wt*, and Oh43*h1h1wt*). Since all the measurements were performed after harvest, throughout the manuscript the six different pollinated inbred lines will be referred to collectively

as “genotypes” and the pairs of inbred lines carrying the same maternal endosperm modifiers, but different paternal contribution will be referred as “near isogenic pairs”.

Each silage genotype was planted in triplicates (three blocks) in four row plots, 4.8 m long and 0.75 m between rows to a resulting planting density of 83,300 plants ha^{-1} at the University of Wisconsin-Madison West Madison Agricultural Research station in the spring of 2019. Isogenic wild-type versions of Oh43 and W64A were planted as borders with two delayed planted sections in the back of the experiment to provide pollen for the cross-pollinated plots. Once the maize plots started showing signs of transitioning to reproductive growth stage ear shoots were covered after emerging from the stalk and prior to silk appearance to avoid uncontrolled pollination. At anthesis, each ear was pollinated three consecutive days, providing excess of pollen during silk elongation to ensure optimal ear filling and minimizing the possibility of Harvest Index (HI) reduction due to lack of fertilization. Shoot bagging, tassel bagging and pollination was performed without removing any leaves from the plant nor breaking the tassel to mimic commercial growing conditions in silage production fields as closely as possible.

At harvest, each ear was visually inspected by pulling the husks, and if the ear was found not to be fully pollinated the whole plant was removed. When plants reached approximately 30% dry matter (DM) the experiment was harvested using a research plot grade self-propelled forage harvester (JD 5830, John Deere, Moline, IL) equipped with a plot harvest sampler (Cibus TRM, Wintersteiger, Ried im Innkreis, Austria). Whole-plants were chopped at a theoretical lengths of cut (TLOC) of 1.9 cm. A full 27-gallon plastic container of WPCS of each chopped plot was immediately taken to the lab where samples of each plot were taken to form the ensiling time treatments. Five-600g samples of chopped WPCS were allotted from each plot and vacuum sealed in nylon-polyethylene standard barrier vacuum pouches (3.5-mil thickness, 25.4 x 35. cm; Doug

Care Equipment Inc., Springville, CA) using an external clamp vacuum machine (Pro-3000; Weston., Southern Pines, NC). Each of the bags or "mini-silos" were stored at room temperature (approximately 20°C) in the dark for 0, 30, 60, 120 or 240 days. After the ensiling time was reached, the bags were immediately frozen and stored at -20C to stop fermentation until they were processed for analysis. All bags, including those ensiled for 240 days, were frozen for at least 21 days to ensure protocol consistency across samples. The experiment consisted of six genotypes x three blocks (reps) x five ensiling time points, which generated 90 mini-silos bags. Mini-silos are the experimental units.

Fermentation Profile, Nutrients and Digestibility Analysis of forage samples.

After all ensiling treatments were completed, including the 21 days treatment homogenization period in the freezer, each bag was opened and subdivided in 200g subsamples for the different wet lab analysis. Nitrogen fraction, fermentation end products and nutrients profile were performed by Dairyland Laboratories Inc (Arcadia, WI). Samples were analyzed for Dry Matter (DM) (oven-drying in forced air at less than 60C until samples contains less than 6% moisture followed by NFTA Method 2.1.4 - Dry Matter by oven drying for 3 hours at 105C), Crude protein (CP)(Combustion method 990.03; AOAC 16th Edition), borate-phosphate buffer soluble CP (procedure described by Licitra et al., 1996), Acid Detergent Fiber (ADF)(method 973.18 AOAC, 1996. Modification includes use of Sea Sand for filter aid as needed), Amylase-Treated Neutral Detergent Fiber (aNDF)(method 2002.04; AOAC, 2005; Modification includes use of Sea Sand for filter aid and Whatman GF/C filter paper for residue collection), Starch (method 2014.10; AOAC, 2014) , *In vitro* Starch Digestibility 7h (IVSD7) (procedure described by

Richards et al., 1995), Volatile Fatty Acids, Lactic Acid and Ethanol (HPLC following Canale et al., 1984).

Kernel sample Collection and Analysis

On the day of harvest, five random plants were selected per plot and excluded from harvest. Ears from these plants were air forced dried at 40°C for three days, then shelled, bulked and ten representative kernels were used to measure vitreousness by manual dissection as described by Correa et al., 2002.

A 200g subsample from each minisilo was used for grain prolamin analysis. The grain portion of each sample was manually separated from the non-grain component, dried at 55°C for 48 hours and then grinded as powder (<1mm) using a Mill grinder (3010-030 Cyclone sample mill, Udy Corporation, Fort Collins, CO). Ground dry kernels were sent to the University of Wisconsin Soil and Forage Analysis Laboratory (Marshfield, WI) for prolamin quantification following the Larson and Hoffman, (2008) protocol.

Statistical Analysis

The experimental design was a Split-plot design where each pollinated genotype plot is the whole-plot factor and ensiling time is the split-plot factor. Data were analyzed using the function "*lmer*" from the package "lme4" (Bates et al., 2015) in R (R core Team, 2019). The linear model fitted was:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ij} + \tau_k + (\alpha\tau)_{ik} + \delta_{ijk}$$

where:

μ = is the population grand mean. α_i = is the fixed whole plot treatment effect of Genotype *i*. β_j = is the random effect of Block *j*. e_{ij} = error term for variation among Genotype. $e_{ij} \sim N(0,\sigma^2_e)$. τ_k = is the fixed subplot treatment effect for Ensiling Time *k*. $(\alpha\tau)_{ik}$ = is the fixed interaction effect between Genotype and Ensiling Time. δ_{ijk} = error term for variation among Ensiling times (Subplot error), $\delta_{ijk} \sim N(0,\sigma^2_\delta)$.

Factor means were determined using Least Square Means and the package "*emmeans*" (Lenth, 2021). Pairwise comparisons were computed using the "*emmeans*" function and the Dunn-Šidák correction was used to counteract the problem of multiple comparison.

When comparing traits at different ensiling times the samples at day 0 were excluded from the model to avoid confounding of ensiling time effects and interactions.

Orthogonal polynomials contrasts were used to evaluate linear and quadratic responses to ensiling time from 30 to 240 d using the function "*contrast*" in the "*emmeans*" package. Because treatments were unequally spaced, contrast coefficients were calculated separately using the function "poly" in R.

Because endosperm vitreousness can only be measured in intact unfermented kernels, this was measured only at ensiling time 0. Likewise, lignin was only measured in unfermented samples, therefore the experiment for these two traits was treated as a Randomized Complete Block Design as there was no subplot (ensiling) factor used.

The reduced linear model for vitreousness and lignin was:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ij}$$

μ = is the population grand mean. α_i = is the fixed treatment effect of Genotype *i*. β_j = is the fixed effect of Block *j*. e_{ij} = error term for variation among Genotype, $e_{ij} \sim N(0,\sigma^2_e)$.

Results and Discussion

Fresh Samples Chemical and Physical Characteristics of Inbred Lines

The average harvest DM % across all plots for the six genotypes evaluated was 28.1% which is considered on the lower limit for optimum moisture for maize WPCS. There was no difference in DM% within isogenic pairs which is expected as the primary contributor of water at harvest is the non-grain portion of the plant. No differences were observed among genotypes for overall CP with an average of 9.86% of DM (Table 3-1).

Soluble CP showed significant differences within the isogenic *o2* pair but not for the other two pairs (Table 3-1). Tsai et al., 1978 reported that *o2* causes about 50% reduction in zein proteins compared to wild type. This type of proteins is innately insoluble in rumen environment (Larson and Hoffman, 2008), therefore, the difference in Soluble CP observed for the two unfermented isogenic o2 samples is likely to be a consequence of reduced zeins of the homozygous version.

aNDF and ADF values (as % DM) were within the expected average range for WPCS hybrids commonly used and those previously reported in the literature including a meta-analysis of the effects of different WPCS hybrids on various fermentation and lactation performance indicators that found on average NDF (% of DM) of 41.5 and ADF (% DM) of 24.2 across 126 and 107 experiments, respectively (Ferraretto and Shaver, 2015). Most of the fiber constituents including cellulose, hemicellulose and pectin are provided by the plant cell wall through the non-grain portion of the mix, with kernels having the lowest concentration of cell wall content compared to other tissues (Coors and Lauer, 2001). This relationship can help explain the high similarity of the fiber parameters within isogenic pairs observed in our study (Table 3-1).

Starch concentrations were low compared to hybrid-based WPCS but not outside the boundaries of possible values for hybrid maize. Low concentration of starch found in our inbred based WPCS

is most likely attributed to a lower Harvest index (HI) of inbred lines compared to heterotic hybrids. Despite significant efforts to exclusively harvest plants with fully pollinated ears, there is still an intrinsic disproportion of reproductive to vegetative tissue ratio between maize inbreds and hybrids probably due to the greater selection pressure that hybrids have been subjected to maintain high HI under modern production methods.

The protein fraction, fiber, lignin and starch levels of the inbred lines found in this study are similar to commonly reported commercial hybrids used for silage performance testing. With the data shown, we propose that under controlled conditions, inbred-based silage systems can be used as reliable models in research to investigate genetically controlled traits in a more effective way than using hybrids.

Endosperm Vitreousness and Prolamins Levels in Unfermented Kernels

The homozygous *o2* kernels had the highest effect on reducing endosperm vitreousness to half the level of its heterozygous counterpart (Table 3-2). This large effect was expected as this recessive mutation has been long and widely characterized to be a strong endosperm texture modifier. The *o2* gene encodes a transcriptional activator that regulates the expression of a number of genes (Schmidt, 1993), including the 22-kD α-zein protein (Schmidt et al., 1990) which concentration in the endosperm is positively associated with vitreousness levels. No significant difference in vitreousness level was observed among the *fl2* isogenic pairs (Table 3-2), with both dosage combinations showing extremely soft endosperm type. *fl2* is a semi-dominant mutation that results from the expression of an abnormal 22-kD α-zein that interferes with protein body assembly (Holding and Larkins, 2006). Unlike *o2* which is fully recessive, *fl2* shows a partial dosage effect

in which two doses of the mutant gene in the endosperm are needed to produce a mutant floury phenotype as is the case in its homozygous state (Jones 1978; Mertz et al., 1964).

Significant decrease of endosperm vitreousness was found for the homozygous *h1* genotype compared to its heterozygous counterpart. This mutation is characterized as generating a "soft endosperm", but to our knowledge it has not been fine mapped nor cloned therefore our expectations of a heterozygous genotype were unknown, as the molecular mechanism of this mutation remains to be elucidated.

Significant reduction in total zein content was observed in the *o2* homozygous genotype compared to the heterozygous counterpart (Table 3-2). This reduction was particularly severe in α-zeins as expected due to the severe reduction of the 22-kD alpha zein family genes causes by the *o2* mutation. No difference was found for total zein content between the *fl2* near isogenic pair, but significant differences were found in α-zeins levels. Contrary to expectations, the homozygous line had higher values of α-zeins than the heterozygote counterpart. For the purpose of this research, this unexpected inversion does not affect the methods, as we are interested in capturing endosperm differences regardless of which allelic dosage combination is the highest or lowest. The *h1* Isogenic lines didn't have significant difference for any of the zein type comparisons (Table 3-2).

The three isogenic pairs analyzed in this experiment shows different associations levels between zeins and vitreousness (Table 3-2). For the *o2* isogenic pair, the homozygous exhibited the well reported response of decreased α-zeins levels associated to a significant reduction in vitreousness, which has been previously reported in several other endosperm mutants such as *Defective endosperm B30* (*De-B30*) and *Mucronate* (*Mc*) (Holding and Larkins, 2006). The *fl2* isogenic pair showed no changes in endosperm vitreousness but a slight increased level of α-zeins was observed

for the homozygous line. Finally, the homozygous *h1* exhibited significant decrease in endosperm vitreousness compared to the heterozygous with no significant change in any of the zein concentrations. Although there is a well described group of endosperm mutations that generate a vitreous endosperm through reduced zeins, the actual connection between vitreousness and endosperm types is just now being elucidated in a limited number of mutants and much remains to be learned from most endosperm mutants. There is a group of endosperm opaque mutations such as O*paque1 (o1) and Floury1 (1)* that does not have reduced zein accumulations (Holding, 2014).

Fermentation Profile and Nitrogen Fraction During Ensiling

Ensiling time was significant for pH ($p = 1.01x10^{-06}$, Table 3-3). The pH of time 0 forage samples rapidly decreased from 4.74 at day 0 to an average of 3.67 at 30 days of ensiling and then increased only slightly from there to 3.71 at 240 days of ensiling (Table 3-4). The narrow range of pH values among isogenic lines is perhaps an indicator of a correct acid balance during ensiling and the similarity among genotypes in terms of composition and particularly HI.

Lactic acid (% DM) was found in the normal range for ensiled WPCS with no significant effect of ensiling time (p = 0.09), nor ensiling time by genotype interaction ($p = 0.59$, Table 3-3) in agreement with the values found by Der Bedrosian et al., 2012 using hybrid maize. Figure 2B shows that no difference was observed between genotypes in the first three ensiling times (30-60-120) but significant differences were detected at 240 days.

Ensiling time and genotype had a significant effect on acetic acid concentration ($p = 4.03x10^{-06}$ and 0.0026 respectively), but no interaction was detected. A general increase trend agrees with values reported by Der Bedrosian et al., 2012. The *fl2* isogenic pair showed higher levels of acetic

acid than the other two pairs. This difference tended to increase more substantially on the later ensiling times (120-240 Figure 3-1C). The *h1* and *o2* isogenic pairs behaved similarly across ensiling periods with no significant difference for any particular ensiling time and always below 2%, which is the optimum recommended for maize silage. Overall, we observed significant effect of ensiling time, genotype and the interaction for ethanol concentration (Table 3-3) with an upward trend towards longer silage times (Figure 3-1D).

No significant difference within isogenic pairs were found at any time of ensiling for acidity, organic acids or ethanol which supports the notion that the allelic dosage treatment applied here didn't interfere with the fermentation performance of the genotypes. These results provide a baseline to establish appropriate comparisons of prolamins degradations and digestibility.

Ammonia-N and soluble CP steadily increased during ensiling time with significant effect of ensiling time in both components ($p<2x10^{-16}$ and $p<2.2x10^{-16}$ respectively). Increased concentration of soluble CP and ammonia-N with prolonged ensiling time were previously reported in the literature (Der Bedrosian et al., 2012; Young et al., 2021) and suggest the occurrence of proteolysis or solubilization of proteins (Hoffman et al., 2011).

Misra and Oaks, 1981 reported increased level of ammonia in homozygous o2 maize compared to a wild type near isogenic line. Although not statistically significant, it's interesting to note that a similar trend was observed for Ammonia-N in our experiment at 0-30-60 and 120 days of ensiling (Figure 3-1E) when comparing the o2 near isogenic pairs.

The observed increase in soluble CP of homozygous *o2* compared to its heterozygous at all ensiling times (Figure 3-1F) might be an indirect consequence of the effect of the mutation on endosperm protein balance. The extensively reported reduction synthesis in non-soluble α-zeins proteins of *o2* mutants comes accompanied by a series of changes including an increased synthesis of soluble

non-zeins proteins. It has been suggested that the increase of non-zein proteins in *o2* endosperm is a consequence of the diversion of N from the zein to the non-zein fraction (Habben et al., 1993). Propionic and butyric acid were below the limit of detection (< 0.01 % DM basis) for all samples at all ensiling times (Data not shown) which is a good indicator of well-fermented silage, probably free of clostridial fermentation (Kung et al., 2018). This was particularly important for this experiment as it is known that forages ensiled in high moisture have a greater risk for clostridial growth. Methanol, 1-propanol, 1,2-propanediol and 2-butanol were below the limit of detection (< 0.01 % DM basis) for all samples at all ensiling times (Data not shown).

Overall, the fermentation values obtained in this study show that under controlled conditions, inbred-based WPCS behaves similarly to hybrid systems.

Prolamin Degradation and Starch Digestibility

Significant differences were found among genotypes, ensiling times and genotype by ensiling time interaction for total zeins (Table 3-3). All genotypes decreased total zeins with increased ensiling time (Figure 3-2A), which agrees with the hypothesis that zein proteins, which form the starch-protein matrix, are degraded in the ensiling process (Philippeau and Michalet-Doreau, 1998; Jurjanz and Monteils, 2005). Although this hypothesis is widely accepted, most of the supporting evidence is based on measurements of starch degradability over ensiling time (instead of zein fraction) which is assumed to be mainly limited by the undegraded starch-protein matrix. Hoffman et al., 2011 compared the zein fraction of two maize hybrids before and after 240 days of ensiling in the context of a high moisture corn (HMC) where only the kernel was ensiled. To our knowledge this is the first time that zeins degradation curves over multiple ensiling times are reported using WPCS rations. The hydrophobic zein proteins in the starch-protein matrix could potentially be

degraded in the ensiling process by solubilization or by proteolytic activity (Hofmann et al., 2011). The largest difference between genotypes of isogenic pairs was found for *o2*, where significant differences were detected at 60, 120 and 240 days after ensiling (Figure 3-2A). The differences in total zeins between the *o2* isogenic pairs seems fairly constant across ensiling times, with almost parallel lines between sampling times. The average difference in total zeins between the *fl2* isogenic lines had intermediate values compared with the others two pairs across ensiling times. No difference in total zeins at any ensiling point was found between the *h1* isogenic lines. Significant effect of genotype, ensiling time and genotype by ensiling time interaction was observed for α - zeins (Table 3-3). Its concentration decreases at advanced different ensiling times similarly to total zeins (Figure 3-2B and 3-2A). The *o2* and *fl2* isogenic pairs showed significant α - zeins difference at all ensiling stages with remarkable parallel trends within pairs. Although significant, genotype by ensiling time interaction was slightly below the significance threshold ($p = 0.04$ Table 3-3) and most likely that probability value was strongly influenced by the crossing over between the W64*o2o2o2* and the *h1* isogenic pairs. The parallel response within isogenic lines with significantly different total zeins and α-zeins its important as it suggests that the differential efficiency of starch degradation inside the rumen will not be compensated by ensiling time, assuming that the negative effect in starch degradation is imposed by the prolamin matrix. Significant genotype and ensiling time effect were found for iVSD (Table 3-3). Only the *o2* isogenic pair had significant difference at 30 and 60 days of ensiling (figure 3-2C). Even with significant differences in α-zeins across ensiling points, the *fl2* isogenic lines were not different in terms of iVSD which is suggesting, that under equal conditions of comparisons, strong differences in zeins profiles are required to generate a significant impact in starch digestibility. Our results indicate that, for the mutations and dosages used in this experiment, the levels of α-zeins plays a

more preponderant role in starch ivSD than the kernel vitreousness endosperm level. This can be seen by comparing the α-zeins and vitreousness level of the three isogenic pairs and their respective ivSD values. The *h1* isogenic pair had significant difference in endosperm vitreousness (Table 3-2), no difference in α-zein levels and therefore no difference in ivSD performance at any ensiling time. Contrastingly, the *fl2* isogenic pair, which had no vitreousness differences but intermediate level of a-zeins difference showed an intermediate ivSD separation on average compared to the other two pairs, although not significant (Fig 3-2C). Finally, *o2* showed the maximum separation having both, high vitreousness differences and the highest α-zein differences within the pairs. These inferences about the effect of endosperm vitreousness and ivSD should be bounded to this specific situation as the preparation for this technique implied grounding at 4mm prior to exposure to the ruminal fluid, which might have attenuated the differences in ivSD of a vitreous endosperm.

The six genotypes had significant differences between day 0 and 30 for ivSD (Figure 3-2D, 3-2E, 3-2F). with the highest slopes in all cases compared to posterior periods. Although all the genotypes increased iVSD between 30 and 240 days of ensiling, only the two *fl2* near isogenic lines were able to surpass the significant threshold. Genotype by ensiling time interaction was not significant ($p = 0.49$, Table 3-3) for iVSD which indicates that fermentation in the silo did not eliminate the difference caused by floury and vitreous endosperm or high or low level of zeins. Similar results were obtained by Ferraretto et al., 2015.

Both total zeins contents and α - zeins showed strong negative correlations with ivSD (Figure 3-3A and 3-3B $r = -069$ and $r = -0.72$, respectively) which agrees with the general hypothesis that the degradation of hydrophobic zeins proteins increases ruminal starch degradability by increasing accessibility of starch granules to rumen microorganisms (Philippeau and Michalet-Doreau, 1998).

iVSD linearly increased (Figure 3-4C and 3-4D) with soluble CP ($r = 0.76$; $p = 3.29x10^{-18}$) and ammonia-N ($r = 0.67$; $p = 7.59x10^{-13}$) which agrees with Ferraretto et at., 2015. The published study compared eight hybrids using the same digestibility assessment used in this work (4-mm ivSD). Similar results were observed for HMC and thought to be related to the lower concentration of zein-proteins associated with greater ammonia-N and soluble-CP concentration after 240 days of ensiling (Hoffman et al., 2011). These data suggest that ammonia-N and soluble CP are good indicators of starch digestibility and may be used in future models as predictors of starch digestibility as proposed by Ferraretto et al., 2015.

Conclusions

We have showed that the use of maize inbred lines is a valid method for WPCS to model fermentation and digestibility changes of maize during ensiling. This is particularly useful to answer mechanistic question using genetics as a source of controlled variation in a simpler way than mixing genetic background to generate hybrids. In the present study, three endosperm mutations combined with two different allelic dosages were used to trigger different endosperm characteristic. The novelty of this method is that it allows the characterization of the true impact of altering zeins and endosperm vitreousness levels while maintaining the non-grain fraction of the ensiling ration unaltered.

The *o2* near isogenic pair showed the highest difference in zeins profile and was the only pair able to significantly show differences in starch digestibility. The *fl2* near isogenic pair was not able to show iVSD differences, even with significant different levels of α-zeins. All genotypes decreased total zeins and α-zeins with ensiling time as expected. Interestingly the two pairs with significant difference at all ensiling points (*o2* and *fl2*) showed a parallel response in terms of zein degradation, which would suggest that extended ensiling times will not alter the negative effects of zeins on ivSD. The present study shows that for 4mm ground ivSD, α-zein levels plays a more preponderant role than vitreousness in terms of starch digestibility. Further research with different grinding sizes, vitreousness and α-zeins combinations would help a more comprehensive understanding of the interaction of these three factors.

The present study shows the first continue measurement of zeins during ensiling times. Soluble protein and ammonia-N concentrations increased through ensiling times for all genotypes supporting the idea that extended fermentation time is beneficial for dairy farmers independently of the hybrid type (Ferraretto et al., 2015). Finally, total zeins, α-zeins, ammonia-N and soluble

CP were all significantly correlated to ivSD, which implies that can be used in future models instead of ivSD, reducing labor and cost for commercial laboratories and dairy farmers.

Tables

Table 3-1 Nutrient composition and fermentation profile of unfermented WPCS of the six genotypes

	W64Ao2o2*WT*	W64Ao2o2o2	Oh43fl2fl2*WT*	Oh43fl2fl2fl2	Oh43h1h1*WT*	Oh43h1h1h1	SE
Nutrient							
DM (% of as fed)	30.19**b**	28.6**ab**	26.5**a**	27.37**a**	28.15**ab**	27.6**a**	0.59
CP (% of DM)	10.07	9.92	9.81	9.65	9.85	9.85	0.15
Soluble CP (% CP)	38.29**a**	46.98**b**	45.26**b**	42.43**ab**	45.65**b**	44.91**b**	1.11
aNDF (% DM)	40.73	43.15	38.17	41.64	40.98	42.99	1.19
ADF (% DM)	22.84**ab**	24.29**b**	21.69**a**	22.22**ab**	23.82**ab**	23.89**ab**	0.6
Lignin (% DM)	2.63	2.64	2.45	2.51	2.08	2.31	0.19
Starch (% DM)	24.04**b**	19.32**a**	22.56**ab**	21.64**ab**	21.45**ab**	20.07**a**	0.89
Fermentation Profile							
pH	4.71**ab**	4.51**ab**	4.8**b**	4.39**a**	4.6**ab**	4.81**b**	0.09
Ammonia-CP (% of CP)	2.09	2.74	1.64	1.71	2.07	2.03	0.32

DM = Dry Matter, CP = Crude Protein, aNDF = ash-free Neutral Detergent Fiber and ADF = Acid detergent fiber.

[a-c] Means in the same row with different letters differs ($p<0.05$). SE = Standard Error

Table 3-2 Endosperm vitreousness, crude protein and zein profile of unfermented kernels of the six genotypes

	W64Ao2o2*WT*	W64Ao2o2o2	Oh43fl2fl2*WT*	Oh43fl2fl2fl2	Oh43h1h1*WT*	Oh43h1h1h1	SE
Vitreousness (% of endosperm)	74.5 **c**	38 **a**	31.5 **a**	33 **a**	56.5 **b**	35 **a**	1.39
CP (% of DM)	13.23 **b**	12.76 **ab**	11.93 **a**	12.12 **a**	12.09 **a**	12.26 **ab**	0.26
Zein Protein profile							
Total zein (% of DM)	6.2**b**	4.7**a**	4.82**ab**	5.15**ab**	5.32**ab**	4.66**a**	0.32
Total zein (% of CP)	46.93	36.9	40.3	42.6	44.02	38.09	3.4
α zeins (% of DM)	4.04 **c**	2.44 **ab**	2.35 **a**	3.17 **b**	2.77 **ab**	2.59 **ab**	0.17
α zeins (% of CP)	30.57**b**	19.17**a**	19.7**a**	26.25**ab**	22.89**a**	21.22**a**	1.73
α zeins (% of zeins)	65.26 **c**	51.86 **ab**	49.03 **a**	61.53 **bc**	51.97 **ab**	55.59 **abc**	2.4
β, γ, δ zeins (% of DM)	2.16	2.26	2.46	1.98	2.55	2.03	0.2
β, γ, δ zeins (% of CP)	16.36	17.73	20.6	16.37	21.13	16.86	2.18
β, γ, δ zeins (% of zeins)	34.73**a**	48.17**bc**	51**c**	38.5 **ab**	48.03 **bc**	44.43 **abc**	2.4

DM = Dry Matter, CP = Crude Protein.[a-c] Means in the same row with different letters differs ($p<0.05$). SE = Standard Error.

Table 3-3 Statistical analysis (p-values) of the effects of genotype, ensiling time and genotype by ensiling time on nutrient composition, fermentation profile, starch digestibility and prolamin fraction in whole plant corn silage (WPCS) for a set of six genotypes with four ensiling times (30-60-120-240).

	Source of variation		
Trait	Genotype	Ensiling Time	Genotype by Ensiling Time Interaction
pH	0.35	1.01E-06	0.38
Lactic Acid (%DM)	0.03	0.09	0.59
Acetic Acid (%DM)	2.56E-03	4.03E-06	0.28
Ethanol (%DM)	0.05	<2.2E-16	0.02
Ammonia-N (% of CP)	0.10	<2.2E-16	0.02
Soluble CP (% of CP)	6.80E-07	<2.2E-16	0.03
CP (% of DM)	0.03	0.12	0.29
Starch (% of DM)	0.01	0.50	0.68
ivSD (% of starch)	3.30E-03	1.41E-06	0.47
Soluble CP (% of DM)	1.08E-11	< 2.2E-16	0.27
Total acids (% of DM)	0.02	0.03	0.54
Ammonia-N (% of DM)	0.13	< 2E-16	0.02
Grain CP (% of DM)	3.14E-05	< 2.2E-16	0.13
Total Zeins (% DM)	6.59E-04	< 2.2E-16	1.68E-03
α- Zeins (% DM)	3.63E-06	< 2.2E-16	0.04
α- Zeins (% CP)	3.80E-05	1.64E-08	6.30E-04
α- Zeins (% Zeins)	2.35E-05	0.254766	1.03E-03
$\beta\gamma\delta$- Zeins (% DM)	0.02	5.26E-13	7.50E-04
$\beta\gamma\delta$- Zeins (% CP)	0.01	1.80E-04	6.10E-04
$\beta\gamma\delta$- Zeins (% Zeins)	2.35E-05	0.26	1.00E-03

DM = Dry Matter, CP = Crude Protein.

Table 3-4 Effect of storage length on nutrient composition, fermentation profile, starch digestibility and prolamin fraction in whole plant corn silage (WPCS) for a set of six genotypes with five ensiling times (0-30-60-120-240) replicated three times.

							P-value*		
Trait	0d	30d	60d	120d	240d	SE	F	E	L	Q
pH	4.74	3.67 **a**	3.68 **ab**	3.69 **b**	3.71 **c**	0.006	<.0001	1.01E-06	<.001	0.9813
Lactic Acid (%DM)	1.26	6.46 **a**	6.85 **a**	6.88 **a**	7.1 **a**	0.228	<.0001	0.093	0.19	0.033
Acetic Acid (%DM)	0.365	1.61 **a**	1.76 **ab**	1.9 **bc**	1.95 **c**	0.052	<.0001	4.03E-06	<.001	0.0049
Ethanol (%DM)	0.201	0.55 **a**	0.69 **a**	1.04 **b**	1.69 **c**	0.064	<.0001	<2e-16	<.001	0.99
Ammonia-N (% of CP)	2.05	4.16 **a**	5.03 **b**	6.23 **c**	7.38 **d**	0.144	<.0001	<2e-16	<.001	0.0002
Soluble CP (% of CP)	43.9	57 **a**	62.6 **b**	65 **c**	67.6 **d**	0.36	<.0001	<2.2e-16	<.001	<.001
CP (% of DM)	9.86	10.1	10.3	10.3	10.3	0.064	0.0246	0.125	0.12	0.18
Starch (% of DM)	21.5	21.7	21.5	22.2	21.7	0.34	0.99	0.5	0.86	0.24
ivSD (% of starch)	53.6	62.6 **a**	62.4 **a**	65.8 **b**	68.4 **b**	0.88	<.0001	1.41E-06	<.001	0.61
Soluble CP (% of DM)	4.32	5.78 **a**	6.45 **b**	6.69 **c**	6.98 **d**	0.06	<.0001	< 2.2e-16	<.001	<.001
Total acids (% of DM)	1.62	8.07 **a**	7.89 **ab**	8.82 **ab**	9 **b**	0.274	<.0001	0.02571	0.0398	0.0231
Ammonia-N (% of DM)	0.202	0.42 **a**	0.52 **b**	0.64 **c**	0.76 **d**	0.014	<.0001	< 2e-16	<.0001	<.0001
Grain CP (% of DM)	12.4	10.41 **a**	9.54 **b**	8.02 **c**	7.04 **a**	0.1	<.0001	< 2.2e-16	<.0001	<.0001
Total Zeins (% DM)	5.51	5.26 **a**	4.91 **a**	3.89 **b**	3.06 **c**	0.127	0.4181	< 2.2e-16	<.0001	0.0009
α- Zeins (% DM)	2.89	2.7 **a**	2.49 **b**	2.01 **c**	1.49 **d**	0.063	0.1334	< 2.2e-16	<.0001	0.0078
α- Zeins (% CP)	23.3	25.8 **a**	25.4 **a**	24.4 **a**	20.4 **b**	0.67	0.0308	1.64E-08	<.0001	0.14
α- Zeins (% Zeins)	55.9	54.9	53.9	56.2	54.1	1.01	0.946	0.254766	0.75	0.2
$\beta\gamma\delta$- Zeins (% DM)	2.25	2.22 **a**	2.08 **a**	1.54 **b**	1.23 **d**	0.0819	0.9975	5.26E-13	<.0001	0.005
$\beta\gamma\delta$- Zeins (% CP)	18.2	21.6 **a**	21.6 **a**	19.1 **ab**	17.1 **b**	0.94	0.0078	0.00018	<.0001	0.6337
$\beta\gamma\delta$- Zeins (% Zeins)	46.8	45	46.1	43.8	45.9	1.05	0.9503	0.256709	0.7387	0.1998

[a-c] Means in the same row with different letters differs ($p<0.05$). Treatment comparison between 30, 60, 120, and 240d.

*F = fermentation effect (0 vs 30 d), E = ensiling effect (from 30 to 240), L = linear effect and Q = quadratic effect comparison between day 30 and 240.

Figures

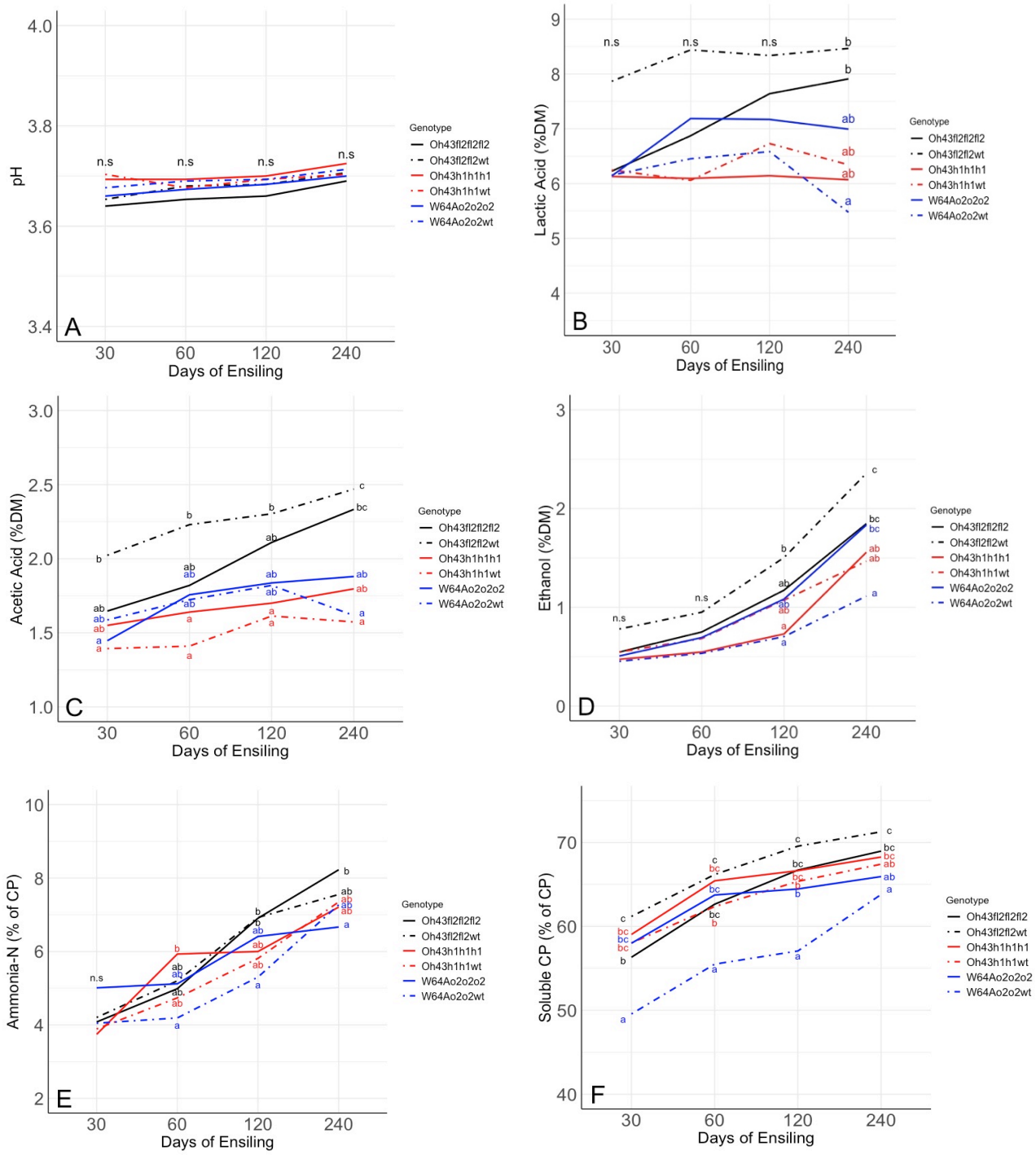


Figure 3-1 pH, lactic acid, acetic acid, ethanol, ammonia – N and Soluble CP

WPCS of the 6 isogenic lines ensiled for 30,60,120, and 240 days. Lines with the same color but different patterns represent isogenic pairs. Means within the same day with different letters (a-c) represent significant differences among genotypes within that ensiling day ($P<0.05$). DM = Dry Matter, CP = Crude Protein.

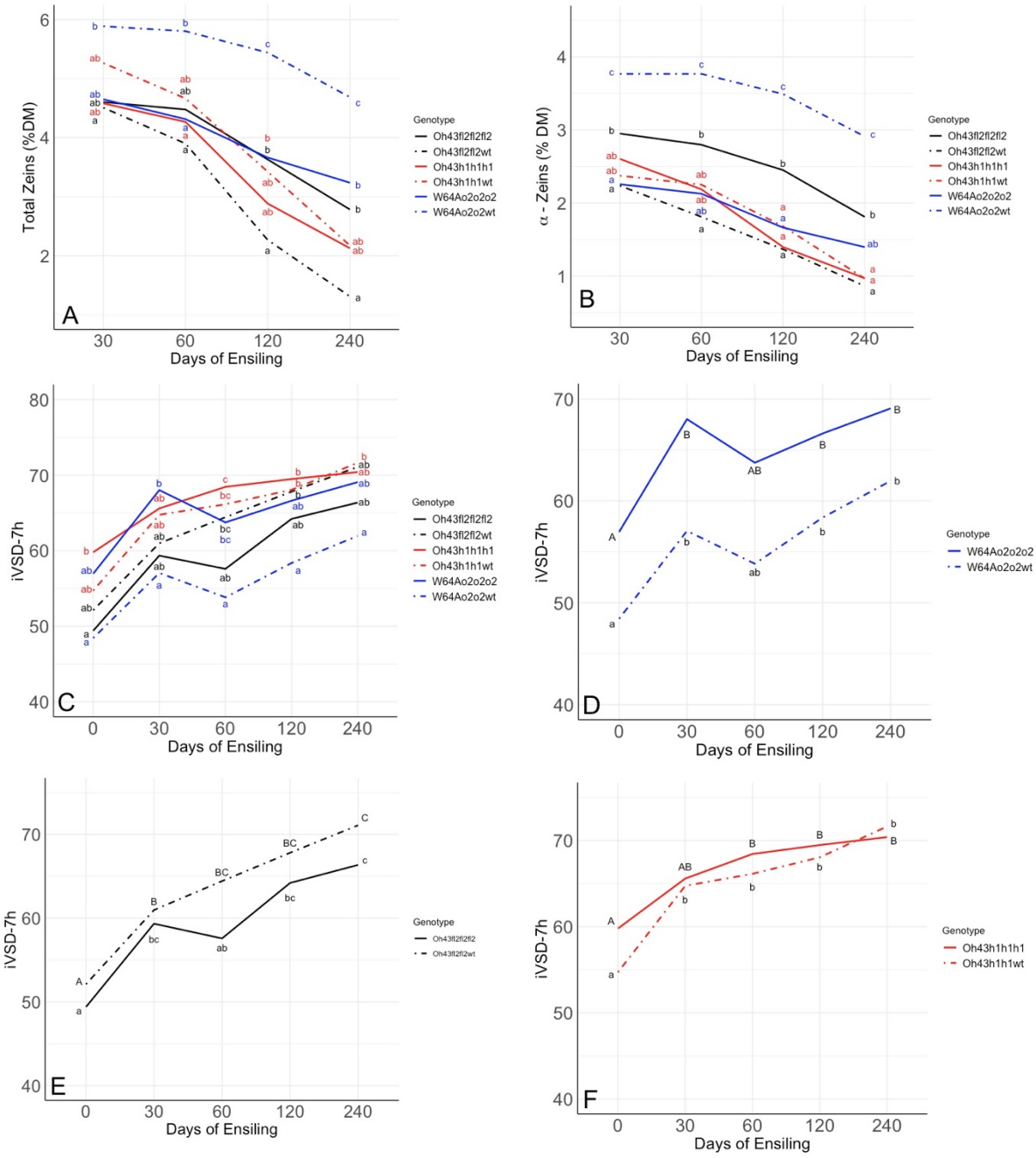


Figure 3-2 Total zeins, α- zeins and in vitro starch digestibility (iVSD-7h) of the 6 isogenic lines ensiled for 0,30,60,120, and 240 days

Panes A, B and C: Total Zeins (% DM), α -Zeins (%DM) and In vitro starch digestibility (IVSD-7h) for the 3 isogenic pairs ensiled for 0,30,60,120, and 240 days. Lines with the same color but different patterns represent isogenic pairs. Means within the same day with different letters (a-c) represent significant differences among genotypes within that ensiling day ($p<0.05$). Panes D, E and F: ivSD for the 3 isogenic pairs ensiled for 0, 30, 60, 120, and 240 days. Means within the same genotype with different letters (a-c) represent significant differences among ensiling times for the genotype ($p<0.05$).

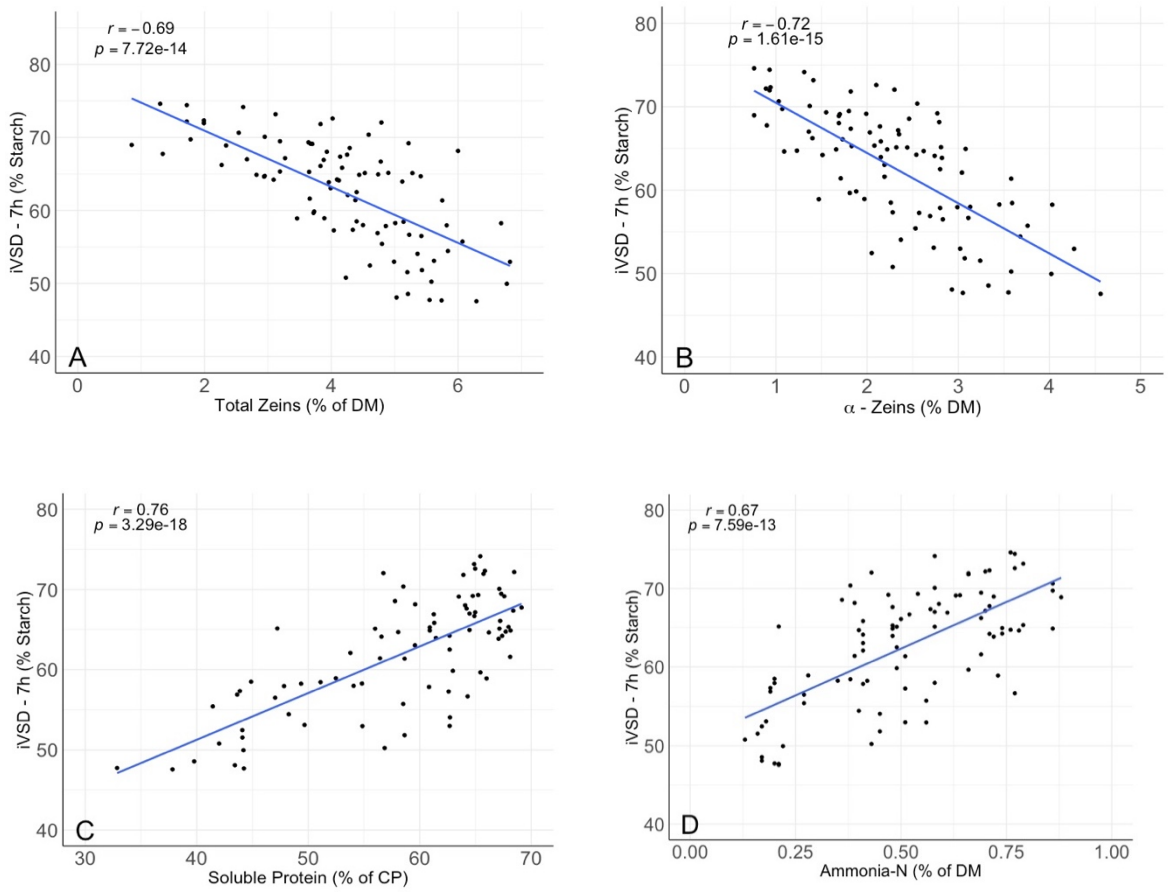


Figure 3-3 Scatterplots of correlations between total zeins (% of DM), α-zeins (% of DM), soluble protein (% of CP), and ammonia-N (% of DM) with in vitro starch digestibility (iVSD)

References

Amylase-treated neutral detergent fiber in feeds. (2005). AOAC Official Method 2002.04

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of statistical software, 67: 1-48. doi: 10.18637/jss.v067.i01.

Canale, A., Valente, M.E., Ciotti, A. (1984). Determination of volatile carboxylic acids (C_1-C_{5i}) and lactic acid in aqueous acid extracts of silage by high performance liquid chromatography. *Journal of the Science of Food and Agriculture*. 35: 1178-1182.

Coors, J.G., Lauer, J.G. (2001). Silage corn. In A.R, Hallauer, (Eds.). *Specialty Corns* (pp.347-392). CRC Press, Boca Raton, FL.

Correa, C.E.S., Shaver, R.D., Pereira, M.N., Lauer, J.G., Kohn, K. (2002). Relationship between corn vitreousness and ruminal in situ starch degradability. *J. Dairy Sci*. *85*: 3008-3012.

Der Bedrosian, M. C., Kung Jr, L., Nestor Jr, K.E. (2012). The effects of hybrid, maturity and length of storage on the composition and nutritive value of corn silage. *J. Dairy Sci*. 95: 5115–5126.

Dietary Starch in Animal Feeds and pet food. (2014). AOAC Official Method 2014.10

Ferraretto, L. F., Crump, P.M., Shaver, R.D. (2013). Effect of cereal grain type and corn grain harvesting and processing methods on intake, digestion, and milk production by dairy cows through a meta-analysis. *J. Dairy Sci*. 96: 533–550.

Ferraretto, L.F., Crump, P.M., Shaver, R.D. (2015). Effect of ensiling time and exogenous protease addition to whole-plant corn silage of various hybrids, maturities, and chop lengths on nitrogen fraction and ruminal in vitro starch digestibility. *J.Dairy Sci*. 98: 8869-8881.

Ferraretto, L. F., Shaver, R.D. (2015). Effects of whole-plant corn silage hybrid type on intake, digestion, ruminal fermentation and lactation performance by dairy cows through a meta-analysis. *J. Dairy Sci*. 98: 2662–2675.

Fiber (Acid Detergent) and Lignin in Animal Feed. (1996). AOAC Official Method 973.18

Firkins, J. L., Eastridge, M. L., St-Pierre, N. R., Noftsger, S. M. (2001). Effects of grain variability and processing on starch utilization by lactating dairy cattle. *J. Anim. Sci*. 79(E. Suppl.): E218– E238.

Habben, J.E., Kirleis, A.W., Larkins, B.A. (1993). The origin of lysine-containing proteins in *opaque-2* maize endosperm. *Plant Molecular Biology,* 23: 825-838

Holding, D.R., Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica.* 51: 243-254

Holding, D. R. (2014). Recent advances in the study of prolamin storage protein organization and function. *Frontiers in Plant Science*. 5: 276. doi: 10.3389/fpls.2014.00276

Hoffman, P. C., Esser, N.M., Shaver, R.D., Coblentz, W.K., Scott, M.P., Bodnar, A.L. (2011). Influence of ensiling time and inoculation on alteration of the starch-protein matrix in high-moisture corn. *J. Dairy Sci*. 94: 2465–2474.

Jones, R.A. (1978). Effect of *Floury-2* Locus on Zein Accumulation and RNA Metabolism During Maize Endosperm Development. *Biochemical Genetics*. 16,:27.

Junges, D., Morais, G., Spoto, M.H.F., Santos, P.S., Adesogan, A.T., Nussio, L.G., Daniel, J.L.P. (2017). Short communication: Influence of various proteolytic sources during fermentation of reconstituted corn grain silages. *J. Dairy Sci.* 100: 9048–9051. Doi: https://doi.org/10.3168/jds.2017-12943.

Jurjanz, S., Monteils, V. (2005). Ruminal degradability of corn forages depending on the processing method employed. *Anim. Res*. 3: 15–23.

Kotarski, S. F., Waniska, R.D., Thurn, K.K. (1992). Starch hydrolysis by the ruminal microflora. *J. Nutr*. 122: 178–190.

Kung, L. Jr., Shaver, R.D., Grant, R.J., Schmidt, R.J. (2018). Silage review: Interpretation of chemical, microbial, and organoleptic components of silages. *J. Dairy Sci*. 101:4020–4033. Doi: https://doi.org/10.3168/jds.2017-13909.

Larson, J., Hoffman, P.C. (2008). Technical Note: A Method to Quantify Prolamin Proteins in Corn that are Negatively Related to Starch Digestibility in Ruminants. *Journal of Dairy Science*. 91: 1-6.

Lenth, R. emmeans: Estimated marginal means, aka Least-Squares Means. R package version 1.7.4-1. https://CRAN.R-project.org/package=emmeans.

Licitra, G., Hernandez, T.M., Van Soest, P.J. (1996). Standardization of procedures for nitrogen fractionation of ruminant feeds. An. Feed. Sci. Tech, 57: 347-358.

Lopes, J. C., Shaver, R.D., Hoffman, P.C., Akins, M.S., Bertics, S.J., Gencoglu, H., Coors. J.G. (2009). Type of corn endosperm influences nutrient digestibility in lactating dairy cows. *J. Dairy Sci*. 92:4541–4548. Doi: https://doi .org/10.3168/jds.2009-2090.

McAllister, T. A., Bae, H.D., Jones, G.A., Cheng, K.J. (1994). Microbial attachment and feed digestion in the rumen. *J. Anim. Sci*. 72: 3004–3018.

Mertz, E. T., Bates, L. S., Nelson, D. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science.* 145: 279.

Misra, S., Oaks, A. (1981). Enzymes of nitrogen assimilation during seed development in normal and high lysine mutants in maize (*Zea mays*, W64A). *Can J Bot*. 59:2735-2743.

Mu-Forster, C., Wasserman, B.P. (1998). Surface localization of zein storage proteins in starch granules from maize endosperm: Proteolytic removal by thermolysin and in vitro cross-linking of granule-associated polypeptides. *Plant Physiology*. 116: 1563-1571.

National Forage Testing Association (NFTA). Method 2.1.4 – Dry matter by oven dryiong for 3 hr at 105 C.

Ngonyamo-Majee, D., Shaver, R.D., Coors, J.G., Sapienza, D., Lauer, J.G. (2009). Influence of single-gene mutations, harvest maturity and sample processing on ruminal in situ and post-ruminal in vitro dry matter and starch degradability of corn grain by ruminants. *Anim. Feed Sci. Technology*. 151: 240-250. Doi: https://doi.org/10.1016/j.anifeedsci.2009.02.002

Philippeau, C., Michalet-Doreau, B. (1998). Influence of genotype and ensiling of corn grain on in situ degradation of starch in the rumen. *J. Dairy Sci*. 81: 2178–2184.
Protein (Crude) in Animal Feed: Combustion Method. AOAC Official Method 990.03 16th Edition, Ch4 pg 18

R Core Team .(2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://R-project.org/

Richards, C. J., Peterson, J.F., Britton, R.A., Stock, R.A., Krehbiel, C.R. (1995). In vitro starch disappearance procedure modifications. *Ani. Feed Sci. Tec*. 55: 35–45.

Schmidt, R.J., Burr, F.A., Aukerman, M.J., Burr, B.(1990). Maize regulatory gene *opaque-2* encodes a protein with a "leucine zipper" motif that binds to zein DNA. *Proc Natl Acad Sc*. 87: 46-50.

Schmidt, R.J. (1993). *Opaque-2* and zein gene expression. In Verma DPS, editor. *Control of Plant Gene Expression*. pp. 337-355. Boca Raton, FL: CRC Press.

Simpson, D. J. (2001). Proteolytic degradation of cereal prolamins -the problem with proline. *Plant Sci*. 161: 825–838. https://doi.org/10.1016/S0168-9452(01)00482-4

Taylor, C. C., Allen, M.S. (2005). Corn grain endosperm type and *brown midrib3* corn silage: Ruminal fermentation and N partitioning in lactating cows. *J. Dairy Sci*. 88: 1434–1442. https://doi.org/10.3168/jds.S0022-0302(05)72811-3.

Tsai, C.Y., Larkins, B.A, Glover, D.V. (1978). Interaction of the *opaque-2* gene with starch-forming mutant genes on the synthesis of zein in maize endosperm. *Biochem. Genet*. 16: 883-896.

Yao, D., Qi, W., Li, X., Yang, Q., Yan, S., Ling, H., Wang, G., Wang, G., Song, R. (2016). Maize *opaque10* Encodes a Cereal-Specific Protein That Is Essential for the Proper Distribution of Zeins in Endosperm Protein Bodies. *PLoS Genetics*. 12: e 1006270

Young, K. M., Lim, J.M., Der Bedrosian, M.C., Kung Jr, L. (2012). Effect of exogenous protease enzymes on the fermentation and nutritive value of corn silage. *J. Dairy Sci.* 95: 6687–6694.

Acknowledgements

The authors wish to thanks Joao Dorea for valuable discussions and Martin Costa for helping processing samples.

Conflict of interest

The authors declare no conflicts of interest.

Chapter 4 Genetic Mapping of Endsosperm Vitreousness and Related Hardness Traits in a Diversity Panel and a Multiparent Population.

Author contributions

José I. Varela, Kathryn J.Michel, Shawn M. Kaeppler and Natalila de Leon.

This was a collaborative project. I was responsible for the WiDivs fields trials, data collection, seed scanning, genetic mapping, and writing the manuscript. MAGIC field trials were supervised by K.M. The deployment of the PHG and generation of the molecular markers for the MAGIC was done by K.M. The work was supervised by N.dL and S.M.K.

Abstract

Endosperm vitreousness affects the quality of maize for human food and for ruminant feed. The level of vitreousness affects starch degradability in the rumen. To investigate the genetic architecture of endosperm maize kernel traits, endosperm vitreousness, protein, density and mass were measured using Near Infrared Spectroscopy with a novel hyperspectral flatbed scanner. A multiparental advanced generation intercross population, the Wisconsin Stiff Stalk MAGIC (WI-SS-MAGIC) and the Wisconsin Diversity Panel (WiDiV) were used to perform multiparent linkage mapping and Genome Wide Association Studies (GWAS). In total 640 DH lines from the WI-SS-MAGIC were tested in two years with two field replicates in Wisconsin and 809 diverse lines from the WiDiv population were tested in two environments (Madison and Arlington, WI) with two replicates. After plants reached physiological maturity, ears were dried, shelled and kernel samples were scanned in the hyperspectral flatbed scanner. Previously developed prediction curves were used to obtain the traits.

Heritabilities for endosperm traits were generally high, ranging from 0.6 for kernel mass to 0.84 for endosperm vitreousness indicating the traits are amenable to genetic analysis. 23 significant associations were found using GWAS and 13 were found with linkage mapping. Several candidates genes were proposed as regulators of endosperm traits based on proximity to the marker associated and functional annotations. Our findings will add to the understanding of the genetic architecture of vitreousness and hardness related traits in maize that could be useful for the genetic improvement of maize varieties with enhanced nutritional composition.

Introduction

Maize (*Zea mays L.*) is a versatile crop with a wide variety of uses including feed, fuel, food, and industrial products. Relative to total production, the vast majority is used as feed for animals and to produce alcohols for fuel use. A smaller but important portion is processed for human consumption and other industrial uses. A few examples of such uses include high fructose corn (mayze) syrup, glucose, dextrose, starch, corn oil, beverage alcohol and industrial alcohol.

Compositional traits are important factors in determining the overall quality of the grain in most, if not all, maize end uses. In sweet corn, for example, the main determinant of flavor is sweetness, which is primarily determined by the amount of sucrose in the endosperm (Reyes et al., 1982) and texture or "mouthfeel", which is due to the amount of starch and the ratio of starch and water-soluble polysaccharides (WSP) (Tracy et al., 2020). Though improving yield has been the main breeding focus for field maize, genetic gain for this trait has brought with it significant changes in kernel composition. Duvick and Cassman (1999) investigated commercial hybrids released from 1960 to 1990 and found that selection for higher yielding hybrids was associated with steady decreases in grain protein content, and increases in starch content.

Whole plant corn silage (WPCS), which is the chopped, compressed, and fermented version of the plant biomass is another important use of this crop as is the primary source of feed for dairy cattle in the United States. Kernel quality plays an important role in WPCS as the kernel component represents between 40-50% of the total plant dry matter in commercial hybrids and presents the highest digestibility parameters compared to other fractions. To increase energy delivery by unit of mass, maize silage varieties are selected to have both a high harvest index (grain to stover ratio) and high percentage of starch in the endosperm.

Another important quality trait is endosperm hardness, also referred to as endosperm vitreousness due to its glassy appearance that is translucent to light compared to non-translucent and opaque appearance of soft endosperms. This trait has substantial relevance in multiple applications of maize. For example, in the case of livestock feed, the bioavailability of starch in the endosperm is highly dependent on endosperm hardness (Philippeau & Michalet-Doreau, 1997; Dias Junior et al., 2016). Furthermore, endosperm vitreousness has been shown to affect resilience during harvest and storage and it has been found to be a factor affecting resistance to insects and fungi, and other practical characteristics (Holding & Larkins, 2006). A reduction in kernel hardness can also cause a higher level of kernel cracking or stress cracks during storage and drying of the grain at harvest, which leads to increased moisture uptake in the kernel (Jackson et al., 1988) with subsequent impacts on final product quality. Maize kernels containing a greater amount of vitreous endosperm have a higher kernel density and test weight and smaller flotation indices (Caballero-Rothar et al., 2019;Gayral et al., 2016).

What creates vitreous endosperm is a longstanding question with many hypotheses, most of which are associated with the synthesis of starch and storage proteins, and the factors affecting them (Wang et al., 2021). Starch and protein are the major storage metabolites present in maize endosperm accounting for ~70% and ~10% of the total dry weight, respectively. Starch is synthesized and accumulated as starch grains (SGs) in amyloplasts. Prolamins, also known as zeins, are the most abundant storage proteins, and they can account for 60%-70% of the endosperm protein (Boston and Larkins, 2008). Although the process of vitreous endosperm formation is not clear yet, considerable evidence suggests that accumulation and packaging of prolamins into endoplasmic reticulum (ER) protein bodies play a central role (Holding, 2014). Multiple mutations affecting endosperm hardness have been identified, and they often show defects in the

accumulation of zeins or their packaging into ER-localized protein bodies. One of the most studied endosperm mutations is *opaque endosperm 2* (*o2*) which increases the lysine content in maize endosperm by decreasing the synthesis of zein proteins and increasing the level of other lysine-rich proteins. The *o2* mutant was used in breeding programs to increase lysine and tryptophan content, but agronomic problems with the resulting soft kernels initially limited the use of *o2* maize worldwide. Breeders at the International Maize and Wheat Improvement Center (CIMMYT) began the quality protein maize (QPM) breeding and made significant progress in the development of selected *o2* lines with more vitreous kernels that retain the protein profile of earlier *o2* lines (Gibbon and Larkins, 2005). Others classic mutations that affect in zeins production are *floury-2* (*fl2*), *Defective endosperm B30* (*De-B30*), and *Mucronate* (*Mc*). Mutant alleles *fl2*, *De-B30*, and *Mc* show pleiotropic effects and result in a general reduction of all zeins, and associated increases in lysine-containing proteins and lobed protein bodies similarly to *o2* (Lending and Larkins, 1992).

The characterization of opaque mutants has shown that vitreous endosperm formation depends on the correct expression and processing of zeins themselves but also on factors that may have indirect roles in zeins protein bodies such as floury-1 (*fl1*) and opaque-1 (*o1*) which do not reduce zein accumulation. *fl1* encodes a protein body membrane protein that affects the distribution of 22-kD α-zeins in the protein bodies (Yao et al., 2016) while *o1* has a reduced number of slightly smaller protein bodies (Wang et al., 2012). Some of these mutations have been deeply studied providing important insights into the biological mechanism that differentiate vitreous and non-vitreous endosperm but its inconrporation in commercial breeding programs has been minimal due to extreme phenotypes and/or undesired detrimental agronomic traits due to pleiotropy.

Current understanding of how quantitative trait loci explain genetic variation in endosperm vitreousness found in natural populations is limited and could potentially bring new understanding of genetic control of the trait with more applicability to commercial breeding.

Genome-wide association studies (GWAS) are a powerful tool to search candidate regions of the genome associated with a particular trait to determine, or at least hypothesize, the genetic control of a trait. The resolution of association mapping depends on the historical recombination that, over time, produces smaller linkage blocks that can be associated with traits (Zhu et al., 2008). Several association studies have been performed for major endosperm constituents such as starch, protein and oils (Wilson et al., 2004; Cook et al., 2012; Renk et al., 2021) but to our knowledge there is no association study performed focused on dissecting endosperm vitreousness specifically. This type of study benefits from high genetic diversity and a historical accumulation of recombination from deep coalescent history. This implies a need of many single-nucleotide polymorphisms (SNPs) required due to the generally rapid linkage disequilibrium (LD) decay. GWAS is not effective at finding causal variants generated by rare alleles, which are present at very unbalanced frequencies in the examined population. These markers are oftentimes removed, as a quality control step, which impedes the detection of association if their rare allele is truly causal.

An alternative approach is the use of structured populations, such as biparental linkage mapping populations. Such populations provide high mapping power but suffer from relatively limited diversity and recombination events due to the limited number of founders and fewer cycles of intermating used to form the population compared to GWAS diversity panels.

Both linkage mapping and GWAS have inherent strengths and weakness, as described above. To alleviate some of the limitations of the two methods, a third approach has been proposed based on multiparent structured populations as a way to bridge the two approaches and increase mapping

resolution by incorporating greater genetic diversity compared to biparental populations and increasing the number of crossing generations and elevating minor allele frequency compared to GWAS panels (Dell' Acqua et al., 2015). Examples of such populations include the Nested Association Mapping (NAM) panels (McMullen et al., 2009) and Multi-parent Advanced Generation InterCrosses populations (MAGIC) (Kover et al., 2009; Huang et al., 2012) populations. MAGIC populations are produced by crossing more than two inbred founder lines by creating panels of recombinant inbred lines that are mosaics of the founder genomes. MAGIC populations have been developed and used to perform genetic mapping for several plant species such as Arabidopsis (Kover et al., 2009), wheat (Huang et al., 2012) and maize (Dell' Acqua et al., 2015; Michel et al., 2022).

The objective of this work is to expand the current knowledge of the genetic architecture underlying key endosperm traits such as vitreousness, kernel density and total protein, using both GWAS to exploit the natural diversity of the Wisconsin Diversity (WiDiV) and the Wisconsin Stiff-Stalk MAGIC population to exploit the power of multiparent linkage mapping.

Materials and Methods

Germplasm and field trials

Multiparental Linkage mapping

The Wisconsin Stiff-Stalk Multiparent Advanced Generation Intercross (WI-SS-MAGIC) was used to perform linkage mapping. This population consist of 782 double haploids (DH) derived from the parental maize inbreds B73, B84, LH145, NKH8431, PHB47 and PHJ40. These parents represent the primary Stiff-Stalks sub-heterotic groups and spans the diversity of this pool. Further details about the population development can be found at Michel et al., (2022). Briefly, to form

the population, the six parents were crossed in a half diallel and then every possible F_1 hybrid combination cross (without reciprocals) was produced. In the event of a F_1 x F_1 failed cross, additional crosses derived from the same four founders lines were added to compensate and keep equal representation of all parents. In subsequent generations, plants from the population bulk were randomly intermated by designating each plant as pollen donor or seed donor parent and using the individual only once for crossing. Balanced bulks were made from the seed harvested from the intermated plants. After two generations of intermating, a subset of the population (hereafter called "Subset A") was sent for doubled haploid (DH) induction, provided as in-kind support by AgReliant Genetics. The remaining balanced bulk was randomly intermated for two additional generations and subsequently sent for DH induction (hereafter "Subset B")

The DH lines were grown in replicated trials at the West Madison Agricultural Research Station in Verona, WI in 2016 and 2017 as single row plots. In 2016 the rows were planted at 5.5 m long separated by 0.76 m with a targeted final population of 60,000 plants ha^{-1} and in 2017 were planted at 3.7m separated by 0.76m with a target final population of 91,000 plants ha^{-1}. Subset A and Subset B groups were organized as subblocks within a randomized complete block design (RCBD) with two replications. After physiological maturity was reached, primary ears were collected from three evenly spaced plants per plot and dried under forced air until reaching approximately 10-15% moisture.

Genome-Wide Association Study (GWAS)

To perform this study, the Wisconsin Diversity Panel (WiDiv) was used. In its full conformation this panel contains 942 maize inbred lines that produce physiologically mature kernels in the Midwest region of the United States (Mazaheri et al., 2019). 899 out of the 942 Inbreds were grown at the UW-Arlington agricultural research station and at the UW-West Madison Agricultural

research station in 2018 and 2020 respectively. Both 2018 and 2020 experiments were planted as single row plots of 9' long separated by 30 inches with a targeted final population of 70,000 plants ha^{-1}. The trial in Arlington was planted on May 25 and ears were harvested on October 15. The trial in West Madison was planted on May 20 and ears were harvested on October 20-24. Fields were arranged in a Randomized Complete Block Design (RCBD) with two blocks per location. After physiological maturity was reached, primary ears were collected from three evenly spaced plants per plot and dried under forced air until reaching approximately 10-15% moisture.

Phenotypic Data Analysis

After drying, ears were shelled and the kernels of each of the three ears per plot were stored in individual envelopes. The kernels from both WiDivs-942 and the WI-SS-MAGIC population were phenotyped using a custom made near infrared (NIR) hyperspectral flatbed scanner. This scanner contains a 12-bit NIR line-scan camera (Specim model FX17e, Oulu, Finland) that collects the reflected light from the samples scattered over a glass bed. The scanner produces hyperspectral images where each pixel in the line registers the energy in 224 wavelength bands between 950 nm and 1700 nm, corresponding to a spectral resolution of 3.3 nm. Prediction curves based on Partial Least Square Regression (PLSr) were constructed to predict endosperm vitreousness (%), total kernel protein (%), total kernel density (g/cm^3) and kernel mass (g) as explained in Chapter 2. Briefly, approximately 80 kernels per ear were placed over the scanner surface using a custom-made 3D printed kernel separator that allowed positioning the kernels without touching each other's in a high throughput fashion. Kernels were illuminated from below the scanning surface with halogen lamps and reflected NIR radiation was collected and stored. Each hyperspectral image collected was processed in Matlab to convert the NIR reflectance raw values into absorbances and the average NIR absorbance per kernel was extracted. The absorbances were

exported to R and submitted to the PLSr prediction curves previously constructed for the traits mentioned.

The traits predicted were endosperm vitreousness, kernel protein, kernel density and kernel mass. Vitreousness percentage is the mass ratio of the vitreous endosperm fraction to total endosperm mass and protein is the total kernel protein percentage.

Genetic Data

Exome Capture Sequencing

Exome capture sequencing was performed on 701 DH lines from the WI-SS-MAGIC using a custom capture design acquired from Roche Diagnostic Corporation (Indianapolis, IN). Probes were designed to target the 5' and 3' ends of the untranslated regions (UTR) of the maize B73_RefGen_v2 genic regions and presence-absence variation (PAV) regions derived from the alignment of whole genome sequencing reads of a core set of 32 inbreds to B73 reference version 2 (Brohammer et al., 2018; Mazaheri et al., 2019). In total, 82,351 genic regions (approximately 26.5 Mb) and 492 PAV regions (approximately 2.8 Mb) of the maize genome were targeted using tiled, variable length probes, with an average probe size of 75 nt. DNA extraction, library preparation, sequencing, and raw sequences quality control were performed as described in Michel et al., 2022.

Genotyping by Sequencing

Additional genotyping was performed on 144 WI-SS-MAGIC DH lines using Genotyping-by-Sequencing (GBS) at the University of Wisconsin Biotechnology Center. Briefly, dual digest GBS was performed with restriction enzymes PstI and MspI on DNA extracted from frozen seedling leaf tissue (Elshire et al., 2011; Poland et al., 2012). DNA was sequenced using an Illumina NovaSeq6000 sequencing system in paired end mode 150 nt and analyzed using bcl2fastq

v2.20.0.422 (San Diego, CA, USA). Read one was demultiplexed and barcodes were removed using Tassel-5-Standalone (Bradbury et al., 2007). Read two was not included in future analysis.

Practical Haplotype Graph (PHG)

The five founder lines of the WI-SS-MAGIC population underwent whole genome assembly (Bornowski et al., 2021) and the six one (reference genome B73) was already available which allowed to infer high density markers to the recombinants DHs using the practical haplotype graph tool as described in Michel et al., 2022.

Briefly, the B73 version 5 genome was used as reference to build the PHG (Bradbury et al., 2021). B73 genes annotations are available (www.maizegdb.org) and were used to make the initial reference ranges endpoints. The other five parental de novo genome assemblies were incorporated to the database and aligned to the reference. After that, the B73 genome was added to allow haplotypes of this founder to be included. Once the consensus haplotypes were generated and indexed in the pangenome, the exome capture and GBS reads could be incorporated to the graph. Using Hidden Markov Model, the paths were identified through the graph for each taxon and SNPs were called in the genic reference region for the progeny population. Exome capture reads were aligned as paired end sequences, while the GBS reads were aligned as single end sequences. Due to the expected homozygosity of the DH lines and parental assemblies, only homozygous SNPs were generated from PHG.

Tassel-5 (Bradbury et al., 2007) was used to select and filter the markers. SNPs with any missing parental data were removed, minor SNP states were set to missing to remove third, fourth and other alleles, and the SNP was removed if the minor allele frequency was less than 0.05. To reduce correlation between SNPs and decrease mapping computational time, 100,000 evenly spaced SNPs were selected across the ten chromosomes and converted to numerical major or minor allele.

RNA-Seq Data

The 809 lines used for the GWAS study have an available set of 899,784 SNP derived from RNA-sequencing identified in Mazaheri et al., (2019), which represents an expansion of the previously published version of the WiDivs (Hansey et al., 2011). The WiDiv-942 panel includes a diverse set of public, expired plant variety protection (exPVP), and germplasm enhancement of maize (GEM)-derived inbreds. This panel represents some of the major North American field corn heterotic groups, including stiff stalks (SS), non-stiff stalks (NSS), and Iodent (IDT), as well as sweet corn, popcorn, and tropical inbreds (Mazaheri et al., 2019).The 899,784 SNPs hapmap file was filtered for lines that contained phenotypic data and for minor allele frequencies greater than 0.05 and converted to numerical format using TASSEL v5 (Bradbury et al., 2007) resulting in 430,695 SNPs.

Statistical analysis

Statistical analysis was conducted using R version 3.6.3 (R Core Team, 2020). Raw values per kernel per image were extracted from the hyperspectral predictions and subjected to outlier detection based on the distribution of kernels values within an image. Endosperm vitreousness, total kernel protein, kernel density and kernel mass values outside 1.5 times the interquartile range (IQR) above the upper quartile (Q3) and below the lower quartile(Q1) were removed (Q1 – 1.5*IQR or Q3+ 1.5 * IQR). This method helped protect the average value per image of possible incorrect predictions triggered by NIR absorbances of kernels that landed in unexpected orientation (cap down) or containing any possible type of sample contamination that could alter the light reflectance. This formula was also used to remove values with circularity or perimeter outside the previously described boundaries, as a method to detect “touching kernels”.

After removing outliers, the average value off all kernels per image was computed as the phenotype per experimental unit.

Both the WiDiV experiments and WI-SS-MAGIC experiments were analyzed as a RCBD with the following model to extract Best Linear Unbiased Estimators (BLUES) for each phenotype:

$\gamma_{ijk} = \mu + G_i + E_k + (G*E)_{ik} + B_{j(k)} + \varepsilon_{ijk}$ (1)

Where:

γ_{ijk} = Response of the ith genotype in the jth rep (Block) within the kth environment (location-year combination).

μ = is the population grand mean.

G_i = The ith genotype effect.

E_k = The kth environment effect.

$(G*E)_{ik}$ = The ith Genotype by kth Environment effect interaction.

$B_{j(k)}$ = The effect of the jth Block nested in the kth Environment.

ε_{ijk} = The error. Assumed to be independent and normally distributes.

All effects in Eq. (1) except μ were considered as random to estimate variance components and were computed by restricted maximum likelihood (REML) using ASReml-R v4 (Butler et al., 2018). Adjusted means for each line were obtained as best linear unbiased estimates (BLUES) considering μ, and Gi in Eq. (1) as fixed effects and the remaining effects as random using ASReml0R v4. Pearson correlation coefficient (r) were calculated among all traits based on the adjusted means of the lines.

Heritability was calculated as follows (Cullis et al., 2006)

$h^2 = 1 - \frac{\bar{v}_{\Delta}^{BLUP}}{2\sigma_g^2}$ (2)

Where σ^2_g is the genotype variance and $\bar{v}_{\Delta}^{BLUP}$ is the average standard error of the genotypic BLUPS.

Phenotypic comparison of WiDivs subgroups:

To compare kernel traits between WiDivs subgroups all the SS and NSS originally assigned by Mazaheri et al., (2019) were combined into one single group, respectively. The "broad origin" and "mixed" group were excluded from this comparison due to lack of information on its usage and unfamiliarity with the members of this group to allow inferences. Tukey's Honest Significant Difference HSD ($\alpha = 0.05$) was applied over the phenotypic BLUES.

Multiparental Linkage mapping

The R package qtl2 (Broman et al., 2019) was used to perform single-marker QTL mapping using the marker dataset derived from the WI-SS-MAGIC PHG. The physical map coordinates of each SNPs were provided, and the genetic map was approximated by converting each SNP' megabase pair position to centiMorgans using the B73 RefGen_v5 chromosomal genome length of 2132 Mbp divided by the composite US-NAM genetic map length of 1456.68 cM (Li et al., 2015). To identify potential sample duplicates, the function *compare_geno()* was used to calculate markers matching for all pairwise comparisons, and any pair of individuals with greater than 95% marker sharing was removed. The conditional genotype probabilities, or the true genotypes giving the observed multipoint marker data, were calculated using the function *calc_genoprob()* which uses Hidden Markov Models (HMM) with an error probability of 0.01. The maximum marginal probability of the parental haplotypes was calculated and the total number of crossover events per individual was identified using the function *count_xo()*. Lines of the subset A with four generations of recombination, with more than 150 theoretical crossovers were removed as well as the lines of subset B (6 generations of recombination) with more than 250 crossovers from further analysis.

This is probably an artifact of germplasm contamination that incapacitates the HMM to deduce the correct underlying parental haplotypes in non-parental regions, and instead frequently switches back and forth among the parental haplotypes.

After quality control, 640 individuals with phenotypic and genotypic information were used for the analysis. The WI-SS-MAGIC population may have complex genetic relationships which will introduce correlation in the model residual that may inflate type I error. This was accounted in the mapping model by including an adjustment for the kinship between the lines. The genotype probabilities were used to calculate a kinship matrix using the "leave one chromosome out" (LOCO) method. In this method each chromosome scan is conducted considering kinship calculated on all the chromosomes but the current one, limiting the correction applied locally to QTL scan. Genome scan was performed using a linear mixed model with the kinship matrix as a random polygenic effect to find association between genotype and phenotype. Significance thresholds to call a QTL were calculated by 1,000 permutation of each trait. The 95th percentile of the permuted logarithm of the odds (LOD) distribution was chosen as a significant threshold using the function scan1perm(). Bayesian credible intervals for QTL peaks were calculated using the function find_peaks(), with LOD thresholds specific to each phenotype and probability of 0.95. Chromosome-wide QTL best linear unbiased predictor (BLUP) effects were calculated using the function scan1blup(), and single locus BLUP effects were estimated using fit1() with "blup=T".

Genome Wide Association

Genome wide association was performed using the R package rMVP 1.0.4 (Yin et al., 2021). Associations were identified implementing the FarmCPU algorithm included in the rMVP package. This algorithm makes an iterative usage of fixed and random effects models to include

associated markers as covariates and to optimize marker covariates in separate steps. For each iteration the kinship matrix calculated internally by rMVP and was included as a random effect. To account for spurious associations, for each phenotype model, the first 3, 4 or 5 principal components were fitted, and Quantile-Quantile plots were visually evaluated to identify the model with the best fit. The maximum iterations allowed for FarmCpu was set to 10 (*mxLoop*=10) and the binning method selected was 'FasT-LMM'. To determine final SNP significance, the Bonferroni corrected threshold (α/(number of markers)) was used to declare significant associations.

Principal component analysis

To visually asses the WiDiv population structure and subpopulations clustering, principal component analysis was performing using the full marker set exported from rMVP. The PCAtools (Blighe, 2019) package in R was used to perform the analysis. To differentiate the subgroups each genotype belongs to, the original classification assigned by Mazaheri et al., (2019) that included heterotic pool subgroups was reduced to main heterotic pools by combining the subgroups. All the Stiff-Stalks were group as one single group (SS) as well as the Non-Stiff-stalks (NSS).

Candidate genes associations examination

Candidate genes for statistically significant SNP hits for the WiDiv GWAS corresponded to physical positions based on anotations of the AGPv4 reference assembly of inbred B73. A window of +/- 500kB from the significant SNP was arbitrary selected to look for downstream and upstream functional annotations. Suggestion of possible candidates was based on functional relationship of genes found in the search window with the trait tested and by checking endosperm expression level using a maize gene expression atlas (Stelpflug et al., 2016).

Candidate genes for statistically significan SNPs for WI-SS-MAGICS linkage mapping was based on annotations of the AGPv5 reference assembly of B73 following the criteria described above.

Results and Discussion

Phenotypic evaluation

Substantial variation was observed for the four kernel compositional traits evaluated (endosperm vitreousness, kernel protein, kernel density and kernel mass) in both populations (Table 4-1). The WIDIV population showed wider ranges than the WI-SS-MAGIC population for endosperm vitreousness, kernel density and kernel mass whereas, interestingly, the WI-SS-MAGIC showed greater range for kernel protein. The range observed for protein was consistent with a previous study that analyzed maize composition on a subset of 501 inbred lines of the WiDivs (Renk et al., 2020). This study used near infrared (NIR) to predict kernel constituents.

Transgressive segregation, or the presence of extreme phenotypes for individuals in the segregating population beyond those observed for the parental founder lines, was found for all traits in the WI-SS-MAGIC. Heritabilities, calculated according to the Cullis method, ranged from 0.60 for kernel mass to 0.84 for endosperm vitreousness in the WI-SS-MAGIC (Table 4-1) and from 0.62 for kernel mass to 0.77 of kernel protein for the WiDivs. Gustafson and de León 2010, reported heritabilities for vitreousness of 0.84 for a set of 175 recombinant inbred lines (RILS) from the intermated B73 and Mo17 (IBM) and a heritability of 0.65 for a set of 123 IBM RILs testcrossed to inbred tester W604S. Cook et al., (2012) reported kernel protein heritability of 0.84 for the Maize Nested Association Mapping population (McMullen et al., 2009) and 0.87 for the Illinois association panel (Flint-Garcia et al., 2005). These values are slightly higher but within the ranges of the values obtained in this study.

The proportion of the variance explained by the genotypes in the WI-SS-MAGIC population ranged from 32% for kernel mass to 63% for kernel vitreousness (Figure 4-1). The proportion of the variance explained by environment was minimal for the four traits in the WI-SS-MAGIC with density being the highest (3.4%) among the four kernel traits. Both environments of this experiments were planted in the same research station during consecutive years, which might explain the almost negligible contribution of environment to the total variance. WiDivs were planted on different locations (Arlington and West Madison) in two different seasons which might explain the higher proportion of the total variance that was associated to the environmental effect compared to the WI-SS-MAGIC. Endosperm vitreousness had the highest proportion of environmental variance compared to the other traits for the WIDIV experiment (Figure 4-1).

To explore the phenotypic correlation among traits, Pearson's correlation coefficients were computed among the four principal kernel compositional traits plus additional phenotypes collected in both seasons of the WI-SS-MAGIC (Figure 4-2). These correlations were calculated for the WI-SS-MAGIC because for that experiment flowering time and plant heigh were also collected. Endosperm vitreousness strongly correlated with kernel density ($r = 0.75$), as expected. Previous studies have shown, through the use of electron micrographs, that vitreous endosperm tend to have more compressed starch grains and protein bodies embedded in the desiccated cytoplasm, while more air spaces is observed between starch grains in floury endosperm (Larkins et al., 2017). Kernel mass was also correlated with endosperm vitreousness ($r = 0.48$), but in a lower magnitude. This lower correlation is probably due to the effect exerted by volume (size) variation over mass. Two kernels with the same vitreousness percentage can have significantly different mass due to dissimilar size. Kernel area and kernel major axis were negatively correlated with endosperm vitreousness. Although the focus of this study does not include the evaluation of

associations with flowering time or plant height, it is relevant to ascertain the potential correlation of different phenological characteristics with the traits of interest. Mural et al., (2022) reported that flowering time datasets in maize tended to identify a disproportionately high number of independent GWAS peaks potentially as a result of a greater proportion of variance among these traits be explained by genetic factors. They also found that genomic intervals associated with flowering time were disproportionately more likely to be associated with a different phenotypic category such as disease resistance, standability and seed composition, among others. In this study, the authors illustrate the example of *ZmMADS69* (syn *Zmm22*)(*Zm00001d042315*), a transcriptional factor located on maize chromosome 3, which has been shown to function as a flowering activator, with a derived allele conferring earlier flowering in maize lines relative to teosinte (Mazaheri et al., 2019; Liang et al., 2019). Mural et al., (2022) found that the same genomic association with the *ZmMADS69* gene appeared significant for multiple other traits such as plant height, extant leaf number, stalk diameter and biomass yield. *ZmMADS69* has been shown to have pleiotropic effects for multiple traits of agronomic importance. In our case we found that only kernel protein and kernel mass were minimally correlated with anthesis GDU ($r=0.1$, $p=0.01$ and $r=0.16$, $p = 2.96x10^{-05}$, respectively).

The 809 inbred lines of the WiDivs used for the association study consist of different maize subpopulations as described by Mazaheri et al., 2019. The initial classification presented by the authors included four subclassifications for the Stiff Stalks (SS) heterotic pool (SS-B73, SS-B37, SS-B14 and SS-BSSSC0), two subclassifications of the Non-Stiff-Stalks (NSS) (NSS-Oh43 and NSS-Mo17), Iodents (IDT), tropical, popcorn, a group of public lines of broad origin and a group of mixed lines. Given the range of types of maize that comprised the WiDiv panel included in this study, we investigated the difference in composition between the most relevant types and their

heterotic pool group, which include SS, NSS, IDT, tropical and popcorn. The popcorn and tropical groups had significantly higher average endosperm vitreousness compared to the other three groups (Figure 4-3). Popcorn and other flint maize types are known for having a large proportion of vitreous endosperm (Larkins et al., 2017). The high proportion of vitreous endosperm in popcorn along with its thick pericarp make this type of maize appropriate to produce edible popped kernels. Flake production is related to a higher ratio of hard to soft starch and a thicker pericarp that can withstand pressure building from steam, which is a trait absent from most other dent maize types (Flint-Garcia, 2017). Similarly, tropical maize lines also contain a high proportion of vitreous endosperm (Miranda et al., 2013).

Popcorn and tropical lines had the highest level of proteins (Figure 4-3). The ranking and magnitude of the differences among means are very similar to those found by Renk et al., (2020). Kernel density and endosperm vitreousness had similar profiles as expected given their high correlation (Figure 4-3). Despite having the highest endosperm vitreousness and density, the popcorn group had the lowest mass per kernel compared to all other groups, except for the IDT group (Figure 4-3). This can be explained by the smaller kernel size typically observed for popcorn types.

Population Structure

Principal component analysis from the molecular marker data of the WiDiv population reveals a clear clustering of the major heterotic pools and related subgroups. The first three principal components explained approximately 11.5% of total variance among genotypes (Supplemental Figure 4-1). As expected, multidimensional scaling (MDS) confirmed the lack of population

structure within the WI-SS-MAGIC population (Supplemental Figure 4 -2). The parental inbreds fell on the perimeter of the point cloud with no discernable clustering of progeny individuals.

Genetic Mapping

Genome Wide Association Study

Twenty-three significant associations were found for the four traits (Supplemental Table 4-1). Two significant associations were observed for endosperm vitreousness (Figure 4-4). The association identified in chromosome 1 is located within the *ao4* (*aldehyde oxisase4*) gene, which is contiguous to *ao3* (*aldehyde oxisase3*) and *ao1* (*aldehyde oxisase1*). The members of the *AO* enzyme family play a role in the last step of the carotenoid catabolism pathway by oxidating abscisic aldehyde into abscisic acid (ABA) (Seo et al., 2000). *AO* enzymes exist in multiple forms and are expressed in leaves, roots and seeds, depending on the plant species (Seo et al., 1998; Seo et al., 2000; Xiong et al., 2001). The relationship between endosperm vitreousness and carotenoids concentration in maize kernels has been suggested due to the particular orange color in some of the flint and tropical lines (mostly hard endosperm) compared to dent maize. Recent work has demonstrated this relationship experimentally. Saenz et al., (2020) showed that kernel hardness is associated with an improved ability to store more carotenoids, which may be related to carotenoids hydrophobic interactions with protein bodies, a central component of vitreous endosperm where carotenoids are mainly stored (Larkins et al., 2017). Recently, Wang et al., (2021) identified *Ven1* (*Vitreous endosperm 1*) as a major QTL that modulates kernel texture in maize by influencing amyloplast envelope integrity. *Ven1* encodes *β-carotene hydroxylase3*, an enzyme that modulates carotenoid composition in the amyloplast envelope. The nonfunctional *Ven1* allele has been shown to lead to a decrease in polar and an increase in non-polar carotenoids, along with an increase in lipid composition which give rise to a persistent amyloplast envelope. These changes impede the

gathering of protein bodies and prevent them from interacting with starch grains, creating air spaces that cause an opaque endosperm phenotype (Wang et al., 2021). With increase evidence supporting the functional relationship between endosperm assembly and carotenoids, and due to the role that the *ao4* enzyme play, this loci could be a potential candidate to explain variation in kernel vitreousness.

The second significant association for endosperm vitreousness is located at 240kB from *defective kernel 44 (dek44)* in maize chromosome 4. Defective kernel (dek) mutants are a major class of maize kernel mutants with more than 20 members. *DEK44* encodes a mitochondrial ribosomal protein that is expressed in a broad range of maize tissues, and it has been shown to be highly expressed in ear and kernels. This protein, however, only accumulates in kernels and is not detected in other tissues. Biogenesis and morphology of mitochondria is strongly affected in *dek44* kernels leading to abnormal endosperm development. Qi et al., (2019) found that overall protein content and the non-zein protein content per weight of mature endosperm were higher in mutant seed compared to wild type. The authors also found that the total starch content was 24% lower per weight in the mutant types. Although the mutant produces small kernels with an embryo lethal phenotype, allelic versions of the gene with a weaker phenotype could not be discarded as a candidate for kernel vitreous types.

Figure 4-5 shows the Manhattan plot for kernel density. The association with the lowest p-value in chromosome 1 is located at 185 kB of the gene *dof36* (*C2C2-DoF-transcriptor factor 36*), *Zm00001d029512*. The *DOF* proteins are a family of plant-specific transcriptional factors that contain a conserved DNA binding Dof domain that recognizes a cis-regulatory element with the common core sequence AAAG (Yanagisazawa, 2002). During the process of maize seed development, *DOF1* has been reported to up-regulate the expression of gamma zeins (Marzabal et

al., 2008). Additionally, the prolamin-box binding factor (PBF) Dof protein associated with seed development has been reported in barley, wheat, and finger millet (Mena et al., 2002; Dong et al., 2007). *Dof36* is specifically expressed in the endosperm between 12 and 20 days after pollination and has been demonstrated that it upregulates the expression of a high number of genes involved in the starch pathway such as *ZmAGPS1a*, *ZmAGPS1b* (glucose pyrophosphorylase), *ZmGBSS1*, *ZmGBSSIIa* (granule-bound starch synthase), *ZmSSIIa* , *ZmSSIV* (soluble starch synthase), and *ZMBE1* (Branching enzyme) (Wu et al ., 2019). Overexpression lines were used to investigate the effect of *ZmDOF36* on starch structure, and the authors found that wild type maize endosperm was filled with smooth, spherical starch granules, while the endosperm of the *ZmDOF36* overexpression transgenic lines had a looser, irregular polyhedron structure packed with larger starch granules (Wu et al., 2019). They also showed that grain length and 1000-grain weights were increased in the transgenic lines. Given that endosperm vitreousness and density are both dependent on the spatial configuration and interactions of starch and storage proteins within the endosperm, and as it has been previously demonstrated that *dof36* modulates starch arrangements, this could potentially be another relevant target gene for density in maize.

The second association for kernel density found in chromosome 6 is located at 290 kB from the *pco087998* (*Zm00001d038304*) gene, which encodes a starch binding domain containing protein and is highly expressed on maize endosperm 27 days after pollination.

The significant association in chromosome 2 for kernel mass is located at 420kB to the *nactf32* gene (*Zm00001d003414*)(Figure 4-6). This transcriptional factor belongs to the *NAC* plant specific transcriptional factor family. Functional studies have demonstrated that *NAC* transcriptional factors play a crucial role in plant growth as well as stress response to drought, salinity, and cold (Wang et al., 2020). Zheng et al., (2019) performed a study to identify the targets of the *miR164*

(MicroRNA164) dependent regulatory pathway during maize seed development and found that *NAC32* was one of its targets. They performed screening of the genes related to *NAC32* by co-expression and found response to cell wall loosening through expansin genes (*EXPB14* and *EXPB15*) followed by transient expression to confirm the relationship. Expansins genes contribute to cell enlargmenet by promoting cell wall loosening during various develomental processes. Zhang et al., (2014) found that eight ZmEXP were expressed in maize endosperm, with five exibiting endosperm-specific expression. Zheng et al., (2019) found that *NAC32* could significantly promote *EXPB14* and *EXPB15* activity. Because expansin genes are related to seed expansion (Bae et al., 2014), Zheng et al., (2019) also assessed if MIR164-transgenic lines altered seed dimensions through downregulation of *EXPB14* and *EXPA15*. They observed significant decreases in kernel length and width, and hundred-grain weight compared to wild type. No significant difference in embryo size among wild type and *MIR164*-transgenic lines was found. This study demonstrated that *miR164* participates in maize seed development via controlling patterns of *NAC32* expression and proposed a possible miR164-NAC-expansin-dependent regulatory pathway involved in maize seed expansion. Considering the role that *NAC32* plays as regulator of seed size and weight, this could be presented as a possible candidate that modulates kernel mass. Fourteen significant association were found for total kernel protein (Figure 4-7).

QTL mapping using a multi-parent population

Several multi-parent mapping populations have been developed in maize given the multiple benefits that they provide over traditional bi-parental mapping populations. Examples of such multi-parent populations include two eight parent MAGIC (Dell'Acqua et al., 2015; Jiménez-

Galindo et al., 2019), four parent populations (Ding et al., 2015; Mahan et al, 2018) and a nested association mapping population (Yu et al., 2008).

The WI-SS-MAGIC population used in this study concentrates the founders within the SS heterotic group. An advantage of focusing on the SS group is that maize breeding relies on recycling genetic materials within heterotic groups to make new parents and subsequently crossing parents between groups to make hybrids (Michel et al., 2022). Thus, the blend of the genomes of parents within a single heterotic pool versus across the diversity of maize creates a more applicable population to study the subset of alleles present within SS seed parent germplasm (Michel et., 2022). Recombining six parents increases the power to examine the effect of rare alleles on phenotypes compared to diversity panels. Mixing multiple founders takes advantage of historical recombination beyond recombinations introduced through structured population development. Multiple founders within a single population also allows the study of allelic series at loci of interest.

Thirteen significant associations were identified across the ten maize chromosomes for the WI-SS-MAGIC population (Supplemental Table 4-2). Four peaks exceeded the LOD threshold for kernel vitreousness. The association on chromosome 2 contains *Opaque-11* (*O11*) within a one-unit LOD interval at 1.9Mbp of the peak. *O11* encodes an endosperm specific basic-helix-loop-helix (bHLH) transcription factor that directly regulates genes involved in carbohydrate metabolism and stress responses (Feng et al., 2018). Some of the *O11* regulation targets are transcriptional factors including, *NKD2*, *ZmDOF3*, *PBF*, and *O2*, which are known key transcriptional factors that play important roles in both endosperm development and nutrient accumulation. *NKD2* regulates a series of biological processes in endosperm development, and *ZmDOF3* is a key regulator of starch accumulation and endosperm development (Yi et al., 2015;

Gontarek et al., 2016; Qi et al., 2017). *O2* and *PBF* play pivotal roles in zein and starch accumulation. *o11* mutant kernels display a opaque and smaller endosperm with smaller starch granules and protein bodies. It is hypothesized that the alterations in starch and protein bodies in *o11* endosperm are responsible for the opaque endosperm in *o11* kernels (Feng et al., 2018).

The significant association in chromosome 9 spans a large region of the chromosome, although the qtl2 software provides options to extract multiple peaks from the same chromosome, those functions should be used with caution as they depend on arbitrary parameters selected by the user (function *findpeaks()* argument peakdrop). The LOD support interval of the region in chromosome 9 is large with no evident peak that could guide a fine search. However, it is important to mention that *dzs10* (*delta zein structural10*) is contained in the support interval. The *dzs10* gene is expressed in the endosperm starting 12 days after pollination and encodes 10 kDa δ-zeins. Compared to other classes of zeins, 10 kD δ-zeins abundance in maize endosperm is low but they interact strongly with 22kD α-zeins to form the core of the protein bodies (Holding and Larkins, 2006). This low abundance and ability to interact with both 22kD and 19kD α-zeins may suggest that a primary function of the 10kD δ-zeins is to mediate the interaction and accumulation of the α-zeins. δ-zeins are rich in methionine and vary in abundance, suggesting that they also function as a sulfur sink (Swarup et al., 1995). The proper formation of protein bodies requires appropriate interaction between the different types of prolamins. Allelic variation in the 10kD δ-zeins might disturb this interaction triggering modifications in vitreousness levels and density.

We investigated the parental effects of the peak on chromosome 9 (Figure 4-9 and 4-10) and found that the PHJ40 allele has a positive effect along with LH145 in endosperm vitreousness an kernel density. PHJ40 is the earliest line with the highest vitreousness among the six parents. Although the background of progenitors of PHJ40 is not known publicly, admixture analysis places this line

primarily as B37 type SS but it represents a relatively unique subgroup within that class because of its diversity. This line has a characteristic hard endosperm commonly found in flint types,

The kernel density LOD profile is similar to the endosperm vitreousness profile (Figure 4-8), as expected due to the high correlation between these two traits. Although the span of the significant association in chromosome 9 is almost identical to the vitreousness one, the maximum positions are different. The peak declared for density is located at 49.1 Mbp which is at 0.64 Mbp from *dzs10*.

A large association was found in chromosome 8 for kernel protein (Figure 4-8). This broad peak suggests the presence of three peaks within that region. The highest peak obtained at 97.4 Mbp is located 84 Kb from the proline responding1 (*pro1*) gene. This gene encodes a Δ1 – pyrroline – 5-carboxylate synthetase that catalyzes the biosynthesis of proline from glutamic acid. Wang et al., 2014 measured protein levels of mature kernels of homozygous for a pro1 mutant allel and found that they contained 75.4% of the wild type protein level. They also examined the zein content in the endosperm of pro1-ref and wild-type kernels and found that the levels of zeins in pro1-ref were only 71.4% of the wild type. Using transmission electron microscopy, they observed a significant reduction in protein body number and size compared to wild type, which is consistent with previous evaluations of seed component analysis (Wang et al., 2014).

Three significant associations were found for kernel mass. The peak in chromosome 9 aligns with one of the peaks of kernel vitreousness and density, which is possible as these three traits are correlated, although the correlation of kernel mass with the vitreousness and density is not as strong given that kernel mass is dependent on volume. To investigate if the results found for kernel mass of our work are consistent with previous studies on maize kernels, we compared our results with the one reported by Zhou et al., 2020. In the study, the authors compiled a QTL atlas for grain

yield and its components traits, including kernel mass and they identified 135 QTL by conducting a meta-analysis of 1,177 QTL generated across 56 studies. The peaks found in chromosome 5 using the WI-SS-MAGICS was within the ranges of a QTL (MTQTL78) identified for kernel mass by Zhou et al., 2020. We hypothesize that any loci that do overlap between populations could be useful further targets for future study.

Conclusions

Kernel vitreousness is an endosperm characteristic associated with hardness that is relevant for multiple maize end uses. This trait has been gaining attention in the last decades in the forage community as it is associated with starch digestibility which is a fundamental requisite for high quality feed.

Genetic mapping was performed for endosperm vitreousness and three related kernel composition traits using a diversity panel and a multiparent population. Both populations showed substantial variation for all traits. Statistically significant differences in means among types of maize in the WiDiv and meaningful correlations between traits in the WI-SS-MAGICS were found to reflect the diversity of maize and the underlying relationships among traits. We presented the first association study in endosperm vitreousness and inferences were made on possible candidate genes based on previous gene annotations. We found 23 candidates regions in WiDiv association panel using GWAS and 13 regions in the WI-SS-MAGIC population through linkage mapping. Insights into the underlying genetic control of these traits and how they are correlated will help breeders in implementing breeding strategies for quality enhancement.

Acknowledgement

Thanks to Anson Siu for his carefull and hard work preparing the samples and operating the hyperspectral scanner.

Tables

Table 4-1 Summary of phenotypic data distribution and estimated heritabilities of the WI-SS-MAGIC and WiDiv populations

	WI-SS-MAGIC						WiDivs					
	N	**Min**	**Max**	**Mean**	**Range**	**h^2**	**N**	**Min**	**Max**	**Mean**	**Range**	**h^2**
Endosperm Vitreousness (%)	640	0.60	1.00	0.82	0.40	0.84	809	0.63	1.14	0.86	0.51	0.68
Kernel Density (g/cm3)	640	1.00	1.25	1.15	0.25	0.83	809	1.03	1.29	1.15	0.26	0.76
Total kernel Protein (%)	640	4.70	14.10	8.24	9.40	0.77	809	6.54	14.9	9.97	8.36	0.77
Kernel Mass (g)	640	0.21	0.44	0.33	0.23	0.60	809	0.18	0.50	0.32	0.32	0.62

Number of genotypes (N), minimum, maximum, mean and heritability estimates for endosperm vitreousness, kernel density, total protein density and kernel mass for the WI-SS-MAGIC experiment and WiDivs experiments.

Figures

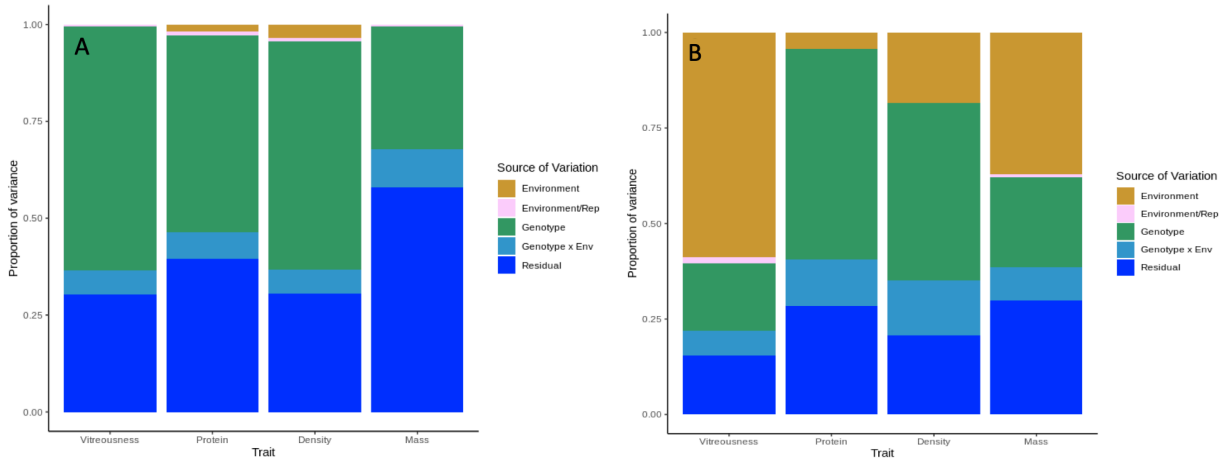


Figure 4-1 Variance components of the WI-SS-MAGICS and the WiDivs.

Proportion of variance components for endosperm vitreousness, kernel protein, density and total mass. A: WI-SS-MAGIC, B: WiDiv population population. Phenotypic information was collected using two field replications per environment in two environments in South Central Wisconsin.

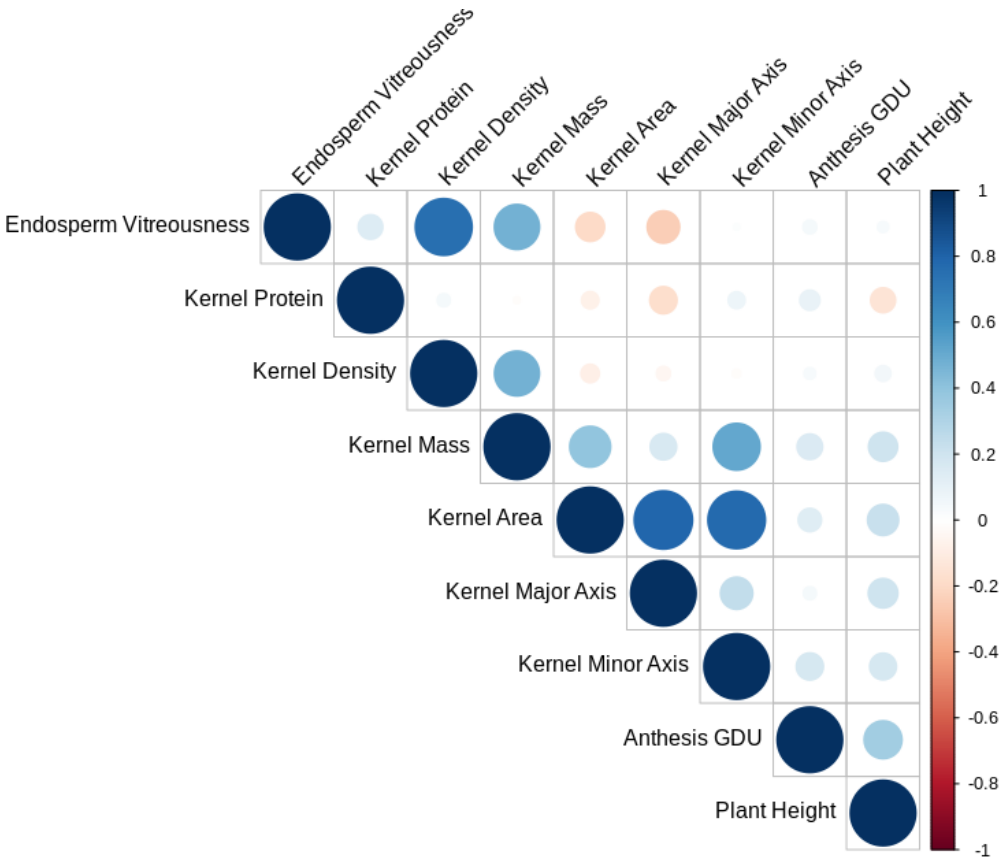


Figure 4-2 Correlation matrix of the WI-SS-MAGIC phenotypes

Pearson correlations were calculated on the BLUES of all phenotypes including the kernel vitreousness (%), kernel density (g/cm^3), protein (%), and mass (g) as well as additional phenological and kernel anatomical characteristics including kernel area (mm^2), kernel major axis (mm), kernel minor axis (mm), anthesis growing degrees units (GDU) and plant height (cm). All traits were evaluated using two field replications per environment in two environments in South Central Wisconsin.

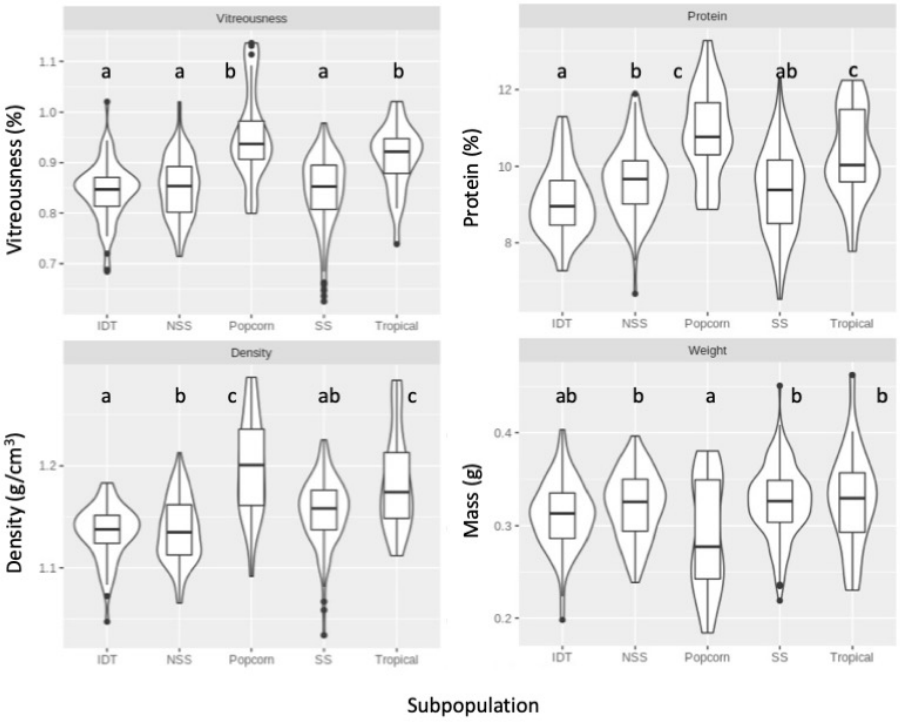


Figure 4-3 Boxplot of phenotypes distributions of the WiDivs by heterotic group and two subgroups

Letters on top of the boxplots represent Tukey's honest significant difference (HSD) between BLUES among heterotic groups and subgroups. Upper left: endosperm vitreousness (%), upper right: kernel protein (%), lower left: kernel density (g/cm3), and lower right: mass per kernel (g). WiDiv association panel set evaluated using two field replications per environment in two environments in South Central Wisconsin.

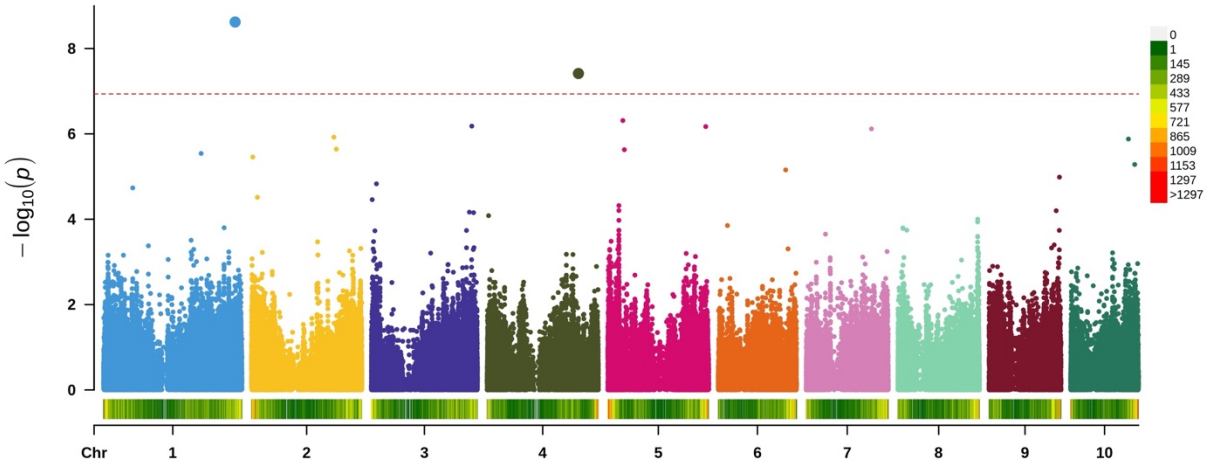


Figure 4-4 Genome wide association study results of endosperm vitreousness.

Manhattan plot of the association study of endosperm vitreousness using the WiDiv population evaluated using two field replications per environment in two environments in South Central Wisconsin. The lower colored bars represent the marker density per Mbp. Red dashed line is the Bonferroni adjusted α-level.

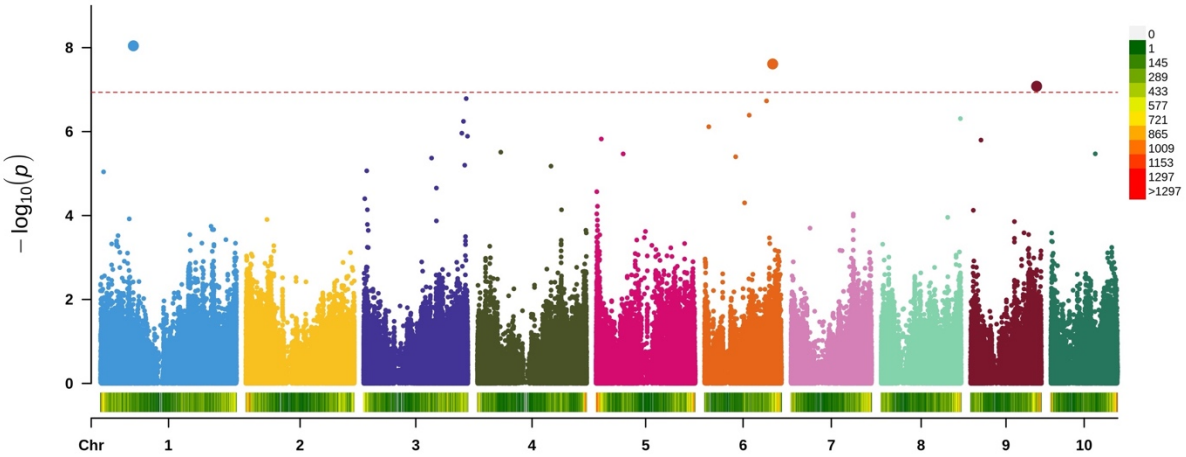


Figure 4-5 Genome wide association study results of kernel density.

Manhattan plot of the association study of kernel density using the WiDiv population evaluated using two field replications per environment in two environments in South Central Wisconsin. The lower colored bars represent the marker density per Mbp. Red dashed line is the Bonferroni adjusted α-level.

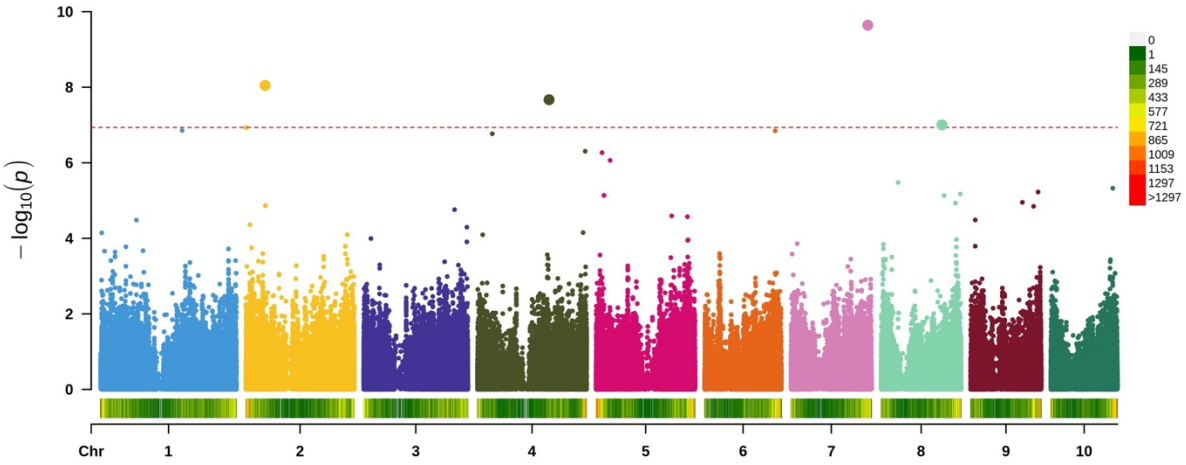


Figure 4-6 Genome wide association study results of kernel mass

Manhattan plot of the association study of kernel mass using the WiDiv population evaluated using two field replications per environment in two environments in South Central Wisconsin. The lower colored bars represent the marker density per Mbp. Red dashed line is the Bonferroni adjusted α-level.

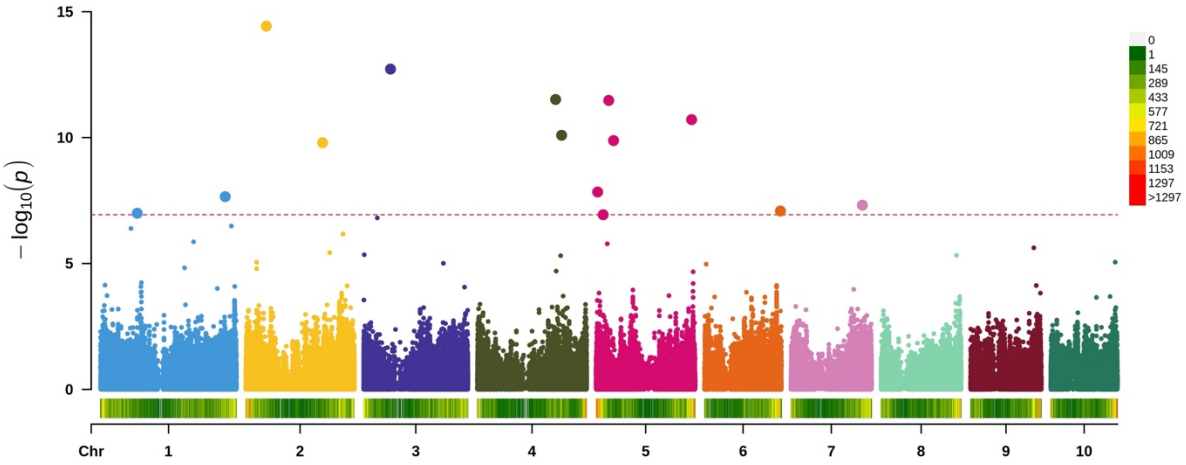


Figure 4-7 Genome wide association study results of kernel protein.

Manhattan plot of the association study of kernel protein using the WiDiv population evaluated using two field replications per environment in two environments in South Central Wisconsin. The lower colored bars represent the marker density per Mbp. Red dashed line is the Bonferroni adjusted α-level.

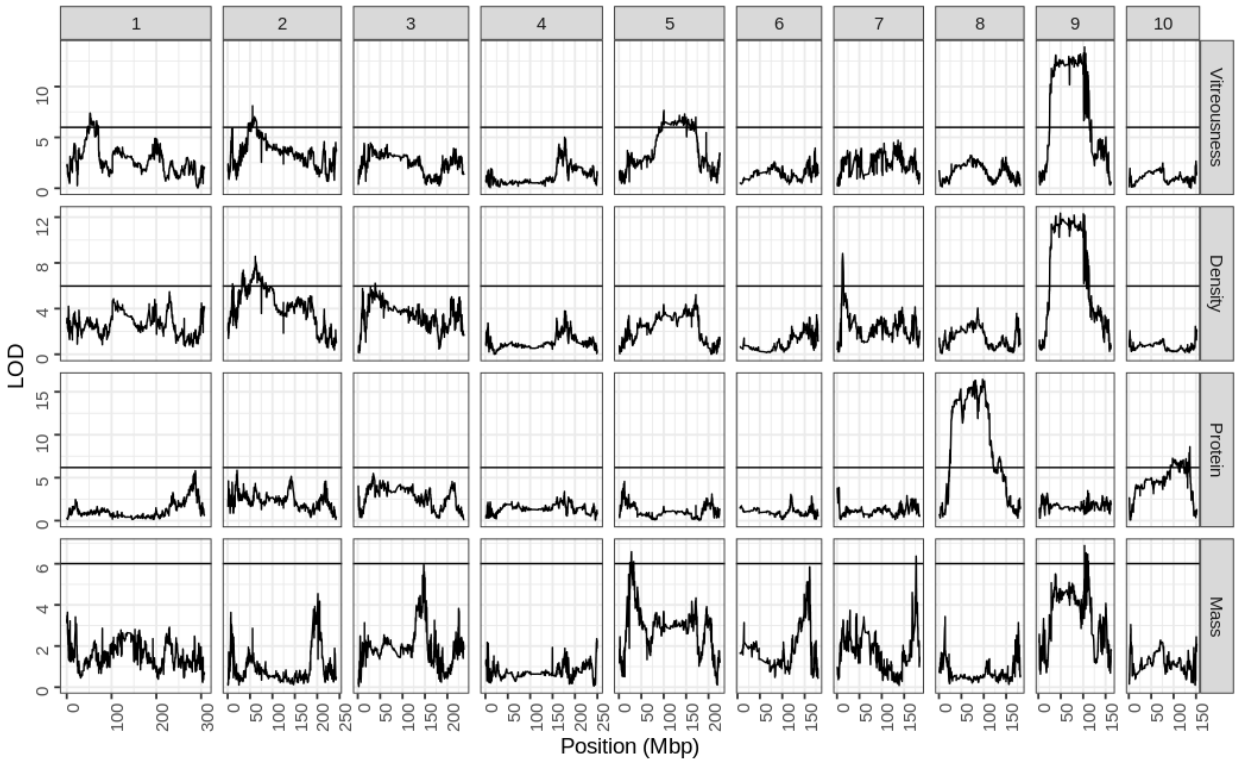


Figure 4-8 LOD score profiles of the WI-SS-MAGIC

LOD scores of endosperm vitreousness, kernel density, kernel protein, and kernel mass, by chromosome for the WI-SS-MAGIC population.

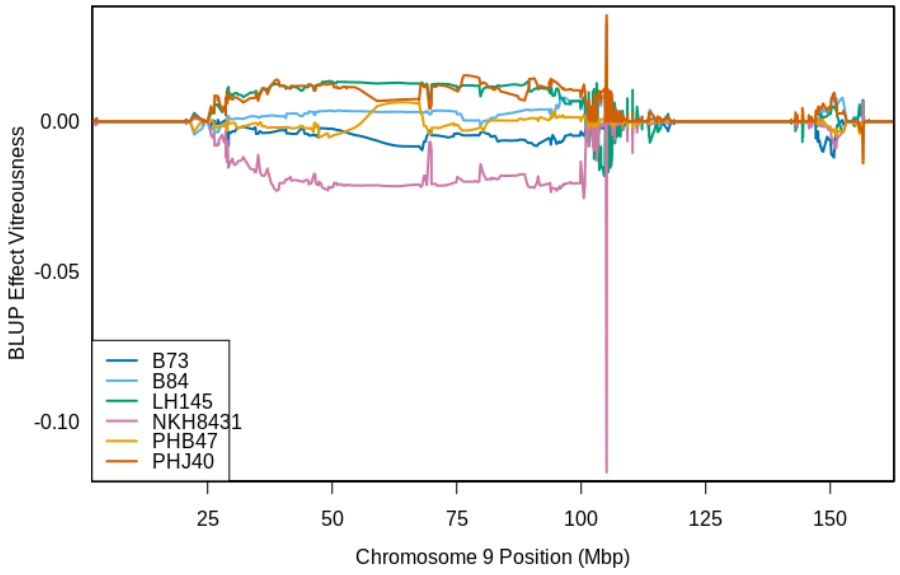


Figure 4-9 WI-SS-MAGICS founders vitreousness QTL BLUP effects

BLUP effects of each parental contribution to kernel endosperm vitreousness in chromosome 9.

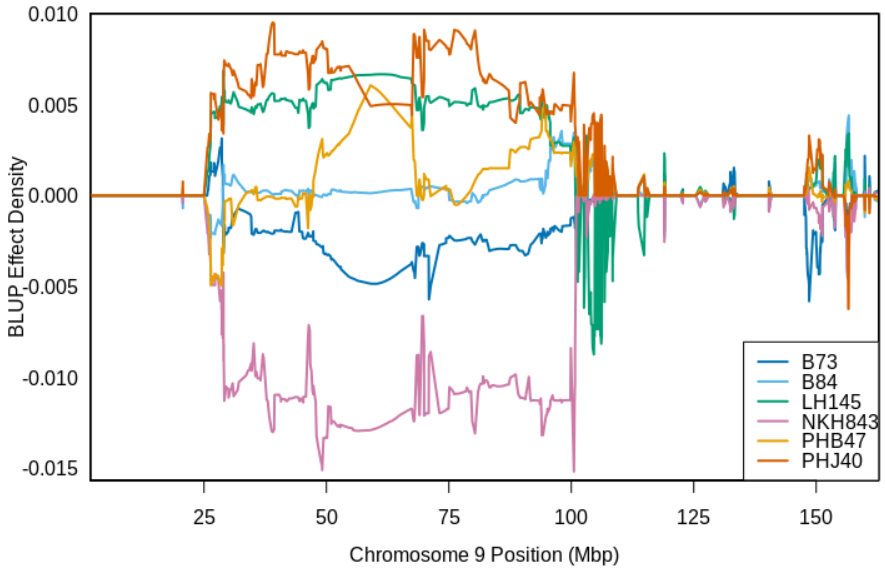
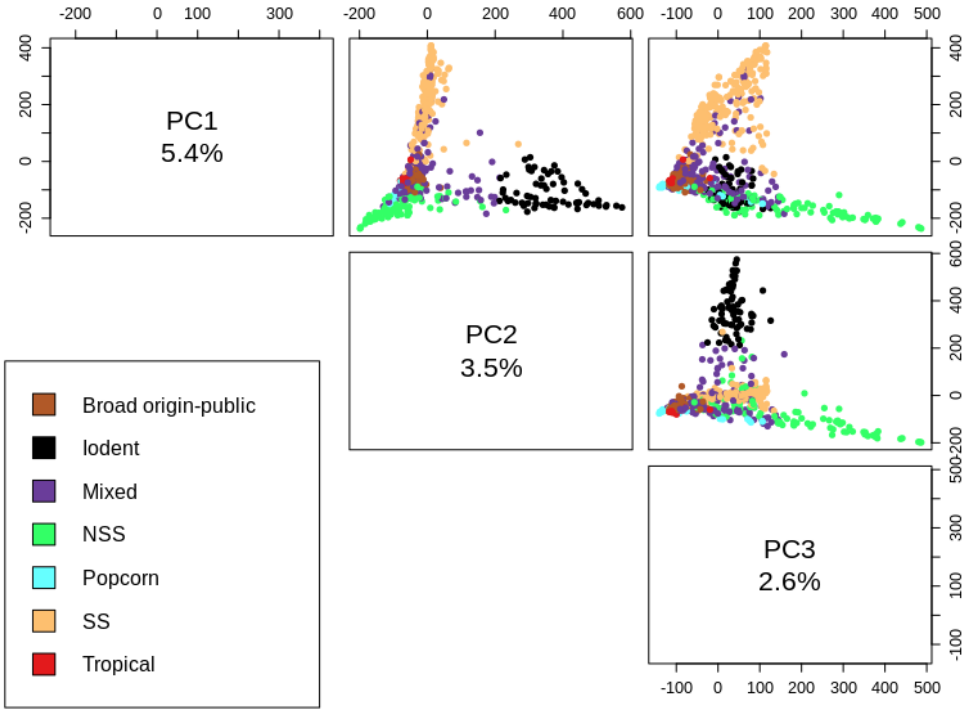


Figure 4-10 WI-SS-MAGIC founders kernel density QTL BLUP effects

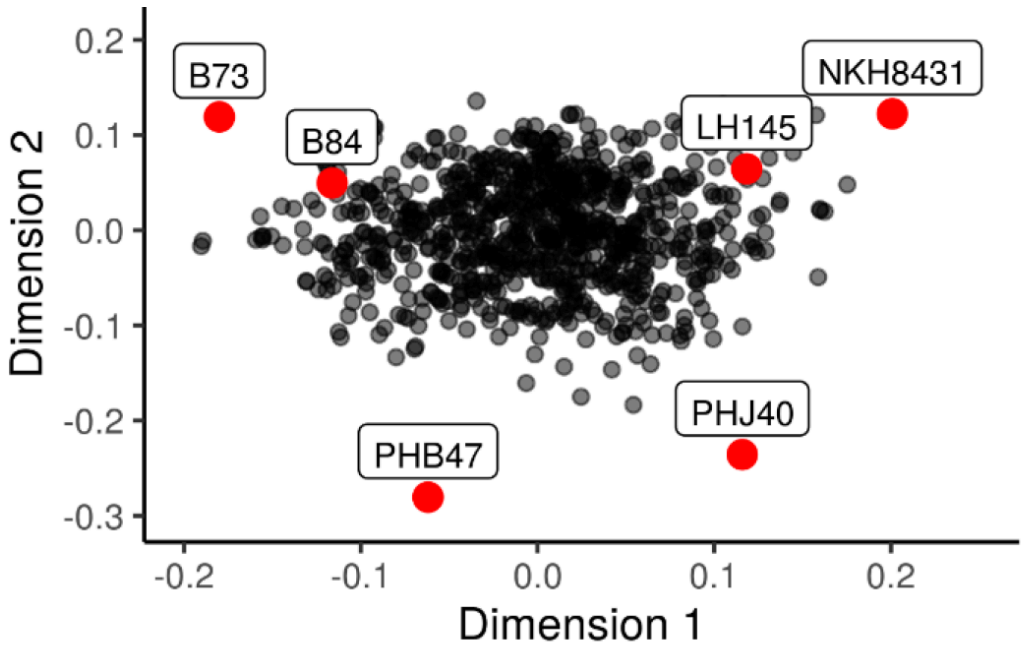
BLUP effects of each parental contribution to endosperm density in chromosome 9.

Supplemental Material



Supplemental Figure 4-1 Principal component analysis of the WiDiv population

Pairwise comparison of the three principal components using 430,695 single nucleotide polymorphisms for the 809 inbred lines used for genome wide association (GWAS) in the WiDiv population.



Supplemental Figure 4-2 Multidimensional scale plot of the WI-SS-MAGIC population

Multidimensional scale plot of the WI-SS- MAGIC lines based on 1.8 million genic SNPs, with the population parents plotted in red. Figure reproduced with permission from Michel et al., 2022.

Supplemental Table 4-1 Significant associations in the WiDivs

Marker	Chromosome	Position	Phenotype	p-value	-log10(p-value)	Effect	SE
rs1_291669572	1	291669572	Vitreousess	2.39E-09	8.62133906	0.013318585	0.002429978
rs4_202627928	4	202627928	Vitreousess	3.85E-08	7.41488515	-0.012887616	0.002561686
rs1_74164047	1	74164047	Density	9.04E-09	8.0437366	0.006039408	0.001141221
rs6_153762466	6	153762466	Density	2.45E-08	7.61135556	-0.007033487	0.00140426
rs9_148924396	9	148924396	Density	8.32E-08	7.07965817	0.0055797	0.001126055
rs2_46121175	2	46121175	Protein	3.75E-15	14.4254657	0.31089844	0.04108047
rs3_60733132	3	60733132	Protein	1.90E-13	12.7203381	-0.21742415	0.03107367
rs4_176600239	4	176600239	Protein	3.07E-12	11.5134463	0.25196132	0.03859558
rs5_28550927	5	28550927	Protein	3.34E-12	11.4759917	-0.19802053	0.03016885
rs5_215519384	5	215519384	Protein	1.94E-11	10.71326	0.33312352	0.05152846
rs4_190121414	4	190121414	Protein	8.09E-11	10.0922933	0.38734449	0.06318274
rs5_39395814	5	39395814	Protein	1.31E-10	9.88223502	0.20316103	0.03362813
rs2_173046942	2	173046942	Protein	1.60E-10	9.79642893	-0.27230905	0.04664379
rs5_3465077	5	3465077	Protein	1.45E-08	7.83893791	-0.15302052	0.02949528
rs1_281348890	1	281348890	Protein	2.22E-08	7.65439183	-0.16971883	0.03292293
rs7_161147611	7	161147611	Protein	4.84E-08	7.3147194	-0.18149393	0.03699902
rs6_171239292	6	171239292	Protein	8.22E-08	7.08505375	0.18088798	0.03949188
rs1_82828274	1	82828274	Protein	1.01E-07	6.99665926	0.2418936	0.04881956
rs5_16001226	5	16001226	Protein	1.16E-07	6.93569479	0.1331535	0.02859098
rs7_173410532	7	173410532	Mass	2.28E-10	9.64211849	-0.00721211	0.001216524
rs2_43285576	2	43285576	Mass	8.98E-09	8.04690709	0.007981579	0.001496867
rs4_161763958	4	161763958	Mass	2.14E-08	7.66902181	-0.012939916	0.002551085
rs8_137410671	8	137410671	Mass	9.93E-08	7.0029629	0.01022944	0.002097731

Significant associations found using Genome Wide Association Stuidies in the WiDiv population evaluated using two field replications per environment in two environments in South Central Wisconsin. SE : Standard error.

Supplemental Table 4-2 Significant associations in the WI-SS-MAGICS

Marker	Chromosome	Position	Phenotype	LOD	CI_LOW	CI_HI	LOD threshold
S1_51967930	1	51967930	Vitreousness	7.389959	45786486	70693537	5.98
S1_55635108	2	55635108	Vitreousness	8.107161	54169650	64584851	5.98
S1_100435764	5	100435764	Vitreousness	7.650565	94908715	171755692	5.98
S1_103107834	9	103107834	Vitreousness	13.888889	37177879	105825873	5.98
S1_97364619	8	97364619	Protein	16.44652	46447894	103805006	6.18
S1_136310343	10	136310343	Protein	8.603477	97604443	136629084	6.18
S1_61744159	2	61744159	Density	8.569425	32919325	71452555	5.97
S1_38868346	3	38868346	Density	6.2148	10517069	227531487	5.97
S1_12693372	7	12693372	Density	8.820569	11455387	14392730	5.97
S1_49134397	9	49134397	Density	12.3475	28979614	103650432	5.97
S1_28063389	5	28063389	Mass	6.586687	21771741	35519817	6.01
S1_176696486	7	176696486	Mass	6.370624	175576924	177449248	6.01
S1_103067265	9	103067265	Mass	6.885856	37177879	111893402	6.01

Significant association found using multiparent linkage mapping in the WI-SS-MAGIC population evaluated using two field replications in two years in South Central Wisconsin. LOD = Logarithm of the odss. CI_LOW = Lowest limit of the confidence interval. CI_HI = Upper limit of the confidence interval.

References

Bae, J.M., Kwak, M.S., Noh, S.A., Oh, M-J., Kim, Y-S., Shin, J.S. (2014). Overexpression of sweetpotato expansin cDNA (IbEXP1) increases seed yield in Arabidopsis. *Transgenic Research* 23:657–667. DOI: 10.1007/s11248-014-9804-1

Bornowski, N., Michel, K, J., Hamilton, J.P., Ou, S., Seetharam, A.S., Jenkins, J., Grimwood, J., Plott, C., Shu, S., Talag, J., et al. (2021). Genomic variation within the maize stiff-stalk heterotic germplasm pool. *Plant Genome*. 14(3):e20114. https://doi.org/10.1002/tpg2.20114

Boston, R.S. Larkins, B.A. (2008). The genetics and biochemistry of maize zein proteins. In: Bennetzen, J.L. and Hake, S.C. (eds.) *The Maize Handbook*. Volume II: History and Practice of Genetics, Genomics and Improvement. Springer, New York, pp. 715–730.

Bradbury, P.J., Zhang, Z., Kroon, D,E., Casstevens, T.M., Ramdoss, Y., Buckler, E,S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23(19):2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brohammer, A,B., Kono, T.J.Y., Springer, N.M., McGaugh, S.E., Hirsch, C.N. (2018). The limited role of differential fractionation in genome content variation and function in maize (Zea mays L.) inbred lines. *The Plant Journal*. 93(1):131–141. https://doi.org/10.1111/tpj.13765

Broman, K.,W, Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., Churchill, G.A. (2019). R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*. 211(2):495–502. https://doi.org/10.1534/genetics.118.301595

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.G., Thompson, R. (2017). Asreml-R reference manual version 4. *VSN International Ltd*, Hemel Hempstead, HP1 1ES, UK.

Caballero-Rothar, N. N., Abdala, L. J., Borras, L., Gerde, J. A. (2019). Role of yield genetic progress on the biochemical determinants of maize kernel hardness. *Journal of Cereal Science*. 87, 301–310

Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E. S., Flint-Garcia, S. A. (2012). Genetic architecture of maize kernel composition in the nested association map- ping and inbred association panels. *Plant Physiology*. 158, 824–834. https://doi.org/10.1104/pp.111.185033

Correa, C.E.S., Shaver, R.D., Pereira, M.N., Lauer, J.G., Kohn, K. (2002). Relationship between corn vitreousness and ruminal in situ starch degradability. *J. Dairy Sci*. *85*: 3008-3012.

Dell'Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A.L., Aung, HH., Nelissen, H., Baute, J., et al. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biology*. 16(1):167. https://doi.org/10.1186/s13059-015-0716-z

Dias Junior, G.S., Ferraretto, L.F., Salvati, G.G.S., de Resende, L.C., Hoffman, P.C., Pereira, M.N., Shaver, R.D. (2016). Relationship between processing score and kernel-fraction particle size in whole-plant corn silage. *Journal of Dairy Science*. *99*(4),2719-2729. http://dx.doi.org/10.3168/jds.2015-10411

Ding, J., Zhang, L., Chen, J., Li, X., Li, Y., Cheng, H., Huang, R., Zhou, B., Li, Z., Wang, J., Wu, J. (2015). Genomic dissection of leaf angle in maize (*Zea mays* L.) using a four-way cross mapping population. *PLOS ONE*. 10(10):e0141619. https://doi.org/10.1371/journal.pone.0141619

Dong, G., Ni, Z., Yao, Y., Nie, X., Sun, Q. (2007). Wheat dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Mol. Biol*. 63, 73–84. doi: 10.1007/s11103-006-9073-9073

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE*. 6(5):e19379. https://doi.org/10.1371/journal.pone.0019379

Feng, F., Lv, Q. W., Yan, T., Xu, S., Yang, L., Yuan, W., Chen, W., Zhao, Y., H., Song, R. (2018). OPAQUE11 Is a central hub of the regulatory network for maize endosperm development and nutrient metabolism. *The Plant Cell*. 30, 375-396

Flint-Garcia, S.A. (2017). Kernel Evolution: From Teosinte to Maize. In B.A. Larkins (Eds), *Maize Kernel Development*. 175-189. CABI.

Flint-Garcia, S.A., Thuillet, A., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *The plant journal*.44,1054-1064

Garrison, E., Marth, G. (2012). FreeBayes. ArXiv preprint 1207.3907. https://doi.org/10.48550/arXiv.1207.3907

Gayral, M., Gaillard, C., Bakan, B., Dalgalarrondo, M., Elmorjani, K, Delluc, C., Brunet, S., Linossier, L., Morel, M-H., Marion, D. (2016). Transition from vitreous to floury endosperm in maize (Zea mays L.) kernels is related to protein and starch gradients. *J. Cereal Sci*. 68, 148–154.

Gibbon, B.., B.A. Larkins. (2005). Molecular genetic approaches to developing quality protein maize. *Trends in Genetetics*. 21:227-233

Gontarek, B.C., Neelakandan, A.K., Wu, H., Becraft, P.W. (2016). NKD transcription factors are central regulators of maize endosperm development. *Plant Cell*. 28: 2916–2936.

Gustafson, T.J., de Leon, N. (2010). Genetic Analysis of Maize (*Zea mays L.*) Endosperm Vitreousness and Related Hardness Traits in the Intermated B73 x Mo17 Recombinant Inbred Line Population. *Crop Science*. 50:2318-2327

Hansey, C.N., Johnson, J.M., Sekhon, R.S., Kaeppler, S.M., de Leon, N. (2011). Genetic diversity of maize association population with restricted phenology. *Crop Science*. 2011;51:704-15. doi: https://doi.org/10.2135/cropsci2010.03.0178.

Hirsch, C.N., Foerster, J.M,. Johnson, J.M., Sekhon, R.S., Muttoni, G., et al. (2014). Insights into the maize pan-genome and pan-transcriptome. *Plant Cell*. 26(1): 121–35. doi: https://doi.org/10.1105/tpc.113.119982.

Holding, D.R., Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica* 51: 243-254.

Holding, D. R., Otegui, M. S., Li, B. L., Meeley, R. B., Dam, T., Hunter, B. G., et al. (2007). The maize floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *Plant Cell.* 19, 2569–2582. doi: 10.1105/tpc.107/053538

Holding, D.R. (2014). Recent advances in the study of prolamins storage protein organization and function. *Frontiers in plant science*. 5,276

Holding, D.R., Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica*. 51, 243-254.

Huang, B.E., George, A.W., Forrest, K,L., Kilian, A., Hayden, M.J., Morell, M.K., et al. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J*. 2012;10:826–39.

Jackson, D.S., Choto-Owen, C., Waniska, R.D., Rooney, L. W. (1988). Characterization of starch cooked in alkali by aqueous high-performance size-exclusion chromatography. *Cereal Chemistry*, 65,493-496.

Jiménez-Galindo, J,C., Malvar, R.A., Butrón, A., Santiago, R., Samayoa, L.F., Caicedo, M., Ordás, B., (2019). Mapping of resistance to corn borers in a MAGIC population of maize. *BMC Plant Biol*. 19(1):431. https://doi.org/10.1186/s12870-019-2052-z

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., et al. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 2009;5, e1000551.

Larkins, B.A., Wu, Y., Song, R., Messing, J. (2017). Maize seed storage proteins. In B.A. Larkins (Eds), *Maize Kernel Development*. 175-189. CABI.

Lending, C. R., Larkins, B. A. (1992). Effect of the floury-2 locus on protein body formation during maize endosperm development. *Protoplasma* 171, 123–133. doi: 10.1007/BF01403727

Liang, Y., Liu, Q., Wang, X., Huang, C., Xu, G., Hey, S., Lin, H-Y., Li, C., Xu, D., Wu, L., Wang, C., Wu, W., Xia, J., Han, X., Lu, S., Lai, J., Song, W., Schnable, P.S., and Tian, F. (2019). ZmMADS69 functions as a flowering activator through the ZmRap2.7-ZCN9 regulatory module and contribute to maize flowering adaptation. *New Phytologist. 221, 2335-2347*

Mahan, A.L., Murray, S.C., Klein, P.E. (2018) . Four-parent maize (FPM) population: development and phenotypic characterization. *Crop Sci*. 58(3):1106–1117. https://doi.org/10.2135/cropsci2017.07.0450

Marzabal, P., Gas, E., Fontanet, P., Vicente-Carbajosa, J., Torrent, M., Ludevid, M. D. (2008). The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Mol. Biol*. 67, 441–454. doi: 10.1007/s11103-008-9325-9325

Mazaheri, M., Heckwolf, M., Vaillancourt, B., Gage, J,L., Burdo, B., Heckwolf, S., Barry, K., Lipzen, A., Ribeiro, C,B., Kono, T.J.Y., et al. (2019). Genome-wide association analysis of stalk biomass and anatomical traits in maize. *BMC Plant Biol*. 19(1):1–17. https://doi.org/10.1186/s12870-019-1653-x

Mena, M., Cejudo, F. J., Isabel-Lamoneda, I., Carbonero, P. (2002). A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone. *Plant Physiol*. 130, 111–119. doi: 10.1104/pp.005561

Miranda, A., Vasquez-Carrillo, G., Garcia-Lara, S., San Vicente, F., Torres, J.L., Ortiz-Islas, S., Salinas-Moreno, Y., Palacios-Rojas, N. (2013). Influence of genotype and environmental adaptation into the maize quality traits for nixtamalization. *Cyta Journal of Food*, 11:54-61

Michel, K., Lima, D.C., Hundley, H., Singan, V., Yoshinaga, Y., Daum, C., Barry, K., Broman, K.W., Buell, C.R., de Leon, N., Kaeppler, S.M. (2022). Genetic mapping of flowering time and platn height in a maize Stiff Stalk MAGIC population. *Genetics*, 221, Issue 2.

McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., et al. (2009). Genetic properties of the maize nested association mapping population. *Science*. 2009;325:737–40.

Philippeau, C., Michalet-Doreau, C. (1997). Influence of genotype and stage of maturity of maize on rate of ruminal starch degradation. *Animal Feed Science Technology*.68, 25-35.

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*. 7(2):e32253. https://doi.org/10.1371/journal.pone.0032253

Qi, W., Lu. L., Huang, S., and Song, R. (2019). Maize Dek44 Encodes Mitochondrial Ribosomal Protein L9 and Is Required for Seed Development. *Plant Physiology*, 180, 2106-2119

Qi, X., Li, S., Zhu, Y., Zhao, Q., Zhu, D., Yu, J. (2017). ZmDof3, a maize endosperm-specific Dof protein gene, regulates starch accumulation and aleurone development in maize endosperm. *Plant Mol. Biol.* 93: 7–20.

Qiu, Y., O'Connor, C.H., Della Coletta, R., Renk, J.S., Monnahan, P.J., Noshay, J.M., Liang, Z., Gilbert, A.M., Anderson, S.N., McGaugh, S.E., Springer, N.M., Hirsch, C.N. (2021) . Whole-genome variation of transposable element insertions in a maize diversity panel. *G3 Genes Genomes Genetics* .11(10): jkab238. doi: https://doi.org/10.1093/g3journal/jkab238.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Renk, J.S., Gilbert, A.M., Hattery, T.J., O'Connor, C.H., Monnahan, P.J., et al. (2021). Genetic control of kernel compositional variation in a maize diversity panel. *Plant Genome.* 14(3): e20115. doi: https://doi.org/10.1002/tpg2.20115.

Reyes, F.G.R., Varseveld, G.W., Kuhn, M.C. (1982). Sugar composition and flavor quality of high sugar (shrunken) and normal sweet corn. *J. Food Sci*. 47:753-755

Saenz, E., Abdala, L, J., Borras, L., Gerde, J.A. (2020). Maize kernel color depends on interaction between hardness and carotenoid concentration. *Journal of Cereal Science*. 91, Article 102901

Seo, M., Akaba, S., Oritani, T., Delarue, M., Bellini, C., Caboche, M., et al. (1998). Higher activity of an aldehyde oxidase in the auxin-overproducing superrootl mutant of Arabidopsis thaliana. *Plant Physiol*. 116, 687–693. doi: 10.1104/pp.116.2.687

Seo, M., Koiwai, H., Akaba, S., Komano, T., Oritani, T., Kamiya, Y., Koshiba, T. (2000). Abscisic aldehyde oxidase in leaves of Arabidopsis thaliana. *Plant Journal*. 23, 481-488. doi: 10.1046/j.1365-313x.2000.00812.x

Stelpflug, S.C, Sekhon, R.S., Vaillancourt, B., Hirsch, C.N., Buell, C.R., de Leon, N., Kaeppler, S.M. (2016). An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. *The Plant Genome*. Volume 9, Issue 1.

Swarup, S., Timmermans, M.C.P., Chaudhuri, S., Messing, J. (1995). Determinants of the high-methionine trait in wild and exotic germplasm may have escaped selection during early cultivation of maize. *Plant Journal*. 8: 359-368.

Tracy, W.F., S.L. Shuler, and H.Dodson-Swenson. (2020). The use of endosperm genes for sweet corn improvement: A review of development in endosperm in endosperm genes in sweet corn since the seminal publication in Plant Breeding Reviews, Volume 1, by Charles Boyer and Jack Shannon (1984). *Plant Breeding Reviews*.

Wang, H., Huang, Y., Xiao, Q., Huang, X., Li, C., Gao, X., Wang, Q., Xiang, X., Zhu, Y., Wang, J., Wang, W., Larkins, B. A., Wu, Y. (2020). Carotenoids modulate kernel texture in maize by

influencing amyloplast envelope integrity. *Nature Communications.* 11, 5346. https://doi.org/10.1038/s41467-020-19196-9

Wang, G., Zhang, J., Wang, G., Fan, X., Sun, X, Qin, H., Xu, N., Zhong, M., Qiao, Z., Tang, Y., and Song, R. (2014). Proline responding1 Plays a Critical Role in Regulating General Protein Synthesis and the Cell Cycle in Maize. *The Plant Cell*. 26m2582-2600

Wang, G., Yuan, Z., Zhang, P., Liu, Z., Wang, T., Wei, L. (2020). Genome-wide analysis of NAC transcription factor family in maize under drought stress and rewatering. *Physiology and Molecular Biology of Plants*. 26,705-717

Wang, G., Wang, F., Wang, G., Wang, F., Zhang, X., Zhong, M.., et al. (2012). Opaque1 encodes a myosin XI motor protein that is required for endoplasmic reticulum motility and protein body formation in maize endosperm. *Plant Cell.* 24, 3447–3462. doi: 10.1105/tpc.112.101360

Wilson, L.M., Whitt, S.R., Ibanez, A.M., Rocheford, T.R., Goodman, M.M., Buckler, E.S.I. (2004). Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*. 16, 2719– 2733.

Wu, J., Chen, L., Chen, M., Zhou, W., Dong, Q., Jiang, H., Cheng, B. (2019). The DOF-Domain Transcription Factor ZmDOF36 Positevely Regulates Starch Synthesis in Transgenic Maize. *Frontiers in Plant Science*. 10, 465.

Xiong, L., Ishitani, M., Lee, H., Zhu, J. K. (2001). The Arabidopsis LOS5/ABA3 locus encodes a molybdenum cofactor sulfurase and modulates cold stress- and osmotic stress-responsive gene expression. *Plant Cell*. 13, 2063–2083. doi: 10.1105/tpc.13.9.2063

Yanagisawa, S. (2002). The dof family of plant transcription factors. *Trends Plant Sci*. 7, 555–560. doi: 10.1016/S1360-1385(02)02362-2362

Yi, G., Neelakandan, A.K., Gontarek, B.C., Vollbrecht, E., Becraft, P.W. (2015). The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. *Plant Physiol*. 167: 443–456.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X. (2021) . rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. *Genomics, Proteomics & Bioinformatics.* 19 (4), 619-628. doi 10.1016/j.gpb.2020.10.007.

Yu, J., Holland, J.B., McMullen, M.D., Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*. 178(1):539–551. https://doi.org/10.1534/genetics.107.074245

Zhang, W., Yan, H., Chen, W., Liu, J., Jiang, C., Jiang, H., Zhu, S., Cheng, B. (2014). Genome-wide identification and characterization of maize expansin genes expressed in endosperm. *Molecular Genetics and Genomics*. 289, 1061-1074.

Zheng, L., Zhang, X., Zhang, H., Gu, Y., Huang, X., Huang, H, Liu, H., Zhang, J., Hu, Y., Li, Y., Yu, G., Liu, Y., Lawson, S.S., Huang, Y. (2019) The miR164-dependent regulatory pathway in developing maize seed. *Molecular Genetics and Genomics*. 294, 501-517

Zhu, C., Gore, M., Buckler, E. S., Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*. https://doi.org/10.3835/plantgenome2008.02.0089

Zhou, Z., Li, G., Tan, S., Li, D., Weiß, T.M., et al. (2020). A QTL atlas for grain yield and its component traits in maize (*Zea mays*). *Plant Breed*. 139(3): 562–574. doi: https://doi.org/10.1111/pbr.12809.